



Universiteit
Leiden
The Netherlands

The effects of UML modeling on the quality of software

Nugroho, A.

Citation

Nugroho, A. (2010, October 21). *The effects of UML modeling on the quality of software*. Retrieved from <https://hdl.handle.net/1887/16070>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/16070>

Note: To cite this publication please use the final published version (if applicable).

The Effects of UML Modeling on the Quality of Software

Ariadi Nugroho

The Effects of UML Modeling on the Quality of Software

Proefschrift

ter verkrijging van de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Prof. Mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 21 oktober 2010
te klokke 13.45 uur

door

Ariadi Nugroho
geboren te Blitar – Indonesia, 1979

Promotiecommissie

Promotor : Prof. dr. Joost N. Kok
Co-Promotor : Dr. Michel R.V. Chaudron
Overige Leden : Prof. dr. Thomas H.W. Bäck
Prof. dr. Frank S. de Boer
Prof. Jon Whittle (University of Lancaster)
Dr. Marcela G. Bocco (University of Castilla-La Mancha)



The work in this thesis has been carried out under the FINESSE project (des7015) supported by the STW (Stichting Technische Wetenschappen), the Netherlands.



The work in this thesis has been carried out under the auspices of the research school IPA (Institute for Programming research and Algorithmics).

ISBN 978-90-9025677-1

This thesis is dedicated to my family...

Publications

During my Phd research, I have co-authored several publications. I list these publications below in chronological order.

1. **A Survey of the Practice of Design – Code Correspondence amongst Professional Software Engineers.**
Ariadi Nugroho and Michel R.V. Chaudron.
Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement (ESEM), September 2007.
2. **On the Relation between Class-Count and Modeling Effort.**
Ariadi Nugroho and Christian F.J. Lange.
Proceedings of Model Size Metrics Workshop (Co-located with MODELS Conference), October 2007.
Received best paper award.
3. **A Survey into the Rigor of UML Use and its Perceived Impact on Quality and Productivity.**
Ariadi Nugroho and Michel R.V. Chaudron.
Proceedings of the Second International Symposium on Empirical Software Engineering and Measurement (ESEM), October 2008.
4. **Managing the Quality of UML Models in Practice.**
Ariadi Nugroho and Michel R.V. Chaudron.
In Model-Driven Software Development: Integrating Quality Assurance. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing; 2008.
5. **Empirical Analysis of the Relation between Level of Detail in UML Models and Defect Density.**
Ariadi Nugroho, Bas Flaton, and Michel R.V. Chaudron.
Proceedings of the 11th International Conference on Model Driven Engineering Languages and Systems (MODELS), September 2008.
Received best papers awards from ACM SIGSOFT and Springer.
6. **Evaluating the Impact of UML Modeling on Software Quality: An Industrial Case Study.**

Ariadi Nugroho and Michel R.V. Chaudron.

Proceedings of the 12th International Conference on Model Driven Engineering Languages and Systems (MODELS), October 2009.

7. Level of Detail in UML Models and its Impact on Model Comprehension: A controlled Experiment.

Ariadi Nugroho

Information and Software Technology Journal, 2009.

8. Assessing UML Design Metrics for Predicting Fault-prone Classes in a Java System.

Ariadi Nugroho, Michel R.V. Chaudron, and Erik Arisholm.

Proceedings of the 7th International Working Conference on Mining Software Repositories (MSR), May 2010.

Contents

1	Introduction	1
1.1	Quality in Modeling	1
1.2	Problem Statement	2
1.3	Objective of the Study	2
1.4	Research Methodology	3
1.5	Contributions and Outline	4
2	Background	7
2.1	The Unified Modeling Language	7
2.1.1	UML Diagram Types	8
2.1.2	The Widespread Use of UML as a Modeling Language	9
2.2	Style and Rigor in UML Modeling	11
2.3	The State of the Art in Quality Assurance of UML Models	13
2.3.1	Quality Models of UML Models	13
2.3.2	Methods and Techniques for Assessing UML Model Quality	17
2.3.3	Predictions of Software Quality based on Design Metrics	20
2.4	Conclusion	23
3	The Use of UML in Practice	25
3.1	Introduction	25
3.2	Related Work	26
3.3	Survey Design	27
3.4	Respondent Background	27
3.5	Results	29

