INVESTIGATING THE ROLE OF UML IN THE SOFTWARE MODELING AND MAINTENANCE

A Preliminary Industrial Survey

Giuseppe Scanniello
Dipartimento di Matematica e Informatica, University of Basilicata, Viale Dell'Ateneo, Macchia Romana, Italy

Carmine Gravino, Genny Tortora
Dipartimento di Matematica e Informatica, University of Salerno, Via Ponte Don Melillo Fisciano, (SA), Italy

Keywords: Software Development, Software Maintenance, UML, Survey.

Abstract: In the paper we present the results of an industrial survey conducted with the Italian software companies that employ a relevant part of the graduate students of the University of Basilicata and of the University of Salerno. The survey mainly investigates the state of the practice regarding the use of UML (Unified Modeling Language) in the software development and maintenance The results reveals that the majority of the companies use UML for modeling software systems (in the analysis and design phases) and for performing maintenance operations. Moreover, maintenance operations are mainly performed by low experienced practitioners.

1 INTRODUCTION

In the software engineering community, there is a growing interest in the use of empirical studies (Basili, 1993) (Kitchenham *et al.*, 2002) such as controlled experiments, qualitative studies, surveys, and archival analyses to investigate processes, methods, and tools for software development and maintenance. Among these kinds of studies, a survey is an investigation performed in retrospect (Wohlin *et al.*, 2000) to gather qualitative and quantitative data very often using questionnaires that are designed through taking a sample which is representative of the population to be studied. The survey results can be analyzed to derive descriptive and conclusions that are only applicable to the population where the sample has been taken.

In order to investigate the state of the practice regarding the use of UML (Unified Modeling Language) in the software development and maintenance we have conducted an explorative survey (Wohlin *et al.*, 2000) among the Italian software companies that employ the greater part of the people with a Master's or a Bachelor's degree from either the University of Basilicata or University of Salerno. The companies that host students for external stages have been involved as well. All the invited companies (i.e., the target population of the study) represent our industrial contact network.

The research questions the presented study aims at answering for the selected population are:

[RQ1] What is the relevance of UML?
[RQ2] What is the used development process model?
[RQ3] Is UML used in the early phases of the software development?
[RQ4] Is UML used to support software engineers in the execution of maintenance operations?
[RQ5] Which software models (i.e., analysis and/or design) are provided to the software engineers to perform maintenance operations?
[RQ6] Which is the more common kind of software maintenance operation?
[RQ7] Which is the mean effort required to perform maintenance operations?

In order to answer the above research questions the following subsequent steps have been accomplished:

Scanniello G., Gravino C. and Tortora G. (2010).
INVESTIGATING THE ROLE OF UML IN THE SOFTWARE MODELING AND MAINTENANCE - A Preliminary Industrial Survey.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Information Systems Analysis and Specification*, pages 141-148
DOI: 10.5220/0002870601410148

1. design of a questionnaire to investigate the research questions;
2. conduction of the survey, leveraging the relative industrial contact network;
3. analysis of the results.

The design and the results of the survey are present and discuss in this paper. In particular, the remainder of the paper is organized as follows. Section 2 provides the notions and concepts we used in the paper and presents the design of the survey as well. Section 3 describes the results achieved, while general findings and threats to validity are reported in Section 4. Final remarks and future work conclude the paper.

2 DEFINITION AND DESIGN

The survey goals can be summarized as follows:

- **Primary Goal:** comprehending the relevance of UML for our industrial contact network (employing the greater part of the students from the Universities of the authors).
- **Secondary Goal:** identifying the core business activity of the interviewed industries between systems design and development and software maintenance.

2.1 Conceptual Model

The conceptual model clarifies the meaning of some terms and describes all the entities of interest for the conducted survey.

- **Project:** a completed software development project.
- **Development Process Model:** a structure imposed on the development of a software product (Pressman, 2005).
- **Software Artefact:** a tangible product created during the software development process. It can be created in the early and final stages of software development (Pressman, 2005).
- **Requirements Analysis:** a subdiscipline of systems engineering and software engineering that is concerned with determining the goals, functions, and constraints of hardware and software systems (Ciolkowski *et al.*, 2003, Laplante, 2009).
- **System Design:** a process to define the architecture, components, modules, interfaces, and data for a software system to satisfy specified requirements.
- **Software Maintenance:** the modification of a software product after delivery to correct faults, to improve performance or other attributes, or to adapt the product to a modified environment (ISO/IEC, 2006) .
- **Ordinary Maintenance:** reactive modifications of a software product performed after delivery to keep a software product usable, to improve performance or maintainability, and to detect and correct latent faults.
- **Extraordinary Maintenance:** modifications of a software product performed after the delivery phase to correct discovered problems.
- **UML:** a formalism to specify, visualize, and document models of software systems, including their structure and design (UML, 2009).

