The 21st Annual ACM Symposium on Applied Computing

Dijon, France
April 23 -27, 2006

SYMPOSIUM CHAIR
Hisham M. Haddad, Kennesaw State University, USA

SYMPOSIUM VICE-CHAIR
Richard Chbeir, Bourgogne University, France

PROGRAM CHAIRS
Roger L. Wainwright, University of Tulsa, USA
Sascha Ossowski, University Rey Juan Carlos, Madrid, Spain

TUTORIALS CHAIR
Kokou Yetongnon, Bourgogne University, France

POSTERS CHAIR
Mathew J. Palakal, Indiana University Purdue University, USA

PUBLICATION CHAIR
(Proceedings Editor)
Lorie M. Liebrock, New Mexico Institute of Mining and Technology, USA

TREASURER, WEBMASTER, & REGISTRAR
Hisham M. Haddad, Kennesaw State University, USA

LOCAL ARRANGEMENT CHAIR
Christophe Nicolle, Bourgogne University, France

Poster Papers

Software Engineering: Sound Solutions for the 21st Century

Track Co-Chairs: Stefan Gruner, Swansea Institute, United Kingdom
Sung Shin, South Dakota State University, USA

Does Object Coupling Really Affect the Understanding and Modifying of OCL Expressions?

Luis Reynoso
Department of Computer Science, University of Comahue
Buenos Aires 1400, Neuquén, Argentina
+54 0299 4490313, 8300
lreynoso@uncoma.edu.ar

Marcela Genero
Mario Piattini
Department of Computer Science, University of Castilla-La Mancha, Paseo de la Universidad 4, Ciudad Real, Spain
+34 926295300, 13071
{Marcela.Genero, Mario.Piattini}@uclm.es

Esperanza Manso
GIRO Research Group, Department of Computer Science, University of Campus Miguel Delibes
E.T.I.C, Valladolid, Valladolid
+34 983423710, 47011
manso@infor.uva.es

ABSTRACT

Early and precise models started to play an increasingly relevant role since models themselves become the primary focus in recent initiatives of Model-Driven Engineering (such as Model-Driven Development and Model-Driven Architecture). However, a precise model cannot be obtained through the use of Unified Modeling Language (UML), due to the limited expressiveness of diagram-based UML notation. A textual add-on to the UML diagrams is needed, such as the Object Constraint Language (OCL), for reaching complete and consistent models and avoiding underspecification. Aware of the proliferation of measures for UML-based models and the lack of measures to capture the quality aspects of UML/OCL combined models we defined a set of measures for measuring the structural properties of OCL expressions. This paper carefully describes an experiment we have conducted to confirm the conclusions and strengthen the external validity of a previous family of experiments, with the purpose of investigating the relationship between object coupling in OCL expressions and the understandability and modifiability of OCL expressions. Empirical evidence that such a relationship exists is reaffirmed and consolidated.

Categories and Subject Descriptors

D.2.8 [**Software Engineering**]: Metrics – *Product metrics*.

General Terms

Measurement, Design, Experimentation.

Keywords

Object-Oriented Measures, UML/OCL Models, OCL Expressions, Understandability, Modifiability, Structural Properties, Coupling, Empirical Validation, Controlled Experiments.

1. INTRODUCTION

In the last years UML models [13] and their quality in early stages started to play an increasingly relevant role. UML models become the primary artifacts [21] in recent initiatives as Model-Driven Software Engineering (e.g. Model-Driven Development (MDD) [1] and Model-Driven Architecture (MDA) [14]). Indeed, in Model-Driven engineering the precision of models is necessary to transform a model from Platform-Independent model (PIM) through a Platform-Specific Model (PSM) to code [22]. However, models built with UML only provide a good view of the software architecture [9] but they are imprecise because diagram-based notation is not expressive enough [6]. Consequently the expressiveness of the modeling technique used (e.g. the notation, etc.) affects one of the most important characteristics of a model, its understandability [19]. For solving this problem a textual add-on to the UML diagrams is needed: the Object Constraint Language (OCL) [12]. OCL allows the specification of expressions along with UML-models, obtaining models of high level of maturity [22] which avoid that the models being severely underspecified. Due to the importance of OCL and aware that formal specification can greatly enhance the quality of produced software [9] [21], we have started to study OCL expressions as a crucial add-on to the UML diagrams. In fact, it was empirically proved that OCL has the potential to significantly improve UML-based model comprehension and maintainability [4].

