

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/66798> holds various files of this Leiden University dissertation.

Author: Fernandez, Saez A.M.

Title: Studying the benefits of using UML on software maintenance : an evidence-based approach

Issue Date: 2018-11-15

Studying the Benefits of Using UML on Software Maintenance: an Evidence-Based Approach

Ana M. Fernández Sáez

Studying the Benefits of Using UML on Software Maintenance: an Evidence-Based Approach

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. mr. CJJM Stolker
volgens besluit van het College voor Promoties
te verdedigen op donderdag 15 november 2018
klokke 10:00 uur

door

Ana M. Fernández Sáez
geboren te Ciudad Real - Spain
in 1986

PROMOTIECOMMISSIE

Promotors: Prof Dr. Joost Kok (University of Leiden, The Netherlands)
Prof. Dr. Michel R.V. Chaudron (Gothenburg University, Sweden)
Prof. Dr. Marcela Genero (University of Castilla-La Mancha, Spain)

Overige Leiden: Prof. Dr. Ir. Fons Verbeek (University of Leiden, The Netherlands)
Prof. Silvia Abrahão (Universitat Politècnica de València, Spain)
Prof. Dr. Guilherme Horta Travassos (Federal University of Rio de Janeiro, Brazil)

Images used in the introduction of each chapter:

Chapter1: <https://goo.gl/aP2Sxw>
Chapter2: <https://goo.gl/YXsPGg>
Chapter3: <https://goo.gl/gJZpYK>
Chapter4: <https://goo.gl/KmYrTK>
Chapter5: <https://goo.gl/p6Sy5H>
Chapter6: <https://goo.gl/qFNvEU>
Chapter7: <https://goo.gl/uRq4Vv>

CONTENTS

CHAPTER 1. INTRODUCTION.....	1
1.1.INTRODUCTION.....	2
1.2.RESEARCH METHODOLOGY	9
1.3.CONTEXT OF THE THESIS.....	17
1.4.CONTRIBUTIONS OF THIS THESIS	20
1.5.OUTLINE OF THE STRUCTURE OF THIS THESIS.....	20
CHAPTER 2. STATE-OF-THE-ART.....	25
ABSTRACT	26
2.1.INTRODUCTION.....	26
2.2.RELATED WORK	29
2.3.PLANNING	30
2.4.CONDUCTING THE REVIEW	35
2.5.REPORTING RESULTS AND DATA SYNTHESIS	40
2.6.DISCUSSION	59
2.7.THREATS TO VALIDITY	62
2.8.CONCLUSIONS.....	63
CHAPTER 3. DOES THE LEVEL OF DETAIL (LOD) OF UML DIAGRAMS INFLUENCE SOFTWARE MAINTENANCE?	67
ABSTRACT	68
3.1.INTRODUCTION.....	68
3.2.RELATED WORK	70
3.3.THE FAMILY OF EXPERIMENTS	76
3.4.RESULTS	91
3.5.SUMMARY AND DISCUSSION OF THE DATA ANALYSIS	105
3.6.IMPLICATIONS OF THE STUDY	109
3.7.THREATS TO VALIDITY	111
3.8.CONCLUSIONS AND FUTURE WORK	113
CHAPTER 4. ARE FORWARD DESIGNED OR REVERSE- ENGINEERED UML DIAGRAMS MORE HELPFUL FOR CODE MAINTENANCE?	117
ABSTRACT	118
4.1.INTRODUCTION.....	118
4.2.RELATED WORK	121
4.3.DESCRPTION OF EXPERIMENT	128
4.4.RESULTS	139
4.5.IMPLICATIONS OF THE STUDY	151
4.6.THREATS TO VALIDITY	152
4.7.CONCLUSIONS AND FUTURE WORK	154

CHAPTER 5. WHAT IS THE PERCEPTION OF INDUSTRY PROFESSIONALS AS REGARDS THE VALUE OF USING UML IN SOFTWARE MAINTENANCE?	157
ABSTRACT	158
5.1.INTRODUCTION.....	158
5.2.SURVEY DESCRIPTION.....	159
5.3.RESULTS	161
5.4.THREATS TO VALIDITY	175
5.5.CONCLUSIONS AND FUTURE WORK	175
CHAPTER 6. WHAT IS THE IMPACT OF USING UML IN SOFTWARE MAINTENANCE IN TERMS OF ITS USE IN A SOFTWARE PROJECT?	179
ABSTRACT	180
6.1.INTRODUCTION.....	180
6.2.RELATED WORK	182
6.3.RESEARCH METHOD.....	185
6.4.RESULTS	196
6.5.RECOMMENDATIONS	234
6.6.SUMMARY OF RESULTS BY RESEARCH QUESTION	237
6.7.THREATS TO VALIDITY	242
6.8.CONCLUSIONS.....	243
6.9.FUTURE WORK.....	244
CHAPTER 7. CONCLUSIONS.....	247
7.1.ACHIEVEMENT OF GOALS AND SUMMARY OF FINDINGS.....	248
7.2.CONTRIBUTIONS OF THE PHD THESIS	254
7.3.FUTURE RESEARCH LINES	258
7.4.PUBLICATIONS	259
BIBLIOGRAPHY.....	263
APPENDIX A. LIST OF PRIMARY STUDIES.....	275
APPENDIX B. DEFINITIONS OF MEASURES	279
APPENDIX C. THE SEARCH STRINGS	283
APPENDIX D. EXAMPLES OF LOW AND HIGH LOD DIAGRAMS	287
APPENDIX E. INTERVIEW QUESTIONNAIRE.....	289
APPENDIX F. BACKGROUND INFORMATION RELATED TO EACH INTERVIEWEE	291
CURRICULUM VITAE	295
SUMMARY.....	297
SAMENVATTING.....	301
ACKNOWLEDGMENTS	305

