

Contents

1	Introduction	1
1.1	Test-First Programming	1
1.1.1	Mechanisms Behind Test-First Programming that Motivate Research	2
1.2	Research Methodology	4
1.2.1	Empirical Software Engineering	4
1.2.2	Empirical Methods	5
1.3	Software Measurement	8
1.3.1	Measurement Levels	8
1.3.2	Software Product Quality	9
1.3.3	Software Development Productivity	12
1.4	Research Questions	13
1.5	Book Organization	13
1.6	Claimed Contributions	14
2	Related Work in Industrial and Academic Environments	15
2.1	Test-First Programming	15
2.2	Pair Programming	20
2.3	Summary	23
3	Research Goals, Conceptual Model and Variables Selection	25
3.1	Goals Definition	25
3.2	Conceptual Model	26
3.3	Variables Selection	28
3.3.1	Independent Variable (IV)	28
3.3.2	Dependent Variables (DVs) — From Goals to Dependent Variables	32
3.3.3	Confounding Variables	36
4	Experiments Planning, Execution and Analysis Procedure	39
4.1	Context Information	39
4.2	Hypotheses	41

- 4.3 Measurement Tools 42
 - 4.3.1 Aopmetrics 42
 - 4.3.2 ActivitySensor and SmartSensor Plugins 43
 - 4.3.3 Judy 43
- 4.4 Experiment ACCOUNTING 44
 - 4.4.1 Goals 44
 - 4.4.2 Subjects 44
 - 4.4.3 Experimental Materials 44
 - 4.4.4 Experimental Task 45
 - 4.4.5 Hypotheses and Variables 45
 - 4.4.6 Design of the Experiment 45
 - 4.4.7 Experiment Operation 46
- 4.5 Experiment SUBMISSION 46
 - 4.5.1 Goals 47
 - 4.5.2 Subjects 47
 - 4.5.3 Experimental Materials 47
 - 4.5.4 Experimental Task 48
 - 4.5.5 Hypotheses and Variables 48
 - 4.5.6 Design of the Experiment 48
 - 4.5.7 Experiment Operation 48
- 4.6 Experiment SMELLS&LIBRARY 50
 - 4.6.1 Goals 50
 - 4.6.2 Subjects 50
 - 4.6.3 Experimental Materials 51
 - 4.6.4 Experimental Tasks 51
 - 4.6.5 Hypotheses and Variables 51
 - 4.6.6 Design of the Experiment 52
 - 4.6.7 Experiment Operation 52
- 4.7 Analysis Procedure 53
 - 4.7.1 Descriptive Statistics 53
 - 4.7.2 Assumptions of Parametric Tests 53
 - 4.7.3 Carry-Over Effect 54
 - 4.7.4 Hypotheses Testing 55
 - 4.7.5 Effect Sizes 55
 - 4.7.6 Analysis of Covariance 56
 - 4.7.7 Process Conformance and Selective Analysis 57
 - 4.7.8 Combining Empirical Evidence 60

- 5 Effect on the Percentage of Acceptance Tests Passed 61**
 - 5.1 Analysis of Experiment ACCOUNTING 61
 - 5.1.1 Preliminary Analysis 61
 - 5.1.2 Selective Analysis 85
 - 5.2 Analysis of Experiment SUBMISSION 101
 - 5.2.1 Preliminary Analysis 101

- 5.2.2 Selective Analysis 110
- 5.3 Analysis of Experiment SMELLS&LIBRARY 116
 - 5.3.1 Preliminary Analysis 117
 - 5.3.2 Selective Analysis 121
- 5.4 Instead of Summary 125

- 6 Effect on the Number of Acceptance Tests Passed per Hour 127**
 - 6.1 Analysis of Experiment ACCOUNTING 127
 - 6.1.1 Descriptive Statistics 128
 - 6.1.2 Non-Parametric Analysis 128
 - 6.2 Analysis of Experiment SUBMISSION 129
 - 6.2.1 Descriptive Statistics 129
 - 6.2.2 Assumption Testing 131
 - 6.2.3 Non-Parametric Analysis 131
 - 6.3 Analysis of Experiment SMELLS&LIBRARY 136
 - 6.3.1 Descriptive Statistics 136
 - 6.3.2 Assumption Testing 138
 - 6.3.3 Non-Parametric Analysis 139
 - 6.4 Instead of Summary 140

- 7 Effect on Internal Quality Indicators 141**
 - 7.1 Confounding Effect of Class Size on the Validity of Object-Oriented Metrics 141
 - 7.2 Analysis of Experiment ACCOUNTING 142
 - 7.2.1 Descriptive Statistics 142
 - 7.2.2 Assumption Testing 145
 - 7.2.3 Mann–Whitney Tests 145
 - 7.3 Analysis of Experiment SUBMISSION 147
 - 7.3.1 Descriptive Statistics 147
 - 7.3.2 Assumption Testing 150
 - 7.3.3 Independent *t*-Test 150
 - 7.4 Analysis of Experiment SMELLS&LIBRARY 152
 - 7.4.1 Descriptive Statistics 152
 - 7.4.2 Assumption Testing 153
 - 7.4.3 Dependent *t*-Test 155
 - 7.5 Instead of Summary 158

- 8 Effects on Unit Tests – Preliminary Analysis 159**
 - 8.1 Analysis of Experiment SUBMISSION 160
 - 8.1.1 Descriptive Statistics 160
 - 8.1.2 Assumption Testing 162
 - 8.1.3 Mann–Whitney Test 163

- 9 Meta-Analysis** 165
 - 9.1 Introduction to Meta-Analysis 166
 - 9.1.1 Combining *p*-Values Across Experiments 166
 - 9.1.2 Combining Effect Sizes Across Experiments 167
 - 9.2 Preliminary Meta-Analysis 171
 - 9.2.1 Combining Effects on the Percentage of Acceptance Tests Passed (PATP) 171
 - 9.2.2 Combining Effects on the Number of Acceptance Tests Passed Per Development Hour (NATPPH) 175
 - 9.2.3 Combining Effects on Design Complexity 177
 - 9.3 Selective Meta-Analysis 184
 - 9.3.1 Combining Effects on the Percentage of Acceptance Tests Passed (PATP) 185
 - 9.3.2 Combining Effects on the Number of Acceptance Tests Passed Per Hour (NATPPH) 187
 - 9.3.3 Combining Effects on Design Complexity 188

- 10 Discussion, Conclusions and Future Work** 197
 - 10.1 Overview of Results 197
 - 10.2 Rules of Thumb for Industry Practitioners 200
 - 10.3 Explaining Plausible Mechanisms Behind the Results 202
 - 10.4 Contributions 205
 - 10.5 Threats to Validity 206
 - 10.5.1 Statistical Conclusion Validity 206
 - 10.5.2 Internal Validity 209
 - 10.5.3 Construct Validity 211
 - 10.5.4 External Validity 213
 - 10.5.5 Threats to Validity of Meta-Analysis 216
 - 10.6 Conclusions and Future Work 217

- Appendix** 219

- Glossary** 223

- References** 227

- Index** 243