Although the proliferation of measures for UML-based models was growing as UML became the de facto standard language [13] there are no measures for assessing the quality aspects of UML/OCL combined models. So we defined in [17] a set of measures for measuring structural properties of OCL expressions, in a methodological way following a process consisting of three main steps [5]: measure definition, theoretical validation and empirical validation.

As many authors have mentioned [2], [8], [11], [18] empirical validation of measures, through experiments is fundamental to assure that the measures are really significant and useful in practice. So, in [15] and [16] we presented two families of experiments to ascertain whether any relation exists between the measures we defined and maintainability sub-characteristics [10], such as understandability (UND) and modifiability (MOD). In the former family of experiments we obtained that OCL expressions UND and MOD are dependent on how far objects coupled to the contextual instance are and how many different objects are coupled to the contextual instance. In the latter family, we prove that the object coupling (defined through navigations and collection operations) defined in an OCL expression is significantly correlated with the UND and MOD of OCL expressions. In particular we proved that object coupling affects in different way on the UND and MOD of OCL expressions. Regarding the UND or MOD Efficiency (Eff): collection operations, their iterators and the number of classes seem to affect the UND Eff meanwhile the length of navigations and number of relationships influences MOD Eff. However we believe that our conclusion regarding MOD tasks are not solid enough because the correctness of the performed MOD tasks was low. We think that this could have happened because we required the subject to perform MOD tasks writing different OCL expressions, and probably not all of them were ready for writing them from scratch. This fact motivated us to undertake a new experiment using the material of the second family but changing their MOD tasks. This new experiment which intends to strengthen the conclusions and external validity of the second family of experiments, will be described in this paper. The experiment is based on the same hypotheses as the previous experiments, it seems that object coupling (defined through navigations and collection operations) has a great influence on the maintainability of OCL expressions.

This paper starts with the definition of the measures for OCL expressions. Following that, in section 2, the design of the new experiment is presented; meanwhile section 3 provides the data analysis and interpretation. Finally the last section presents some concluding remarks and outlines directions for future research activities.

2. MEASURES FOR OCL EXPRESSIONS

Our hypothesis is that structural properties of an OCL expression within an UML/OCL model (artifacts) have an impact on the cognitive complexity of modelers (subjects), and high cognitive complexity leads the OCL expression to exhibit undesirable external qualities on the final software product [10], such as less understandability or a reduced maintainability [7]. We thoroughly defined in [17] a suite of measures for structural properties of OCL expressions. Table 1 only shows a brief description of the measures we used in the experiment presented in this paper.

We decided to validate object coupling in OCL expressions because coupling is the most complex software attribute in object oriented systems [3] and a high quality software design should obey the principle of low coupling. Furthermore, scanty information of object coupling is available in early stages of software development which only use UML graphical notations, and many times, many coupling decisions are made during implementation [22]. However, at early stages it would be useful the availability of more information about coupling, e.g. to decide which classes should undergo more intensive verification or validation. We believe that a UML/OCL model reveals more coupling information than a model specified using UML only, due to the fact that OCL navigation defines coupling between the objects involved [22], and the coupled objects are usually manipulated in an OCL expression through collections and its collection operations (to handle its elements).

3. DESCRIPTION OF THE EXPERIMENT

Although we followed the experimental process suggested by Wohlin et al. [23] for the sake of brevity we will only show their main characteristics.

Forty six students enrolled on a Software Engineering course at the Technical University of Valencia (UPV) participated in a seminar of 10 hours about OCL in May, 2005. They are students at the 4th year of Computer Science. As an inducement to do the seminar, students were informed that they would do an assessment and its result would be considered to obtain an extra credit as part of the course approval process. The assessment in fact was the experiment. The seminar was conducted by the same teacher who supervised the experiment in a laboratory. In the following we will summarize the main characteristics of the experiment.

