The Effect of Coupling on Understanding and Modifying OCL Expressions: An Experimental Analysis

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

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

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

Abstract

Not all the model constraints can be defined using only UML graphical features due to the limited expressiveness of diagram-based UML notation. To solve this problem, the OCL language was defined as a textual add-on to the UML diagrams, allowing the specification of a wide range of constraints of objects.

Aware of the lack of metrics to capture the quality aspects of UML/OCL models and the importance of models in recent initiatives of model-Driven software engineering (such as Model-Driven Development and Model-Driven Architecture), we define a set of metrics for measuring the structural properties of OCL constraints in UML/OCL models. Many of these metrics are defined in terms of navigations, a core concept of OCL that defines coupling between objects. This paper carefully describes a family of experiments we have conducted to ascertain whether any relationship exists between object coupling (defined through metrics related to navigations and collection operations) and two maintainability subcharacteristics: understandability and modifiability of OCL expressions. Empirical evidence that such a relationship exists is shown in the obtained results, however they must be considered as preliminaries results. Further validation is needed to strengthen the conclusions and external validity.

1. Introduction

Within the Object Oriented (OO) software development process, the importance of models is gradually becoming an essential aspect. This fact is corroborated by recent initiatives such as Model-Driven Development (MDD)[1] and the Model-Driven Architecture (MDA) [19], which are based on the assumption that models are the basis of the software development, and they constitute its primary focus and products [23]. Currently, the Unified Modelling Language (UML) [18] is the standard language in software development. However, UML models only provide a good view of the software architecture [13] and they are imprecise because diagram-based notation is not expressive enough [10]. The expressiveness of the modelling technique used (e.g. the notation, etc.) affects one of the most important characteristics of a model, its understandability [23].

Modellers can only obtain models of a high level of maturity using the combination of UML and the OCL language [17], otherwise their models would be severely underspecified [26]. Due the importance of OCL, and aware that formal specification can greatly enhance the quality of produced software [13] [25], 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 [6].

As our interest was to evaluate quality aspects of UML/OCL combined models, we defined in [21] a set of metrics for OCL expressions in a methodological way. We followed a process consisting of three main steps [7]: metric definition, theoretical validation and empirical validation. As many authors have mentioned [2], [12], [15], [22] empirical validation of metrics, through experiments is fundamental to assure that the metrics are really significant and useful in practice. So, in [20] we presented a family of experiments (an experiment and its two replicas) to ascertain whether any relation exists between the navigation depth (measured by DN) and the quantity of different objects coupled (NNC) of an OCL expression and its understandability and maintainability. Through experimentation we obtained that OCL expressions understandability and modifiability are dependent on how far objects coupled to the contextual instance are and how many different objects are coupled to the contextual instance.

We believe that the coupling defined in an OCL expression is significantly correlated with the understandability and modifiability of OCL expressions, and we still need to focus our efforts on the empirical proof of new results. We decided to continue validating object coupling because coupling is the most complex software attribute in object oriented systems [5] 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 decision are made during implementation [26]. However, it would be useful the availability of more coupling information of a model at early stages, 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 [26], and the coupled objects are usually manipulated in an OCL expression through collections and its collection operations (to handle its elements). Therefore, the goal of this paper is to carefully describe a second family of experiments we have undertaken to ascertain if any relationship exists between the object coupling (defined through navigations and collection operations), and two maintainability sub-characteristics [14]: understandability and modifiability of OCL expressions.

This paper starts with a description of the definition of the metrics for OCL expressions. Following that, in section 3 a description of a family of experiments is presented. Section 4 provides the data analysis and interpretation. Finally the last section presents some concluding remarks and outlines directions for future research activities.

2. Metrics for OCL Expressions

Because our intention is the metric definition, and traditional metrics can be supported by the fact they are clearly related to cognitive limitations [4] we have considered the cognitive techniques applied by modellers during OCL comprehension and modification in the metric definition. In this way, we have taken into account the cognitive complexity (the mental burden of a person when he/she deals with artifacts) of modellers when they use 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 modellers (subjects), and high cognitive complexity leads the OCL expression to exhibit undesirable external qualities on the final software product [14], such as less understandability or a reduced maintainability [11].