3.5.1	UML Model Completeness	29
3.5.2	Level of Detail Used in UML Models	31
3.5.3	Model – Code Correspondence	33
3.5.4	The Use of UML and Productivity	39
3.5.5	The Use of UML and Software Quality	41
3.6	Discussion	42
3.6.1	Targeting Completeness & Level of Detail	42
3.6.2	Toward a Better Correspondence	43
3.6.3	Improving Productivity and Quality with UML Modeling	43
3.6.4	Threats to Validity	44
3.7	Conclusions and Future Work	45
4	The Effects of Using UML on Defect Density and Fixing Effort	47
4.1	Introduction	47
4.2	Related Work	48
4.3	Design of Study	49
4.3.1	Objective	50
4.3.2	Hypothesis Formulation	50
4.3.3	Measured Variables	51
4.4	Case Study	53
4.4.1	Project Context	54
4.4.2	Data Collection and Preprocessing	55
4.4.3	Analysis Methods	57
4.5	Testing Hypothesis 1: The Effect of UML Modeling on Defect Density	57
4.5.1	Descriptive statistics	57
4.5.2	Controlling for the Confounding Factors	58
4.5.3	Assessing the Effect	59
4.6	Testing Hypothesis 2: The Effect of Modeling on Defect-fixing Effort	61
4.6.1	Descriptive Statistics	61
4.6.2	Assessing the Effect	65
4.7	Interpretation of Results	69
4.7.1	Threats to Validity	71
4.8	Conclusions and Future Work	72

5	The Impact of Level of Detail in UML Models on Comprehension	73
5.1	Introduction	73
5.2	Related Work	74
5.3	Level of Detail	76
5.4	Experiment Planning	77
5.4.1	Subjects in the Experiment	77
5.4.2	Variables in the Experiment	77
5.4.3	Hypotheses Formulation	77
5.4.4	Experimental Design	78
5.4.5	Instrumentation	80
5.5	Experiment Operation	86
5.6	Data Analysis and Interpretation	87
5.6.1	Analysis Procedure	87
5.6.2	Experiment Results	88
5.6.3	In-depth Analyses	94
5.7	Discussion	98
5.7.1	Reflection on the Results	98
5.7.2	Implications for Research and Practice	99
5.7.3	Threats to Validity	100
5.8	Conclusion and Future Work	101
6	LoD in UML Models and its Relation with Defect Density	103
6.1	Introduction	103
6.2	Related Work	104
6.3	Level of Detail in UML Models	107
6.3.1	Definition	107
6.3.2	Level of Detail Measures	107
6.4	Design of Study	112
6.4.1	Research Questions	112
6.4.2	Measured Variables	112
6.4.3	Other Factors to Be Controlled	113
6.4.4	Analysis Method	113
6.5	Case Study	113

6.5.1	Data Collection and Processing	113
6.6	Data Analysis and Results	115
6.6.1	Analysis Procedure	115
6.6.2	Descriptive Statistics	116
6.6.3	Correlation Analyses between LoD and Defect Density	120
6.6.4	The Functional Form of LoD - Defect Density Relationship	129
6.6.5	The Relationship between LoD and Defect-fixing Effort	131
6.7	Discussion	134
6.7.1	Interpretation of Results	134
6.7.2	Some Remarks for Modeling Practices	135
6.7.3	Threats to Validity	137
6.8	Conclusion and Future Work	138
7	UML Metrics for Predicting Fault-prone Java Classes	139
7.1	Introduction	139
7.2	Related Work	140
7.3	UML Design Metrics	142
7.4	Design of Study	144
7.4.1	Measured Variables	144
7.4.2	Case Study and Data Collection	144
7.4.3	Analysis Method	145
7.5	Results	146
7.5.1	Univariate Analysis	146
7.5.2	Multivariate Analysis	147
7.5.3	Goodness of Fit of the Prediction Models	148
7.5.4	Assessing Cost-effectiveness	151
7.6	Discussion	153
7.7	Threats to Validity	155
7.8	Conclusion and Future Work	155
8	Conclusions	157
8.1	Summary of Findings	157