Table 1. Measures for OCL expressions defined within UML/OCL models

Measures	Measure Description
NNR	Number of Navigated Relationships
NAN	Number of Attributes referred through Navigations
NNC	Number of Navigated Classes
WNCO	Weighted Number of Collection Operations
DN	Depth of Navigations
WNN	Weighted Number of Navigations
NEI	Number of Explicit Iterator variables
NKW	Number of OCL KeyWords
NES	Number of Explicit *Self*
NCO	Number of Comparison Operators

3.1 Independent and Dependent Variables

The independent variable (IV) is the object coupling of OCL expressions. The dependent variables (DVs) are the UND and MOD of OCL expressions.

3.2 Experimental Materials

The experimental objects were six UML/OCL combined models. Each model includes only one OCL expression. We designed them covering a wide range of the measure values (except in the case of the measures NES, NWK, and NCO). But really, it is impossible to cover all of the possible combinations of measure values. Fifteen models were initially designed, but we thought that some models were quite similar, and the fact of having many models of the same complexity could bias the experiment result. For that reason we carried out a hierarchical clustering of the 15 models to group them into three groups according to their measure values: Low, Medium or High Coupling (we identify each level of coupling by using the acronyms LC, MC, HC respectively). Finally, we obtained two models of each group, i.e. six models.

Each model had a test enclosed including the following tasks:

- Understandability Tasks (UND Tasks): The subjects had to answer a questionnaire consisting of 4 questions that reflected whether or not they had understood the OCL expression attached to the class diagram.
- Modifiability Tasks (MOD Tasks): Two different modifications were required, expressed in natural language. For each modification, the subjects had to select one of three OCL expressions which represent the modification of the original OCL expression (the one associated to the model) according to the new requirement (a multiple choice task). The OCL expressions which should be selected by the subjects had the same measures' values as the original expression associated to the model, i.e. present the same structural properties.
- Rating Tasks: After finishing each task (UND or MOD Tasks) the subject used a scale of five linguistic labels to rate their complexity (e.g. for UND-Tasks we use as the "Easily understandable", "Quite easy to understand", "Normal", "Quite difficult to understand", "Barely Understandable" labels). This rate indicates the perception of the subjects of how complex it was for them to do UND-Tasks or MOD-Tasks.

We assigned to each subjects six tests. The first three tests (and the second three) assigned to each subject had a model of different levels of coupling, i.e. HC, MC or LC. However, the tests were assigned to the subjects in such a way that there are no two subjects doing the six tests in the same order.

In this paper we identify as C_1 the collection of the first tests performed by all the subjects, C_2 the second collection, and so on. It is important to notice that all the six models were examined by the same number of subjects in each C_i.

The IV was measured through the measures shown in Table 1. We used NNR, NNC, WNN, DN, WNCO, NES and NAN measures, because in all of them an aspect of object coupling is captured in their intent [17] through navigation or collection operation concepts. We also use the NEI measure which is related to the collection operation iterator variables, and allows us to define the context inside the collection operations. The rest of the measures NWK (number of keywords) and NCO (number of comparison operators) were not related to collection operations but they are needed to define simple OCL expressions. Because we are not interested in studying the last two measures we try to keep their value as constant as possible. For example all the OCL expressions used as experimental objects were defined with three OCL keywords.

We think that the time the subjects spent doing the required tasks (i.e., UND Time and MOD Time) is not the most accurate measure for the DVs. Therefore we used, the Understandability Efficiency (UND Eff) and the Modifiability Efficiency (MOD Eff), defined as: correct answers / (UND or MOD) Time.

Through the rating tasks we obtained subjective measures of UND and MOD called UND Subjective Complexity (UND SubComp) and MOD Subjective Complexity (MOD SubComp), respectively. These measures are essential to estimate the cognitive load of subjects dealing with UML/OCL combined models.

3.3 Experiment Hypotheses

We formulated different hypotheses along with distinct beliefs:

- Belief 1: The Efficiency of the subjects would be different according the level of object coupling of the UML/OCL models they should to manipulate.

  Hypotheses 1: $H_{0,1}$ The ranks of the (UND or MOD) Eff do not differ from their expected value, i.e. the mean efficiency is the same for all the models. $H_{1,1} = \neg H_{0,1}$
- Belief 2: The object coupling in OCL expressions influences the degree of correctness of the performed tasks per time, i.e. the subject's efficiency (UND Eff or MOD Eff). The greater the object coupling the lower the subjects' efficiency.