We have also hypothesized that during the comprehension of an OCL expression the modellers concurrently and synergistically apply two cognitive techniques [8]: "chunking" and "tracing". The former involves the recognition of a set of declarations and the extraction of information from them, which is remembered as a chunk (a single mental abstraction), whereas the latter involves scanning, either forward or backwards, in order to identify pertinent chunks. So, studying OCL expressions as a chunk unit, we have defined a set of metrics considering the OCL concepts related to these cognitive techniques. Analysis of each of these techniques in turn leads to identification of structural properties which can be measured. In order to identify the broad set of OCL concepts, and not omit any of them, we have studied the OCL metamodel.

We thoroughly defined in [21] a suite of metrics for structural properties of OCL expressions. Table 1 only shows a brief description of the metrics we used in the family of experiments presented in this paper and the cognitive technique they are related to. In the fourth column of Table 1 we partially show the result of the theoretical validation (only for the metrics used in this experiment) carried out following Briand et al.´s frameworks.

Metric	Cognitive technique	Metric Description	Theoretical Validation		
			IBC*	S*	L*
NNR	Tracing	Number of Navigated Relationships	Yes		
NAN	Tracing	Number of Attributes referred through Navigations	Yes		
NNC	Tracing	Number of Navigated Classes	Yes		
WNCO	Tracing	Weighted Number of Collection Operations	Yes		
DN	Tracing	Depth of Navigations			Yes
WNN	Tracing	Weighted Number of Navigations		Yes	
NEI	Chunking	Number of Explicit Iterator variables		Yes	
NKW	Chunking	Number of OCL KeyWords		Yes	
NES	Chunking	Number of Explicit *Self*		Yes	
NCO	Chunking	Number of Comparison Operators		Yes	
* IBC stands for *Interaction Based for Coupling*, S stands for *Size* and L stands for *Length*					

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

3. Family of Experiments

Relevant results can only be obtained by families of experiments rather than individual experiments. In other words, simple studies rarely provide definite answers [16] [3]. So, in order to fulfil the experiment goal previously defined in the introduction, we ran a family of experiments, consisting of three experiments, executed in three universities of different countries. Although we followed the experimental process suggested by Ciolkowski et al. [9] and Wohlin et al. [27], for the sake of brevity we will only show their main characteristics:

- First Experiment (April 2004): We invited the third-year students of Computer Science at the University of Alicante (UA, Spain) to do a short seminar about OCL (only 5 hours) and to do an experiment as part of the seminar. Sixty undergraduate students agreed to take part in a course. They were motivated to participate in the experiment because they would be able to obtain an extra point in the final score of the Software Engineering course if and only if they completed a test. The collected data was called “UAE”.
- First Replica (October 2004): Twenty six students who participate in a course of the Eighth International School of Computer Science (celebrated in La Matanza University, Argentina) were the subjects of the first replica. The duration of the course was 20 hours and during the last two hours we ran the experiment replica. The subjects were undergraduate students of different universities, graduate students and teachers. The data obtained in this replication, was called “ULME” data.
- Second Replica (November 2004): Twenty nine students of fifth year enrolled on a Software Engineering course of the Austral University of Chile participated in a course of 20 hours about OCL. As an inducement to do the course, students were informed that they would do a test and its result would be considered as a point of the course of Software Engineering. The collected data was called “UAChE”.

The training sessions of the experimental subjects, seminar or courses, were conducted by the same teacher. The three experiments were carried out with supervision in a laboratory.

3.1. Common Characteristics of the Family

In this section we will summarize the main experimental process steps common to the three experiments.

Independent and dependent variables: The independent variable (IV) is the object coupling of OCL expressions. The dependent variables (DVs) are two maintainability sub-characteristics: understandability and modifiability.

Experimental Material: The experimental objects were nine UML/OCL combined models, each model having one OCL expression. We designed them covering a wide range of the metric values (except in the case of NES, NWK, and NCO). But in reality, it is impossible to cover all of the possible combination of metrics 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 metric values: Low, Medium or High Complexity (we identify each complexity by using the acronyms LC, MC, HC respectively). Finally, we obtained three models of each group.

1. 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.
2. Modifiability tasks (MOD-Tasks): The subjects had to modify the OCL expressions according to a new requirement expressed in natural language.
3. Rating Tasks: After finishing each task (UND or MOD Tasks) the subject uses a scale of five linguistic labels to rate them (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.