We identified three areas of interest where we wanted to collect data:

- **Demographic Information:** we collected demographic information about the respondents and their organizations.
- **Relevance and Typology:** we collected information about the relevance (i.e., number of organizations that use UML with respect to the complete sample) and quantity (i.e., the number) of projects where UML is used to develop and maintain software products. Great care was taken to collect information about the UML notations mainly used in the software development and maintenance.

Demographic information includes the respondent's organization. In particular we collected: business domain, size of the organization and respondent's group/business unit, typical duration and kind of projects, average experience and skill of group members. In addition, we collected information about the respondents themselves, such as: age, gender, educational qualification, role in the organization (e.g., IT manager, Project manager).

To evaluate the company relevance the respondents were also asked to specify whether their organizations use UML for software development and maintenance. Accordingly, we divided the respondents' organizations in the following groups, which depend on whether the organizations have:

- **(Group 1)** used UML for modelling software systems in the software development and for performing maintenance operations;
- **(Group 2)** used UML only in the early phases of the software development (i.e., analysis and design);
- **(Group 3)** used UML for performing maintenance operations;
- **(Group 4)** used UML neither in software development nor maintenance.

For the first and second group we asked the typical size of a developed software product and the adopted development process model. For the second and the third group, questions were also asked on the employees involved in the software development and maintenance, respectively. These questions were also asked to the respondents of group 1. For the organizations of the fourth group the questionnaire finished after collecting the demographic information.

2.2 Identification of the Target Population of the Sample

The target population consisted of decision makers in software developing organizations. Indeed, we considered ICT organizations that develop, sell, and maintain software as a main part of their business (e.g., software house) or develop software as an integral part of their products or services (e.g., commerce in the healthcare domain).

The selection of the organizations (sampling) has been conducted using the network contacts of the groups of the researchers who conducted the survey. Within this network there were the companies that host students for external stages and employ the majority of the persons with the Master or the Bachelor degree in Computer Science from the University of Salerno and the University of Basilicata.

2.3 Questionnaire Design and Data Collection

We have developed the questionnaire following the standard schema proposed by Ciolkowski *et al.* (2003). Figure 1 shows the design of the defined questionnaire, which according to the conceptual model consists of four different paths. The questionnaire was divided in 3 sections and was structured such that the total number of questions depended on whether the respondent's organization uses UML to develop and maintain software systems. The first part (Section 1), common to all paths, was used to get information on both the respondent and his/her company (i.e., demographic information). The questions of Section 1 were 12, while Section 2 included 8 questions. Finally, Section 3 included 9 questions. The questionnaire mainly contained closed questions. Only a few questions required filling in numbers or text.

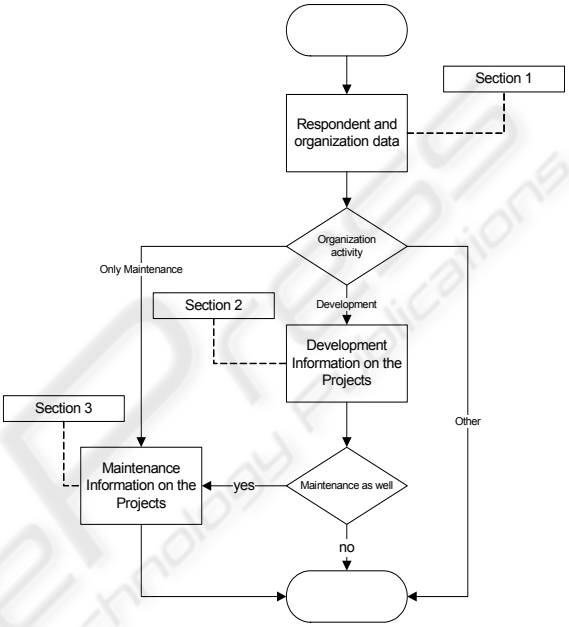

Figure 1: Designed questionnaire.