  Hypotheses 2: $H_{0,2}$ There is no significant correlation between the measures defined for OCL expressions, related to object coupling and their (UND or MOD) Eff. $H_{1,2} = \neg H_{0,2}$
- Belief 3: The object coupling in OCL expressions influences the subjective rate provided by subjects (UND SubComp or MOD SubComp). This means that if the object coupling increases the subjects perceive the tasks more difficult.

  Hypotheses 3: $H_{0,3}$ There is no significant correlation between the OCL expression measures related to object coupling and the (UND or MOD) SubComp. $H_{1,3} = \neg H_{0,3}$
- Belief 4: The subjective perception of subjects when they have to rate tasks has been influenced by the UND (or MOD) Time. For example, we expect subjects to rate time-

consuming UND tasks as “quite difficult to understand” or “barely understandable”.

Hypotheses 4: $H_{0,4}$ The UND or MOD SubComp are not correlated with the UND and MOD Time. $H_{1,4} = \neg H_{0,4}$

– Belief 5: We believe the degree of correctness of the tasks performed per time, i.e. the UND Eff or MOD Eff, could be an indicator of the subjective rating given by the subjects about the complexity of the required tasks. This means that the perception of the subjects about the complexity of the tasks is influenced by their efficiency when performing such tasks.

Hypotheses 5: $H_{0,5}$ The (UND or MOD) SubComp is not correlated with the UND and MOD Eff . $H_{1,5} = \neg H_{0,5}$.

4. DATA ANALYSIS AND INTERPRETATION

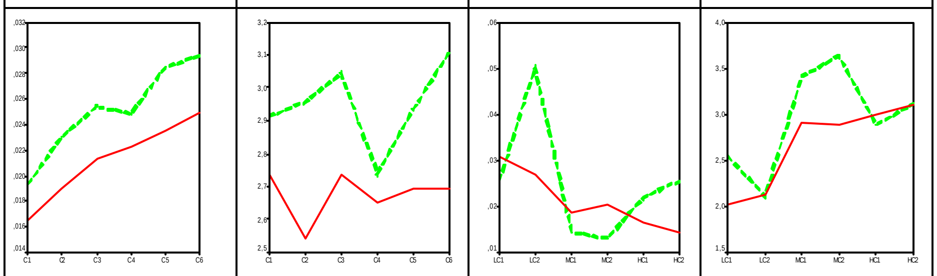
In this section we will summarize the main aspects of the analysis of the empirical data, carried out with SPSS [20]. We are not able to include the statistics tables here for the sake of brevity; nevertheless further information of analysis can be provided by request to the leading author. First we will carry out descriptive and exploratory studies (section 3.1). Later on, we will test the formulated hypotheses.

As all the formulated hypotheses are concerned with dependency degree between two variables, a correlation analysis can be used. Coefficients such as Spearman or t of Kendall, work with pairs of observation, (X_i, Y_j), over *n*-objects (in our case 6 models), but observations must be independent. That means for example, if we study a dependent variable, said UND Eff, of the subject k in the *i*-diagram we are not allowed to consider any other observation of the same *k*-subject. So, the descriptive analysis and the correlations of the formulated hypotheses are tested for each C_i (the *i*-tests performed by all the experimental subjects), ranging i from 1 to 6.

4.1 Descriptive and Exploratory Studies

We depict in the first graph of Table 2 the subjects’ efficiency along with each C_i. In this figure we show that subjects are more efficient in MOD tasks (dash line) than in UND tasks (solid one). Indeed, the subjects were more efficient than in the previous family. As it was previously explained we change MOD tasks, instead of writing a new OCL expression according to a new requirement, the subjects should select one of three proposed OCL expressions which represented an OCL expression modeling the new requirement. However in both kinds of tasks subjects improved their efficiency as time goes on. In relation to the subjective complexity (SubComp) during the time, see the second graph of Table 2, it seems that subjects rated MOD tasks to be more difficult than UND tasks.