8.1.1	RQ1: How is the UML used in practice; and do software engineers perceive any benefits of using UML?	157
8.1.2	RQ2: What is the effect of using UML for modeling a software system on quality and productivity?	158
8.1.3	RQ3: What is the effect of the quality of UML models on model comprehension?	159
8.1.4	RQ4: Does the quality of UML models correlate with the quality of the resulting software?	159
8.1.5	RQ5: Can we predict fault-prone software modules or components based on the quality of the software model?	160
8.2	Contributions	161
8.3	Notes on Additional Case Studies	162
8.4	Recommendations	165
8.5	Concluding Remarks	166
8.6	Future Research	166
A	UML Survey Questionnaire	169
B	Model Comprehension Questionnaire	175
C	Descriptive Statistics of the NS-OFI Data Set	181
	Summary	193
	Samenvatting	195
	Curriculum Vitae	197
	Acknowledgements	199

List of Tables

2.1	Formal Released Versions of UML	8
2.2	Chidamber and Kemerers metrics suite for object-oriented design	18
2.3	Overview of Quality Prediction Studies	21
3.1	List of statements on levels of detail in UML models	31
3.2	Spearman correlation coefficient between model completeness and strictness in implementing modeling constructs	38
3.3	Spearman correlation coefficient between strictness in implementing modeling constructs and productivity across development phases	40
4.1	Measured Variables	52
4.2	Project Summary	54
4.3	Descriptive statistics of all Java classes	58
4.4	Descriptive statistics of the randomly sampled Java classes from the IPS project comparing NMC and MC	59
4.5	Mann-Whitney test - Ranks of the measured variables of the NMC and MC groups	60
4.6	Mann-Whitney test - The significance of differences in the measured variables between the NMC and MC groups	60
4.7	Results of assessing the impact of UMLforCLASS on DDENS using ANCOVA	61
4.8	Descriptive statistics of sampled defects from IPS comparing NMD and MD .	62
4.9	Mann-Whitney test - Ranks of the measured variables of the NMD and MD groups	66
4.10	Mann-Whitney test - The significance of differences in the measured variables between the NMD and MD groups	66
4.11	Results of assessing the effect of UMLforDEFECT on FixEffort using ANCOVA— accounting for the effects of the confounding factors	69

5.1	LoD treatments in the UML model	81
5.2	Descriptive Statistics of subjects' knowledge/experience, comprehension correctness, and comprehension efficiency across groups	88
5.3	Ranks of knowledge/experience score across groups	90
5.4	The results of the Mann-Whitney test showing the insignificance of the difference in knowledge/experience between the L-LoD and H-LoD groups	90
5.5	Group statistics for comprehension correctness and comprehension efficiency .	92
5.6	The results of the independent t-test showing the significant effects of LoD on comprehension correctness and comprehension efficiency	92
5.7	The results of the two-way ANOVA for CO - Ability (using normalized data). The effect of LoD on comprehension correctness remains significant after the effects of Ability and LoD*Ability are accounted for	94
5.8	The results of the two-way ANOVA for EF - Ability. The effect of LoD on comprehension efficiency is not significant after the effects of Ability and LoD*Ability are accounted for	94
6.1	Distribution of defects across faulty classes	119
6.2	Descriptive statistics of implementation classes modeled in class diagrams . .	120
6.3	Descriptive statistics of implementation classes modeled in sequence diagrams	120
6.4	Correlation between independent variables of class diagram LoD (Spearman's)	122
6.5	Correlation between independent variables of sequence diagram LoD (Spearman's)	123
6.6	Results of univariate regression for class diagram LoD measures	123
6.7	Results of multivariate regression for class diagram LoD measures	125
6.8	Results of multivariate regression analysis for sequence diagram LoD measures	125
6.9	Results of univariate regression for sequence diagram LoD measures	127
6.10	Results of univariate regression for class diagram LoD measures — NS-OFI data set	128
6.11	Results of univariate regression for sequence diagram LoD measures — NS-OFI data set	129
6.12	Results of multivariate regression for class diagram LoD measures — NS-OFI data set	130
6.13	Results of multivariate regression analysis for sequence diagram LoD measures — NS-OFI data set	130
6.14	Results of curve estimations between SD_{msg} and defect density	131
6.15	Correlation between averaged MsgLoD and FixEffort (Spearman's)	133