The questionnaire was introduced with a brief motivation sketching the general problem of the investigation. The importance of this study and our objectives were inserted in an accompanying letter attached to the questionnaire and sent to the respondents. Furthermore, we clarified that all the information was considered confidential and that the data were used only for research purposes and revealed only in aggregated form. Great care was also taken to ensure ethical requirements and privacy rules imposed by the Italian regulations (these were inserted at the end of the questionnaire).

We collected the questionnaire via e-mail. The main drawback of collecting the data in such a way is that information concerning the time to fill in the questionnaire is not available. However, this information may consider useless for the study presented here.

2.4 Survey Preparation, Execution and Analysis Procedure

The survey has been performed according to the

following four activities:

1. **Preparation and Design of the Questionnaire.** We used similar questionnaires to identify a set of questions.
2. **Invitation to Participate.** The organizations of the industrial contact network were invited by phone (to ask for their availability) and then an e-mail was sent to provide detailed instructions on the state of the practice survey and its aim. The questionnaire and the information related to it were attached to this e-mail.
3. **Collecting.** One of the authors collected the filled in questionnaire that the respondents returned by e-mail.
4. **Analysing.** The answers of the collected questionnaire were analyzed with respect to the goals of the industrial survey.

We collected 22 questionnaire correctly filled in. Note that the invited company were 53, thus obtaining 42% as response rate.

3 ANALYSIS

Due to the nature of our study the analysis is mainly based on basic descriptive statistics. In the following we summarize the obtained results according to the design of the developed survey questionnaire (see Section 2.3).

3.1 Respondents' Background and Companies' Characteristics

The respondents' age ranges from 28 to 56 years old with an average of 41 years. The female are only 5 and the majority of respondents work for a company located in the South of Italy (15). The companies are geographically distributed as follows: 16 are from the south of Italy, 4 are from the central, and 2 are from the north. Note that southern Italy is better represented, also because the universities that conducted the survey are located in southern Italy. On the other hand, further information on the respondents and their companies are summarized in Table 1.

Regarding the role of the respondents, 15 of them stated that they are project managers, while the others declared to be developers (2) or architects (4). Only one respondent involved in the study is a company administrator. Concerning the education, 3 respondents declared to have a Bachelor degree, while 18 respondents have a Master degree, and 1 only one respondent has a first level Master's degree. None has a PhD. As expected, the majority of the respondents (17) has a specific IT degree. 14 companies are independent, while 8 are controlled by a bigger and more powerful company. Among those companies, 13 are S.p.A, (i.e., companies quoted on the Italian stock exchange), while 9 are private companies. With respect to the size of the companies and according to the recommendation 2003/361/EC (micro companies have less than 10 employees; small companies have from 10 to 50 employees; medium companies have from 50 to 250 employees; large companies have more than 250 employees), 13 of the survey responses come from small and medium-sized companies and 8 come from large ones. Only one is a micro company. Regarding the business units, we have 4 micro, 8 small, and 9 medium sized units. Only 1 can be considered a large business unit.

Respondents also provided information about the experience and the education of the employees. The analysis reveals that on average 11% of the employees have less than one year of experience, 52% of the employees have less than five years of experience, and 37% of the employees have more than five years of experience.

Regarding the education, on average 29% of the employees have a Bachelor degree, 44% of the employees have a Master degree, and 27% of the employees have a High School diploma.

Information on the employees hired during the last year was also specified by the respondents. In particular, on average the percentage of employees with an ICT Bachelor degree hired every year is 40%, while 45 is the percentage of employees with an ICT Master degree. 15% is the percentage of employees without an ICT degree on average hired. Moreover, on average the 68% of the employees hired during last year had no previous experience in the ICT field. The remaining (32%) have on average 3 years of experience in the ICT field.

Regarding the industrial domains, about 10 work in the area of software consultancy. The same number of companies works as software house/vendors, while only 2 companies declared to be Web agency.

Concerning the main activities of the involved companies, 20 focus on the development and maintenance of software systems. A company only focuses on software development.

Table 1: Section 1 answers (Respondent and Organization Data).