Table 2. Descriptive statistics of mean values of DVs

(1) Efficiency in Ci	(2) SubComp in Ci	(3) Efficiency in Model Coupling	(4) SubComp in Model Coupling
			
Simple lines represent UND tasks, Dashed lines represent MOD Tasks			

The collected data grouped by the level of coupling (LC, MC, HC) is depicted in the third and fourth graphs of Table 2, according to the subject efficiency and subjective complexity respectively. The UND Eff is decreasing as the complexity of models increase (note that the horizontal axis shows the two models of each complexity, from low to high complexity). The same was expected to prove in MOD Eff, however as we found in a previous study [16], the subjects were less efficient performing MC models than HC models.

The findings regarding the subjects’ efficiency (grouped by the model complexity) is similar to the findings of their subjective complexity. An increasing subjective complexity is observed in UND Tasks as model complexity increase, however in MOD Tasks, MC models are rated as more difficult than HC models. The main difference between MC and HC models is that in the former the complexity is mainly based on combined navigations, (see the value of WNN) whereas in the latter the complexity is mainly based on an intertwining collection operations (see the value of WNCO). We believe that for the subjects it was more difficult to identify and trace which relationships they should use (its rolename, attribute name, etc.) in MOD Tasks, instead of identifying which operation collections should be used to modify the expression.

Finally, through Shapiro-Wilk tests, we found that the dependent variables do not follow a normal distribution.

4.2 Testing Hypotheses 1

To test the first hypothesis, we used Friedman chi-square test (a non-parametric test for multiple related samples) which tests the null hypothesis that means Eff is the same in all the considered models. The results were significant (p-values were equal to 0.000, less than 0.05), i.e. the mean efficiency of subjects when performing UND and MOD tasks is different according to the complexity of the diagrams.

4.3 Testing Hypotheses 2 and 3

To test the hypotheses 2 and 3, a correlation analysis was performed using Spearman's correlation coefficient with a level of significance a = 0.05, which means the level of confidence is 95% (i.e. the probability that we accept H_0 when H_0 is true is 0.95). We studied the correlation for independent observations, i.e. for each C_i, as it was justified at the beginning of this section. Table 3 summarizes the quantity of significant coefficients we found at level 0.05 between measures and DVs for the C_i . For example, reading the intersection between UND Eff's raw and NNR column, we found in five C_i a significant correlation (below 0.05) between the NNR measure and the UND Eff (in C_2 = 0.0018, C_3 = 0.0000, C_4 = 0.0158, C_5 = 0.0011, C_6 = 0.0022).

In relation to the second hypotheses we concluded that: (1) the NNR, NNC, WNN, DN, WNCO and NEI measures have a strong correlation with the UND Efficiency for almost the six models; (2) the NNR, WNN, DN and NCO have a strong correlation with the MOD Efficiency for almost the six models. It seems that many factors influences the efficiency of UND tasks, as classes, relationships, the navigations, the collection operations and the iterators variables, but only relationships, collection operations and the depth of navigations influences the efficiency of MOD tasks. These results are similar to the one obtained in [16] nevertheless in the results of the current experiment we found a bigger quantity of significant coefficients for the C_i than in the previous family of experiments.

Regarding the third hypothesis: the NNR, NNC, WNN, DN, WNCO and NEI measures are correlated with the subjective complexity of the subject for UND tasks in more than four C_i (more than the half of the C_i, they are six) and NNR, WNN, DN, WNCO and NCO measures are correlated with the subjective complexity of the subject for MOD tasks in more than four C_i.

4.4 Testing Hypotheses 4 and 5

In order to test the 4th and 5th hypotheses, we transformed the variables UND SubComp and MOD SubComp, assigning numbers to the linguistic labels: ranging from 1 (assigned to "Easily understandable/modifiable") to 5 (which correspond with "Barely understandable/ modifiable"). After the data was transformed we used a Kendall's t coefficient to contrast the hypotheses $H_{0,4}$ and $H_{0,5}$. As the results were significant (p-values ranged from 0.000 and 0.018) we can conclude that it seems to exists a statistically significant correlation between the UND/ MOD SubComp variable and the UND/MOD Time, and between UND/MOD SubComp and UND/MOD Eff. Coefficients are negatives for the efficiency and positives for Time, i.e. those tasks rated as difficult were time-consuming tasks and the subjects were less efficient. Moreover, the observed p-values of UND/MOD Eff are smaller (and coefficients are greater) than UND/MOD Time, meaning that the relationship of efficiency and the subjective complexity is stronger than the influence between time and subjective complexity.