7.1	Results of univariate analysis	146
7.2	Results of multivariate analysis	147
7.3	The Confusion Matrix	148
7.4	Classification Accuracy of the Prediction Models (cross-validated using LOOCV)	148
7.5	The Goodness of Fit of the Multivariate Models (probability threshold = 0.5)	149
7.6	A Comparison of the goodness of fit	151
8.1	Project Summary	163
8.2	Results of multivariate analysis of sequence diagram LoD measures for the BHN project. The results show that SD_{msg} is not a significant predictor of defect density, after controlling for the effects of coupling and complexity. . .	164
C.1	Descriptive statistics of classes modeled in class diagrams – NS-OFI data set	181
C.2	Descriptive statistics of classes modeled in sequence diagrams – NS-OFI data set	181

List of Figures

2.1	The Taxonomy of UML Diagram Types	9
2.2	Use Case Diagram	10
2.3	Class Diagram	11
2.4	Sequence Diagram	12
2.5	Boehm's Quality Model	14
2.6	The ISO 9126	15
2.7	SQA and Software Quality Assessment in Software Development	15
2.8	Framework for quality of UML models	16
3.1	Developers' country of origin (percentage of respondents)	28
3.2	Developers' project involvement in the past 10 years	29
3.3	Developers' perception on model completeness in projects	30
3.4	The occurrence of imperfections in UML models	31
3.5	Respondents' agreement over approaches in using detail in models	32
3.6	The importance of correspondence	34
3.7	Developers' strictness in implementing different constructs	35
3.8	Methods used in maintaining correspondence	36
3.9	Factors driving deviation in an implementation	37
3.10	The use of UML and productivity	39
3.11	The use of UML and its impact on some software quality properties	41
4.1	Defect solving procedure	56
4.2	Box-plots of defect density in NMC and MC group	59
4.3	Box-plots of FixEffort in the NMD and MD groups	63

4.4	The distribution of defect types of the 86 sampled defects—also showing the frequency of modeled and not modeled defect per type	64
4.5	Box-plots of FixEffort, ModFile, TCBO, and TMCC in the NMD and MD groups (after data transformation)	67
4.6	Box-plots of TKSLOC, DPRIO, and TPERS in the NMD and MD groups (after data transformation)	68
5.1	Example of level of detail in a class diagram	76
5.2	A class diagram in the UML model with low LoD (the M-Low model)	83
5.3	A class diagram in the UML model with high LoD (the M-High model) . . .	84
5.4	Example of sequence diagrams created using different LoD treatments	85
5.5	A sample question of the model comprehension questionnaire	86
5.6	The profile of subjects' knowledge and experience	89
5.7	Box-plots of comprehension correctness between groups	91
5.8	Box-plots of comprehension efficiency between groups	92
5.9	Score of all questions in both groups	95
5.10	Subjects' perception on the UML model	97
6.1	Identical class diagrams modeled with different LoD	108
6.2	Overview of steps in collecting and processing data	115
6.3	Defect type distribution of the sampled findings	117
6.4	Histogram of defect distribution across the 122 faulty classes	119
6.5	Box-plots of defect density, complexity, and coupling of the 122 faulty classes	120
6.6	Box-plots of the 23 faulty classes modeled in class diagrams (after log transformation)	121
6.7	Box-plots of the 30 faulty classes modeled in sequence diagrams (after log transformation)	122
6.8	Scatterplot showing the correlation between SD_{msg} and defect density	128
6.9	Estimations of the form of functional relationship between SD_{msg} and defect density	132
6.10	Scatterplot showing the correlation between MsgLoD and FixEffort	133
6.11	The MetricView tool showing the LoD profile of classes in a class diagram. Classes with low LoD are marked with red borders. The color represents the value of an LoD metric. The stronger the color, the lower is the metric value.	136
7.1	Accuracy of the prediction models represented using ROC curves	150

7.2	Cost-effectiveness of the Prediction Models	152
8.1	Cause-effect diagram summarizing the main findings of the study	161
8.2	The relationships between SD_{msg} and defect density in the IPS and BHN projects	164