QUESTION	ANSWER				
Respondent information					
[1.1.1] Role in the business unit	IT manager	Project manager	Software architect	Software developer	Other (role)
	0	15	4	2	1 (company administrator)
[1.1.2] Education level	Bachelor degree	Master degree	First level Master's degree	Second level Master's degree	PhD/Doctorate
	3	18	1	0	0
[1.1.3] Qualification in ICT field	Yes	No			
	17	5			
Company information					
[1.2.1] Type of Company	Self-supporting	Subsidiary			
	14	8			
[1.2.2] Type of Company	Stock company	Company controlled by a private entity		Governmental organization	
	13	9		0	
[1.2.3] Number of employee	< 10	[10-50]	[50-250]	> 250	
	1	3	10	8	
[1.2.3] Number of employee of the business unit	< 10	[10-50]	[50-250]	> 250	
	4	8	9	1	
[1.2.4] Employee experience	[0 – 1] years	[1 – 5] years	> 5 years		
	11%	52%	37%		
[1.2.5] Employee Degree	Bachelor degree	Master degree	High Scholl Diploma	Other	
	29%	44%	27%	0%	
[1.2.6] Education level of employee hired in the last year	ICT Bachelor degree		ICT Master degree	High Scholl Diploma	Other
	40%		45%	15%	0%
[1.2.7] Experience of employee hired in the last year	No experience in ICT field		Experience in ICT field (average number of year)		
	68%		32% (3)		
Company activity information					
[1.3.1] Company is	Software house/ Software vendor		Software/IT consultancy		Other
	10		10		2
[1.3.2] The main activity of the company is	Software development and maintenance		Software development	Software maintenance	Other
	20		1	0	1

3.2 Development Activity

The analysis of the answers of Section 2 (see Table 2) reveals that the greater part of the companies (14) usually develop software systems whose sizes range from 10,000 to 100,000 LOCs. Software systems with sizes ranging from 100,000 to 500,000 LOCs are developed by 5 companies. On the other hand, the size of the software systems developed by the 2 remaining companies is less than 10,000 LOCs and greater than 500,000 LOCs, respectively.

Regarding the development process model, 4 respondents indicated that the waterfall model is typically used, while 4 respondents specified that agile methodologies are employed. The remaining 13 respondents selected the Unified Process as widely used development process model used in their companies.

On average 18% of the employees has less than one year of experience, 51% has less than five years of experience. 32% of the new employees have more than five years of experience. Regarding the education, on average 42% of the employees has a Bachelor degree, 42% has a Master degree, and 16% has a High School diploma.

All the respondents indicated that UML is typically used during the development phase. Indeed, on average UML has been employed in the 74% of the completed software projects. In the remaining 26% of the projects UML was not used.

Concerning the employed UML diagrams, almost all companies typically employ Use Case Diagrams (19), Class Diagrams (20), and State Diagrams (17) to model software systems. The Component Diagrams and the State Diagrams are instead used by 14 and 13 companies, respectively. On the other hand, 8 companies use the Activity Diagrams.

The respondents were also asked to specify the UML diagrams adopted in the analysis and the design phases. The analysis of the answers has revealed that the diagrams more used in the analysis phase are the Use Case diagrams (13), the Class Diagrams (10), and the State Diagrams (13). However, the Sequence Diagrams (5) and the Activity Diagrams (8) are used as well. As expected

Table 2: Section 2 answers (Development Information on the Projects).

QUESTION	ANSWER						
[2.1] Size of developed software system	<10,000 LOC	[10,000 - 100,000] LOC		[100,000 - 500,000] LOC		> 500,000 LOC	
	1	14		5		1	
[2.2] Development process model	Waterfall	Unified Process		Agile Methodologies		Other	
	4	13		4		0	
[2.3] Employee Degree	Bachelor degree	Master degree		High Scholl Diploma		Other	
	42%	42%		16%		0%	
[2.4] Employee experience	< 1 year	[1 – 5] years		> 5 years			
	18%	51%		32%			
[2.5] Use of UML	Yes			Percentage of software systems where UML has been used			
	20			74%			
[2.5.1] UML diagram usually employed	Use Case diagram	Class Diagram	Component Diagram	Sequence Diagram	Activity Diagram	Deployment Diagram	State Diagram
	19	20	14	13	8	0	17
[2.5.2] UML diagram used during analysis phase	Use Case diagram	Class Diagram	Sequence Diagram	Activity Diagram		State Diagram	
	13	10	5	8		13	
[2.5.3] UML diagram used during design phase	Class Diagram	Component Diagram		Sequence Diagram		Deployment Diagram	
	19	14		10		0	

for the design phase the widely employed notations are the Class Diagrams (19), the Component Diagrams (14), and the Sequence Diagrams (10).