Table 3. Quantities of times we found significant correlations (at 0.005 level) between measures and DVs for the C_i

	NNR	NNC	WNN	DN	WNCO	NAN	NEI	NES	NCO
UND Eff	**5**	**6**	**5**	**5**	**6**	**2**	**6**	0	**1**
MOD Eff	**5**	0	**6**	**6**	**2**	0	**1**	0	**6**
UND SubComp	**5**	**4**	**5**	**5**	**6**	0	**4**	0	**1**
MOD SubComp	**6**	0	**6**	**6**	**4**	**1**	**1**	0	**6**

5. CONCLUSIONS

In the current paper we run a new experiment in order to confirm the results of previous experiments [16] in which we investigated if the object coupling (defined through navigations and collection operations) affects the UND and MOD of OCL expressions. The experiment was run in May 2005 at the Technical University of Valencia with forty six students of the 4th year of Computer Science.

In order to study the UND and MOD of the OCL expressions we have considered not only the time subjects spent on tasks related to this activities, but also their efficiency and their subjective perception of their activities. We think that quantitative (UND and MOD efficiency) and qualitative (subject's rating of their cognitive load) information is important to obtain more solid findings.

From the descriptive and exploratory analyses, we can conclude that efficiency of subjects when they have to understand and modify the OCL expressions, increases as time passes. We obtain a considerable improvement of subjects' efficiency in MOD tasks, and this was the most important fact to run a new experiment (in this new experiment MOD tasks differ considerably from those in the previous experiments where subjects' MOD efficiency was low). Moreover, subjects were more efficient in MOD tasks than in UND tasks. The UND and MOD efficiency is different if we arrange the collected data according to the model complexity. The higher the complexity of the OCL expression, the lower is the UND efficiency spent by subjects. However, models of Medium complexity (MC) were more difficult to modify for the subjects than models of High complexity (HC). The former (MC models) were models in which students should have traced from different navigations. In the latter (HC models) subjects should identify which collection operations they have to use and how to combine them. We think that their experience in collection operations and the activities of tracing relationships between classes could produce this situation.

We found that there is a statistically significant relationship between many measures and the UND and MOD Efficiency. For example, the number of classes (NNC) used in navigations, the number of collection operations (WNCO) and the number of collection operation's iterator variables (NEI) influences the subjects' UND efficiency. The number of navigations used in navigation (NNR), WNN and more important the length of navigations (DN) has a correlation with the cognitive load when subjects rate MOD Tasks.

The subjects' subjective ratings (UND or MOD rating) are influenced by the time they spent for understanding or modifying the OCL Expressions, i.e. both times seem to affect their perception of the level of complexity of an OCL expression. Furthermore the subjective ratings are also influenced by the UND and MOD efficiency.

As the results reveals there is empirical evidence that object coupling defined in an OCL expression through navigations and collection operations is significantly correlated with the maintainability of OCL expression. So, we have planned to replicate this experiment with students at the Complutense University of Madrid. In this way we will strengthen the conclusions and external validity respectively. The empirical validation of the rest of the measures is also pending. Moreover, we will work in a generalization of the benefits of the set of measures defined for OCL expressions, trying to obtain a global complexity of UML/OCL models (note that all the proposed measures are defined in terms of a single OCL expression).

6. ACKNOWLEDGMENTS

This research is part of the ENIGMAS project (PBI-05-058) financed by "Consejería de Ciencia y Tecnología de la Junta de Comunidades de Castilla-La Mancha", the CALIPO project supported by "Dirección General de Investigación del Ministerio de Ciencia y Tecnología (Spain)" (TIC2003-07804-C05-03) and the network VII-J-RITOS2 financed by CYTED and MEC-FEDER project TIN2004-03145.

7. REFERENCES

[1] Atkinson, C., Kühne, T.: Model-Driven Development. A Metamodeling Foundation, *IEEE Software*, 20(5), 2003, pp. 36- 41.