All three tests assigned to any subject had three different complexities, i.e. HC, MC or LC, which means there is no subject doing two tests of the same complexity. However, the tests were randomly assigned to the subjects. 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 C_3 the third one. It is important to notice, that all the nine models are examined by the same number of subjects in each C_i.

The independent variable was measured through the metrics shown in Table 1. We used NNR, NNC, WNN, DN, WNCO, NES and NAN metrics, because in all of them an aspect of the navigation concept is captured in its intent [21]. We also use the NEI metric which is related to the collection operation iterator variables, and allows us to define the context inside the collection operations. The rest of the metrics 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 metrics 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 each subject spent doing each 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:

- UND Eff = correct answers/UND Time
- MOD Eff = correct modifications/MOD Time.

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

Experiment Hypotheses: We formulated different hypotheses along with distinct beliefs:

- Belief 1: The structural properties related to 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 hypotheses are:
  Hypotheses 1: $H_{0,1}$ There is no significant correlation between the metrics defined for OCL expressions (see Table 1), related to object coupling and their UND Eff /MOD Eff. $H_{1,1} = \neg H_{0,1}$
- Belief 2: The structural properties related to object coupling in OCL expressions influences the subjective rate provided by subjects (UND SubComp or MOD SubComp) tasks. If so, we will be able to find an early indicator of the subject's cognitive load. The hypotheses are:
  Hypotheses 2: $H_{0,2}$ There is no significant correlation between the OCL expression metrics related to object coupling and the SubComp Eff. $H_{1,2} = \neg H_{0,2}$
- Belief 3: The subjective criteria 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". The hypotheses are:

Hypotheses 3: $H_{0,3}$ The UND or MOD SubComp are not correlated with the UND and MOD Time. $H_{1,3}$: ¬ $H_{0,3}$

- Belief 4: 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, The hypotheses are:
  Hypotheses 4: $H_{0,4}$ The UND or MOD SubComp is not correlated with the UND and MOD Eff . $H_{1,4}$: ¬ $H_{0,4}$

4. Data Analysis and Interpretation

In this section we will summarize the main aspects of the analysis of the empirical data, carried out with the SPSS software [29]. Further information of the family of experiments, can be provided by request to the leading author.
The analysis of the empirical data is laid out as follows

- First we will carry out a descriptive and exploratory study (section 4.1).
- 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 Tau of Kendall, work with pairs of observation, (X_i, Y_j), over n-objects (in our case 9 diagrams), but observations must be independent. That means for example, if we study a dependent variable, said UND Eff, of the subject "j" in the i-diagram we are not allowed to consider any other observation of the same j-subject. So, the correlations of the formulated hypotheses are tested for each C_i (which represents the i-tests performed by all the experimental subjects). In same way, studying the correlation for each C_i will indicate whether our hypotheses are dependent on the learning curve of subjects during the experiment.

4.1. Descriptive and Exploratory Study

The fact that the dependent variables do not follow a normal distribution was corroborated using the Shapiro-Wilk tests. Outliers were removed for all the DVs. As previously described, the set of C_i represents the order of the performed tasks, which allows us to show how the time spent on each task decreases as new tasks are solved by subjects. After analysing the UND and MOD time as time passed, we realized the time decreased during the experiment's execution. In the case of UND Eff and MOD Eff, we expected the subject rump up efficiency but it does not improve as time goes on, except in the UA experiment for UND Eff Time. However if we arrange the collected data according to their complexity (HC, MC or LC), the UND Time and UND Eff improves as the complexity diminish. This is not the case for MOD Time and MOD Eff because the Medium Complexity (MC) tasks were more difficult to modify than the tasks corresponding to High Complexity. This situation occurs in the three experiments. The main difference between MC and HC models is that in the former the complexity is mainly based on combined navigations, whereas in the latter the complexity is mainly based on an intertwining collection operations. 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. We think that complementing the UML class diagram with a natural language description of the intent of the relationship would have been useful to the subject to realize what relationship they must use.

The descriptive statistics for the mean of the UND Time and the mean of the MOD Time have higher values in UAChE compared with ULME and UAE, and between the last two, the smallest mean values are from UAE. Chilean students have low experience in UML, so they required a certain amount of necessary extra time to undertake any task. Although UAE presents a higher mean UND Time than ULME their UND Eff are similar, if we compare the C_i.