3.3 Maintenance Activity

Table 3 summarizes the information regarding the maintenance operations performed by the respondents' organizations (i.e., the answers of the questions of Section 3). In particular, the respondents specified that 64% of systems on which their companies perform maintenance operations are Objected-Oriented applications. Furthermore, the analysis of the answers revealed that 13 companies usually perform maintenance tasks on software systems developed by other companies. 11 respondents indicated that the typical maintenance operations performed by their companies are ordinary maintenance, while the remaining 9 companies usually carry out also extraordinary maintenance operations.

Respondents also provided information about the experience and the education of the employees performing maintenance tasks. Regarding the education, on average 43% of the employees has a Bachelor degree, 26% has a Master degree, and 31% has a High School diploma. The analysis revealed that on average 17% of the employees has less than one year of experience, 55% has more than 1 year and less than five years of experience, and 28% has more than five years of experience.

Moreover, 14 respondents declared that a typical ordinary maintenance operation requires less than 5 hours, while 5 respondents specified that less than 10 hours are required to perform this kind of maintenance operation. On the other hand, the majority of respondents (17) answered that a typical extraordinary maintenance operation requires from 10 to 50 hours. The remaining companies need less than 10 hour to perform a typical extraordinary maintenance operation.

Regarding the documentation exploited to perform extraordinary maintenance tasks, the majority of the respondents (14) specified that both analysis and design documents are usually used. It is worth to noting that 3 respondents declared that only source code is usually exploited to perform ordinary maintenance operations. The remaining respondents answered that only analysis documentation is used by the companies to perform ordinary maintenance. The results for extraordinary maintenance operations are nearly the same.

4 RESULTS

In the following subsections we summarize the main findings of the study and highlight the threats that may affect the obtained results.

4.1 General Findings

The analysis presented in the previous section has revealed that the majority of the companies (20) use UML in the analysis and design phases. In particular, the respondents declared that UML was employed in 74% of the completed projects in order to model objected-oriented software systems. Moreover, UML is also employed to carry out maintenance tasks since 75% of the respondents revealed that analysis and design models, produced during the software development process, are widely

Table 3: Section 3 answers (Maintenance Information on the Projects).

QUESTION	ANSWER			
[3.1] Maintenance operations performed on	O-O System		Other	
	64%		36%	
[3.2] Maintenance operations performed on system developed by other companies	Yes		No	
	13		7	
[3.3] Typical maintenance operations	Ordinary		Extraordinary	
	11		9	
[3.4] Employee Degree	Bachelor degree	Master degree	High Scholl Diploma	Other
	43%	26%	31%	0%
[3.5] Employee experience	[0 – 1] years	[1 – 5] years	> 5 years	
	17%	55%	28%	
[3.6] Typical ordinary maintenance operations takes	< 1 person/hours	[1 – 5] person/hours	[5 – 10] person/hours	> 10 person/hours
	0	14	5	0
[3.7] Typical extraordinary maintenance operations takes	< 10 person/hours	[10 – 50] person/hours	[50 – 100] person/hours	> 100 person/hours
	3	17	0	0
[3.8] Documentation used for ordinary maintenance	Analysis and design models	Analysis models	Design models	Nothing (only source code)
	14	3	0	3
[3.9] Documentation used for extraordinary maintenance	Analysis and design models	Analysis models	Design models	Nothing (only source code)
	15	4	0	1

used by the company maintainers. Maintenance operations are mainly performed by low experienced maintainers. In fact, the respondents revealed that 72% has less than 5 years of professional experience. Furthermore, on average 43% has a Bachelor degree in Computer Science. Employees with a Master degree in Computer Science are less frequently used to maintain existing software systems (on average 26% of maintainers).

Furthermore, the average effort to perform maintenance operations ranges from 1 to 5 person hours for an ordinary maintenance operation for 70% of the companies. On average, 85% of the companies spend from 10 to 50 person hours in case of extraordinary maintenance operations.

4.2 Threats to Validity

In the following the threats to validity (i.e., internal, external, and construct) that could affect the conducted study are presented and discussed in increasing priority order.

Internal validity threats regard external factors that may affect the observed results. In industrial surveys, it is usually impossible to know whether the respondents answer truthfully, or whether other effects may bias the results. Also, the respondents' motivations could affect the answers and then the survey results. Another factor that may have influenced this validity is the number of invited organizations that did not answer the questionnaire (less than 50%). Another negative factor could be the difficulty to comprehend the questions (e.g., ambiguous, not clear, not well formulated). However, we designed the questionnaire to minimize this problem. The reader may also object to the fact that the organizations within our industrial network may influence the internal validity as well. However, one of the goals of this survey concerned the study of the productive reality where the greater part of the students graduated at the University of Basilicata and at the University of Salerno are employed.