[2] Basili, V. R., Rombach, H. D. The TAME project: towards improvement-oriented software environments. *IEEE Trans. on Softw. Eng.*Vol.14 Nº 6, 1998, pp. 758-773.

[3] Briand, L. C., Bunse, L. C., Daly, J. W. A Controlled Experiment for evaluating Quality Guidelines on the Maintainability of Object-Oriented Designs. *IEEE Trans. on Softw. Eng.*, Vol. 27 Nº 6, 2001, pp. 513-530.

[4] Briand, L. C., Labiche, Y.,Yan, H. D., Di Penta, M. A controlled Experiment on the Impact of the OCL in UML-based Maintenance. *IEEE Int. Conference on Softw. Maintenance*, 2004.

[5] Calero, C., Piattini, M., Genero, M. Method for Obtaining Correct Metrics. In *Proc. of the 3rd Int. Conference on Enterprise and Information Systems* (ICEIS`2001), 2001, pp. 779-784.

[6] Cook, S., Kleepe, A., Mitchell, R., Rumpe, B., Warmer, J., Wills, A. The Amsterdam Manifiesto on OCL. *Advances in Object Modelling with the OCL*, Springer, Berlin, LNCS 2263, 2001, pp. 115-149.

[7] El-Eman, K. *Object-Oriented Metrics: A Review of Theory and Practice.* National Research Council Canada. Institute for Information Technology. 2001.

[8] Fenton, N. E., Pfleeger, S.L. *Software Metrics: A Rigorous and Practical Approach.* Chap-man & Hall, London, 2nd Edition. International Thomson Publishing Inc. 1997.

[9] Giese, M., Heldal, R. *From Informal to Formal Specification in UML.* UML 2004, LNCS 3273, pp. 197-211, 2004.

[10] ISO/IEC 9126. *Software Product Evaluation-Quality Characteristics and Guidelines for their Use.* Geneva.

[11] Kitchenham, B., Pflegger, S., Fenton, N. Towards a Framework for Software Measurement Validation. *IEEE Trans. of Softw. Eng.*,Vol. 21 Nº 12, 1995, pp. 929-944.

[12] Object Management Group. *UML 2.0 OCL 2nd revised submission. OMG Document.* Available at http://www.omg.org

[13] Object Management Group. *UML Specification Version 1.5, OMG Document.* Available at http://www.omg.org

[14] Object Management Group. *MDA The OMG Model Driven Architecture.* Available: http://www.omg.org./mda/, 2002.

[15] Reynoso, L., Genero, M., Piattini, M. Validating OCL metrics through a Family of Experiments. *Jornadas de Ingeniería del Sofware y Base de Datos.* 2004.

[16] Reynoso, L., Genero, M., Piattini, M. Assessing the impact of coupling on the understandability and modifiability of OCL expressions within UML/OCL combined models. *Metrics 2005*.

[17] Reynoso, L., Genero, M., Piattini, M. Measuring OCL Expressions: An approach based on Cognitive Techniques. *In Metrics for Conceptual Models.* Imperial College Press, UK. 2005.

[18] Schneidewind, N. F. Methodology for Validating Software Metrics. *IEEE Trans. of Softw. Eng.*, Vol. 18 Nº 5, 1992, pp. 410-422.

[19] Selic, B. The Pragmatics of Model-Driven Development. *IEEE Software.* Vol. 20, Nº5, 2003, pp 19 –25.

[20] SPSS, 2002 SPSS 11.5. *Syntax Reference Guide.* Chicago. SPSS Inc. 2002.

[21] Vinter, R., Loomes, M., Kornbrot R. Applying Software Metrics to Formal Specifications: A Cognitive Approach. 5th. *Int. Symposium on Softw. Metrics*. March 20 - 21, 1998. pp 216-223. Bethesda, Maryland

[22] Warmer, J., Kleppe, A. *The Object Constraint Language. Second Edition. Getting Your Models Ready for MDA.* Addison-Wesley, Massachusetts, 2003.

[23] Wohlin, C., Runeson, P., Höst, M., Ohlson, M., Regnell, B., Wesslén, A. *Experimentation in Softw. Eng.: An Introduction*, Kluwer Academic Publishers, 2000.