4.2. Testing hypotheses 1 and 2

To test the first two hypotheses, a correlation analysis was performed using Spearman's correlation coefficient with a level of significance $\alpha = 0.05$, which means the level of confidence is 95% (i.e. the probability that we accept H_0 when

	NNR	NNC	WNN	DN	WNCO	NAN	NEI	NES	NCO
UAE UND Eff C_1	0.250	**0.021**	0.517	0.263	**0.028**	0.124	0.252	0.360	0.903
UAE UND Eff C_2	**0.035**	**0.042**	**0.027**	0.430	**0.026**	0.641	**0.029**	0.194	**0.047**
UAE UND Eff C_3	0.446	**0.002**	0.810	0.843	**0.000**	**0.000**	**0.000**	**0.019**	0.051
UAChE UND Eff C_1	0.152	**0.001**	0.938	0.590	0.057	**0.011**	**0.005**	**0.037**	**0.005**
UAChE UND Eff C_2	0.175	0.154	0.072	0.201	**0.030**	0.808	0.099	0.911	0.710
UAChE UND Eff C_3	0.404	0.696	0.769	0.488	0.585	0.674	0.670	0.765	0.747
ULME UND Eff C_1	0.278	0.150	0.279	0.484	0.066	0.147	0.053	0.350	0.698
ULME UND Eff C_2	0.440	0.993	0.677	0.982	0.748	0.762	0.970	0.456	0.132
ULME UND Eff C_3	0.987	0.338	0.760	0.311	0.126	**0.048**	0.083	**0.017**	0.296
UAE MOD Eff C_1	0.201	0.403	0.061	**0.000**	0.329	0.061	0.316	**0.000**	**0.015**
UAE MOD Eff C_2	0.479	0.851	0.794	0.689	0.072	**0.049**	0.059	0.118	0.584
UAE MOD Eff C_3	0.335	0.230	0.052	**0.001**	0.273	**0.011**	0.264	**0.000**	**0.004**
UAChE MOD Eff C_1	0.117	0.364	0.685	0.413	0.532	0.907	0.953	0.954	0.751
UAChE MOD Eff C_2	**0.031**	0.810	**0.029**	**0.010**	0.545	0.381	0.400	0.557	**0.037**
UAChE MOD Eff C_3	**0.005**	0.084	0.130	0.116	0.824	0.694	0.617	0.857	0.270
ULME MOD Eff C_1	0.166	0.374	0.479	0.057	0.903	0.680	0.977	0.241	0.831
ULME MOD Eff C_2	**0.028**	0.485	0.081	**0.010**	0.485	**0.035**	0.395	**0.021**	0.181
ULME MOD Eff C_3	0.353	0.825	0.241	0.638	0.471	**0.032**	0.186	0.312	0.543

Table 1. Spearman´s correlation coefficient between Metrics and UND/MOD Eff (significant coefficients at level 0.05 are shown in bold font).

H_0 is true is 0.95). Table 2 show the p-value of the Spearman's significant coefficient between metrics and efficiency' DVs. The conclusions are:

Hypotheses 1: All the metrics present a negative correlation coefficient, except several metrics as NAN and NCO in MOD Eff and NES and NCO in UND Eff in same observations within subjects. A negative coefficient means that the subjects are less efficient when the values of a metrics a high, otherwise they are more efficient.

- The NNC, WNCO and NEI metrics have several correlations with the UND Eff in the UAE and UAChE. This is logical, meaning that the number of classes involved in the OCL expressions (NNC), the number of collection operation (WNCO) and the number of collection operation's iterator variables (NEI) influences the subjects' efficiency. This influence seems to be independent of the order of the tasks performed for UAE because we find a correlation for most of the C_i.
- The length of the navigation (DN) has correlations with the MOD Eff in the three experiments. NNR, NAN, NES and WNN have also correlations with the MOD Eff, but not for the three experiments. NAN, NES and NCO have a positive correlation coefficient, i.e. the subjects are more efficient when the values of the metrics are higher.

Hypotheses 2: All the metrics present a positive correlation coefficient except several values of NAN and NCO in MOD SubComp and NES and NCO in UND SubComp.

- We found few correlations between metrics and the UND SubComp. From the set of metrics that present a correlation just one of them is correlated twice. The significance levels were between 0.002 and 0.038.
- DN, WNN and NNR are correlated with the MOD SubComp in the three experiments. The significance levels were between 0.000 and 0.041. DN has the stronger correlation in UAE, independently of the order of the tasks. However in this experiment, the correlation of NNR and WNN is stronger as time goes on.