INDEX OF FIGURES

Figure 1.1. An example of a UML class diagram.	3
Figure 1.2. An example of a UML use case diagram.	4
Figure 1.3. An example of a UML sequence diagram.	4
Figure 1.4. Research plan.	7
Figure 1.5. Phases of a Systematic Mapping Study.	10
Figure 1.6. Global view of the experiment process.	13
Figure 1.7. Research project timeline during PhD thesis development.	20
Figure 1.8. Chronology of publications and stays of this PhD Thesis.	21
Figure 2.1. Selection process.	39
Figure 2.2. Factors that influence the maintainability of a system.	56
Figure 2.3. Number of papers per year.	57
Figure 3.1. Chronology of the family of experiments	77
Figure 3.2. Example of Understandability task.	85
Figure 3.3. Example of Modifiability tasks for System B.	86
Figure 3.4. Meta-analysis of <i>UEffec</i>	97
Figure 3.5. Meta-analysis of <i>UEffic</i>	97
Figure 3.6. Meta-analysis of <i>MEffec</i>	98
Figure 3.7. Meta-analysis of <i>MEffic</i>	98
Figure 3.8. Interaction between LoD and System for <i>UEffec</i>	99
Figure 3.9. Interaction between LoD and System for <i>UEffic</i>	99
Figure 3.10. Interaction between LoD and System for <i>MEffec</i>	100
Figure 3.11. Interaction between LoD and System for <i>MEffic</i>	100
Figure 3.12. Subjects' perception of the experiment.	102
Figure 3.13. Subjects' difficulties when reading diagrams.	103
Figure 3.14. Subjects' difficulties when reading source code.	103
Figure 3.15. Subjects' opinion of LoD.	104
Figure 4.1. Chronology of the family of experiments.	129
Figure 4.2. Example of adaptive maintenance task.	134
Figure 4.3. Example of corrective maintenance task.	134
Figure 4.4. Answer sheet related to the element "Class", filled with an example.	135
Figure 4.5. Meta-analysis results for Maintainability Effectiveness (MEffec). ...	143
Figure 4.6. Meta-analysis results for Maintainability Efficiency (MEffic).	143

Figure 4.7. Interaction between Origin and Ability for Maintainability Effectiveness (MEffec).....	144
Figure 4.8. Interaction between Origin and Ability for Maintainability Efficiency (MEffic).....	145
Figure 4.9. Subjects' answers as regards adequacy of time provided.....	146
Figure 4.10. Subjects' answers as regards difficulty of task.....	146
Figure 4.11. Subjects' answers as regards adequacy of the LoD.....	146
Figure 4.12. Subjects' answers as regards to difficulties when reading diagrams.....	147
Figure 4.13. Subjects' answers as regards usefulness of class diagram used.....	147
Figure 4.14. Subjects' answers as regards usefulness of sequence diagrams used.....	148
Figure 5.1. Use of UML diagrams during software maintenance per role.....	163
Figure 5.2. Use of UML diagrams during software maintenance by educational level.....	163
Figure 5.3. Use of UML diagrams by software development methodology.....	168
Figure 5.4. Relationship between size of company and use of UML in software maintenance.....	170
Figure 5.5. Relationship between role of software in the company and use of UML in software maintenance.....	172
Figure 5.6. UML diagrams used during software maintenance.....	173
Figure 5.7. UML diagrams created during software development.....	174
Figure 6.1. Matrix structure of the ICT organization.....	191
Figure 6.2. Flow of documentation in the company and responsibility roles.....	192
Figure 6.3. The type of case study based on Yin's definition (Yin, 2002).....	193
Figure 6.4. Case study execution and analysis procedure.....	196
Figure 6.5. "Baseline" theory.....	198
Figure 6.6. Legend of the theory-diagrams.....	198
Figure 6.7. Responses regarding the main reasons for the use of UML.....	201
Figure 6.8. Theory summarizing the purposes of the use of UML.....	203
Figure 6.9. Summary of findings about cost-types.....	205
Figure 6.10. Influences of the SE process in the modelling approach.....	209
Figure 6.11. Summary of reasons for not maintaining documentation, and risks involved in not maintaining documentation.....	214
Figure 6.12. Summary of findings about Practice.....	222
Figure 6.13. Summary of findings about Style.....	224
Figure 6.14. Needs of modelling tools detected in the case study.....	230
Figure 6.15. Influence of outsource/offshored maintenance on the documentation approach.....	231