External validity threats concerns the generalization of the results. These threats are always present in case of industrial survey. For our survey, although the companies belonged to a variety of domains and covered small, medium and large company from north, centre, and south of Italy, we cannot be sure that our sample is representative of ICT Italian industry, in general.

The metrics used in the study as well as the designed questionnaire may threaten the construct validity. In our case, the questionnaire was designed using standard ways and scales (Ciolkowski *et al.*, 2003). Furthermore, the questions were formulated to minimize possible ambiguities.

Note that due to the nature of the conducted study, it is only through replications that we can gain more confidence in the survey results. Accordingly, we have planned to extend our investigation involving industries that do not belong to our industrial contact network.

5 FINAL REMARKS AND FUTURE WORK

This study revealed a set of interesting outcomes that can be summarized as follows. The greater part of the interviewed companies uses UML for modelling software systems (in the analysis and design phases) and for performing maintenance operations. Maintenance operations are mainly performed by low experienced practitioners. In fact, the companies generally use professionals, who have a Bachelor degree in Computer Science and have less than 5 years of experience. Professionals with a Master degree in Computer Science and with less than five years of experience are also used. However, they are less frequently used for maintaining existing software systems with respect to persons with a Bachelor degree.

Another interesting point concerns the average effort to perform maintenance operations. It ranges from 1 to 5 person hours for an ordinary maintenance operation (e.g., corrective changes), while it is from 10 to 50 person hours in case of an extraordinary maintenance operation (e.g., perfective or adaptive changes).

The results of this study represent an interesting starting point for further empirical investigations. In fact, we have been designing and conducting controlled experiments to assess the effectiveness of UML in the requirements engineering process and in the software maintenance as well.

We are also going to encourage replications to increase the confidence in the achieved results. This will increase our body of knowledge on the role of UML in the software development and maintenance. In fact, at the best of our knowledge the presented study is the first investigation on the UML usage in the Italian software industry regarding the use of UML in the software maintenance. Differently, Dobin and Parsons (2006) conducted a survey on how the UML diagrams are used in the analysis phase. The authors in the survey did not consider both the develop process used within the interviewed companies and the role of UML in the execution of maintenance operations.

Concluding the survey results also positively answer an interesting question from the pedagogical point of view: “Are the students that took a Bachelor or a Master degree either at the University of Basilicata or at the University of Salerno ready to be employed in the industries of our contact network?”

ACKNOWLEDGEMENTS

We want to thank all the companies that participated to the survey.

REFERENCES

Basili V. R., “The Experimental Paradigm in Software Engineering”, *Proc. International Workshop, Experimental Software Eng. Issues: Critical Assessment and Future Directions*, LNCS 706, Springer-Verlag, 1993.

Ciolkowski, M. Laitenberger, O., Biffl, S.,. “Software reviews: The state of the practice”. In *IEEE Software*, vol. 20, no 6, 2003, pp. 46 51.

Dobing B., Parsons J., “How UML is used”; In *Communications of the ACM*, vol. 49, no 5, 2006, pp. 109-113.

Kitchenham B. A., Pfleeger S.L., Pickard L.M., Jones P.W., Hoaglin D. C., El Emam K., Rosenberg J., “Preliminary Guidelines for Empirical Research in Software Engineering”, *IEEE Transactions on Software Engineering*, vol. 28, no. 8, pp. 721–734.

UML: Introduction to OMG's Unified Modeling Language™ available, 2009 http://www.omg.org/gettingstarted/what_is_uml.htm.

ISO/IEC 14764:2006 available at http://www.iso.org/

Laplante, P., *Requirements Engineering for Software and Systems* (1st ed.). Redmond, 2009: CRC Press. ISBN 1-42006-467-3.

Oppenheim, N., *Questionnaire Design, Interviewing and Attitude Measurement*, Pinter Publishers, 1992

Pressman R. S., *Software Engineering: A Practitioner's Approach*, Sixth Edition, United States, McGraw Hill, 2005.

Wohlin C., Runeson P., Host M., Ohlsson M. C., Regnell B., Wesslen A., *Experimentation in Software Engineering* - An Introduction, Kluwer, 2000.