4.3. Testing hypotheses 3 and 4

In order to test the 3rd and 4th hypotheses, we study the correlation using measures for ordinal data. 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 Tau coefficient to analyze the correlation of $H_{0,3}$ and $H_{0,4}$. The statistics for ordinal measures are not included here for the sake of brevity; nevertheless they can be obtained from the laboratory package. We conclude the following:

- UND SubComp and UND Time: In the UAE and UAChE there is a statistically significant relationship between the SubComp variable and the UND Time. However in the ULME we only found correlation in one trial (C_2).
- MOD SubComp and MOD Time: Regarding the MOD Time, the same results as the previous case are obtained.
- UND/MOD SubComp and UND/MOD Eff: there is a statistically significant relationship between UND SubComp and UND Eff and, between MOD SubComp and MOD Eff, in the case of UAE and UAChE experiments. In the ULME we found that MOD SubComp is correlated with the MOD Eff.

5. Conclusions

We launched a family of experiments in order to analyse the effect of coupling (measured by metrics) on the understandability and modifiability of OCL expressions. Through a thorough analysis of the collected empirical data we obtained the following findings:

- There seems to be a statistically significant correlation between many metrics, especially those related to tracing, and the Understandability Efficiency and Modifiability Efficiency. Moreover, coupling affects in different way on the understandability and modifiability of OCL expressions. Regarding the UND or MOD Eff: collection operations, their iterators and the number of classes seems to affect the UND Eff meanwhile the length of navigations and number of relationships influences MOD Eff. The MOD SubComp (the cognitive load when subjects rate MOD Tasks) seems to be affected by the length of navigations, the number of relationships and how the navigations are combined in collection operations.
- In the UA and UACh experiments the subjects’ subjective ratings (understandability or modifiability rating) are influenced by the time they used to understand or modify the OCL expressions, i.e. both times seems to affect their appreciation of the level of complexity of an OCL expression. In these two experiments the UND or MOD Eff are also correlated with UND and MOD SubComp, in stronger way. The reason the same results are not obtained in ULME could be the subjects’ heterogeneity, they were students of different universities (a threat to internal validity).

Learning effects could have also affected the internal validity of the results related to UND. In fact, in the UND tasks the time was lower as the time passed, and some metrics are only correlated in some observations.

Moreover, we had to go further in studying MOD Tasks because their correctness and efficiency was not good enough as in the case of UND Tasks, and for the subjects 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. In spite of these findings, we are aware that further validation is needed to obtain stronger results about the effect of coupling on understanding and modifying OCL expressions. For that reason our immediate goal is to study in depth these issues.

Acknowledgements

This research is part of the MESSENGER project (PCC-03-003-1) 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.

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] Basili, V. R., Shull, F., Lanubile, F.: Building knowledge through families of experiments. IEEE Trans. on Softw. Eng., Vol. 25 Nº 4, 1999, pp. 456-473.

[4] Boehm-Davis, D.A., Fox, J. E., Philips, B.: Techniques for Exploring Program Comprehension. Empirical Studies of Programmers, Sixth Workshop. Norwood, NJ: Ablex, 1996, pp. 3-37.

[5] 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.

[6] 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.

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

[8] Cant, S.N., Henderson-Sellers, B., Jeffery, D.R.: Application of Cognitive Complexity Metrics to Object Oriented Programs. Journal of Object Oriented Programming, Vol. 7 Nº 4, 1994, pp. 52-63.

[9] Ciolkowski, M., Shull, F. and Biffl, S. A Family of Experiments to Investigate the Influence of Context on The Effect of Inspection Techniques. Proceedings of the 6th Int. Conference on Empirical Assessment in Softw. Eng. (EASE), 2002, pp. 48-60.

[10] 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.

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

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

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

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

[15] 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.

[16] Miller, J.: "Applying Meta-Analytical Procedures to Software Engineering Experiments", Journal of Systems and Software, 54, 2000, pp. 29-39.

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

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

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

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

[21] Reynoso, L., Genero, M., Piattini, M.: Measuring OCL Expressions: An approach based on Cognitive Techniques. Imperial College Press, UK. 2005.

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

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

[24] SPSS, 2002 SPSS 11.5. "Syntax Reference Guide". Chicago. SPSS Inc. 2002.

[25] 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

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

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