Figure 6.16. Influence of legacy systems on the modelling approach.....	232
Figure 7.1. Contextual factors that affect the adoption of UML diagrams.....	249
Figure 7.2. Main factors studied and relations among them.....	254
Figure 7.3. Overview of the impacts studied on this PhD Thesis.....	255

INDEX OF TABLES

Table 1.1. Projects that form the context of the thesis.	18
Table 2.1. Search string.	32
Table 2.2. Summary of search strategy.	32
Table 2.3. Summary of selection strategy.	33
Table 2.4. Data extraction form.	34
Table 2.5. Quality checklist.	36
Table 2.6. Systematic Mapping Study outline.	38
Table 2.7. Results per type of diagram.	41
Table 2.8. Variables and measures for class diagrams.	42
Table 2.9. Variables and measures for statechart diagrams.	44
Table 2.10. Variables and measures for sequence diagrams.	44
Table 2.11. Variables and measures for collaboration diagrams.	44
Table 2.12. Variables and measures for activity diagrams.	45
Table 2.13. Variables and measures for use case diagrams.	45
Table 2.14. Variables and measures for the UML diagrams, in general.	45
Table 2.15. Variables and measures for the system (source code and diagrams)	46
Table 2.16. Results per duration.	47
Table 2.17. Results per empirical method.	48
Table 2.18. Results per context.	49
Table 2.19. Results per type of subject.	49
Table 2.20. Results per object to maintain.	50
Table 2.21. Results per type of system.	50
Table 2.22. Treatments in the empirical studies.	51
Table 2.23. Quality of primary studies.	53
Table 2.24. Number of papers per type of publication.	58
Table 2.25. Diagrams obtained from RE.	58
Table 3.1. Summary of related works.	74
Table 3.2. Levels of detail in UML diagrams.	78
Table 3.3. Description of the systems received.	79
Table 3.4. Experimental design.	84
Table 3.5. Summary of modifiability tasks.	86
Table 3.6. Post experiment questions.	87
Table 3.7. Descriptive statistics for <i>UEffec</i>	91
Table 3.8. Descriptive statistics for <i>UEffic</i>	92

Table 3.9. Descriptive statistics for <i>MEffec</i>	92
Table 3.10. Descriptive statistics for <i>MEffic</i>	92
Table 3.11. Wilcoxon tests results for <i>UEffec</i>	93
Table 3.12. Wilcoxon tests results for <i>UEffic</i>	94
Table 3.13. Wilcoxon tests results for <i>MEffec</i>	95
Table 3.14. Wilcoxon tests results for <i>MEffic</i>	95
Table 3.15. U-Mann Whitney tests results for the influence of Order.....	100
Table 3.16. U-Mann Whitney tests for the influence of Ability.....	101
Table 3.17. Summary of results of the family of experiments.....	106
Table 4.1. Summary of the related work.....	126
Table 4.2. Metrics of the software system used in the family of experiments.....	130
Table 4.3. Experimental design.....	133
Table 4.4. Summary of maintenance tasks.....	135
Table 4.5. Post-Experiment Survey.....	136
Table 4.6. Descriptive statistics for <i>MEffec</i>	140
Table 4.7. Descriptive statistics for <i>MEffic</i>	140
Table 4.8. Statistical relation between Origin of diagram (RE/FD) and Maintainability Effectiveness (<i>MEffec</i>).....	141
Table 4.9. Statistical relation between Origin of diagram (RE/FD) and Maintainability Efficiency (<i>MEffec</i>).....	141
Table 4.10. Statistical relation between Ability and Maintainability Effectiveness (<i>MEffec</i>) and Efficiency (<i>MEffic</i>).....	144
Table 4.11. Usage of UML diagrams for solving tasks.....	149
Table 4.12. Summary of the results of the family of experiments.....	150
Table 5.1. Frequency of use of source code only.....	165
Table 5.2. Effort of consulting UML diagrams for UML users, and effort of consulting documentation for non UML group.....	167
Table 5.3. Relationship between geo-distribution of the maintenance teams and use of UML in software maintenance.....	169
Table 5.4. Relationship between geo-distribution of companies and use of UML in software maintenance.....	170
Table 5.5. Relationship between size of ICT department and use of UML in software maintenance.....	171
Table 5.6. Relationship between team size and use of UML in software maintenance.....	171
Table 5.7. Relationship between size of system maintained and use of UML in software maintenance.....	172

Table 6.1. Summary of related work.	186
Table 6.2. Cost factors related to the use of UML.	204
Table 6.3. Summary of cost of modelling or not modelling.....	207
Table 6.4. Presence of diagrams in the documentation.	216