EVIDENCE-BASED SOFTWARE ENGINEERING AND SYSTEMATIC REVIEWS

Barbara Ann Kitchenham
Keele University, Staffordshire, UK

David Budgen
Durham University, UK

Pearl Brereton
Keele University, Staffordshire, UK
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>xv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xvii</td>
</tr>
<tr>
<td>Preface</td>
<td>xix</td>
</tr>
<tr>
<td>Glossary</td>
<td>xxiii</td>
</tr>
<tr>
<td>I Evidence-Based Practices in Software Engine</td>
<td>1</td>
</tr>
<tr>
<td>1 The Evidence-Based Paradigm</td>
<td>3</td>
</tr>
<tr>
<td>1.1 What do we mean by evidence?</td>
<td>4</td>
</tr>
<tr>
<td>1.2 Emergence of the evidence-based movement</td>
<td>7</td>
</tr>
<tr>
<td>1.3 The systematic review</td>
<td>10</td>
</tr>
<tr>
<td>1.4 Some limitations of an evidence-based view of the world</td>
<td>14</td>
</tr>
<tr>
<td>2 Evidence-Based Software Engineering (EBSE)</td>
<td>17</td>
</tr>
<tr>
<td>2.1 Empirical knowledge before EBSE</td>
<td>17</td>
</tr>
<tr>
<td>2.2 From opinion to evidence</td>
<td>19</td>
</tr>
<tr>
<td>2.3 Organising evidence-based software engine</td>
<td>23</td>
</tr>
<tr>
<td>2.4 Software engineering characteristics</td>
<td>25</td>
</tr>
<tr>
<td>2.5 Limitations of evidence-based practices in software engineering</td>
<td>27</td>
</tr>
<tr>
<td>2.5.1 Constraints from software engineering</td>
<td>27</td>
</tr>
<tr>
<td>2.5.2 Threats to validity</td>
<td>28</td>
</tr>
<tr>
<td>3 Using Systematic Reviews in Software Engine</td>
<td>31</td>
</tr>
<tr>
<td>3.1 Systematic reviews</td>
<td>32</td>
</tr>
<tr>
<td>3.2 Mapping studies</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Meta-analysis</td>
<td>37</td>
</tr>
</tbody>
</table>
4 Planning a Systematic Review

4.1 Establishing the need for a review ........................................ 40
4.2 Managing the review project .................................................. 43
4.3 Specifying the research questions ........................................... 43
4.4 Developing the protocol .......................................................... 48
  4.4.1 Background ........................................................................ 49
  4.4.2 Research questions(s) ....................................................... 49
  4.4.3 Search strategy ................................................................. 49
  4.4.4 Study selection ................................................................. 50
  4.4.5 Assessing the quality of the primary studies ....................... 50
  4.4.6 Data extraction ................................................................. 51
  4.4.7 Data synthesis and aggregation strategy ............................. 51
  4.4.8 Limitations ....................................................................... 52
  4.4.9 Reporting ......................................................................... 52
  4.4.10 Review management ......................................................... 52
4.5 Validating the protocol ............................................................. 52

5 Searching for Primary Studies .................................................... 55

5.1 Completeness .......................................................................... 56
5.2 Validating the search strategy .................................................. 59
5.3 Methods of searching ............................................................... 62
5.4 Examples of search strategies ................................................... 64

6 Study Selection ............................................................................ 67

6.1 Selection criteria ...................................................................... 67
6.2 Selection process ...................................................................... 69
6.3 The relationship between papers and studies ............................ 71
6.4 Examples of selection criteria and process ............................... 72

7 Assessing Study Quality ................................................................. 79

7.1 Why assess quality? ................................................................. 79
7.2 Quality assessment criteria ....................................................... 82
  7.2.1 Study quality checklists ..................................................... 83
  7.2.2 Dealing with multiple study types ....................................... 86
7.3 Procedures for assessing quality .............................................. 86
7.4 Examples of quality assessment criteria and procedures .......... 88

8 Extracting Study Data ................................................................. 93

8.1 Overview of data extraction ..................................................... 93
8.2 Examples of extracted data and extraction procedures .......... 95
9 Mapping Study Analysis 101

9.1 Analysis of publication details 102
9.2 Classification analysis 103
9.3 Automated content analysis 106
9.4 Clusters, gaps, and models 110

10 Qualitative Synthesis 111

10.1 Qualitative synthesis in software engineering research 112
10.2 Qualitative analysis terminology and concepts 113
10.3 Using qualitative synthesis methods in software engineering systematic reviews 116
10.4 Description of qualitative synthesis methods 117
10.4.1 Meta-ethnography 118
10.4.2 Narrative synthesis 120
10.4.3 Qualitative cross-case analysis 121
10.4.4 Thematic analysis 123
10.4.5 Meta-summary 124
10.4.6 Vote counting 127
10.5 General problems with qualitative meta-synthesis 129
10.5.1 Primary study quality assessment 129
10.5.2 Validation of meta-syntheses 130

11 Meta-Analysis 133

with Lech Madeyski

11.1 Meta-analysis example 134
11.2 Effect sizes 135
11.2.1 Mean difference 136
11.2.2 Standardised mean difference 138
11.2.2.1 Standardised mean difference effect size 138
11.2.2.2 Standardised difference effect size variance 140
11.2.2.3 Adjustment for small sample sizes 141
11.2.3 The correlation coefficient effect size 141
11.2.4 Proportions and counts 142
11.3 Conversion between different effect sizes 144
11.3.1 Conversions between d and r 144
11.3.2 Conversion between log odds and d 144
11.4 Meta-analysis methods 145
11.4.1 Meta-analysis models 145
11.4.2 Meta-analysis calculations 146
11.5 Heterogeneity 148
11.6 Moderator analysis 151
11.7 Additional analyses 152
## Contents

11.7.1 Publication bias ........................................ 152  
11.7.2 Sensitivity analysis ..................................... 153  

12 Reporting a Systematic Review .......................... 155  
12.1 Planning reports ........................................ 157  
12.2 Writing reports ......................................... 158  
12.3 Validating reports ....................................... 162  

13 Tool Support for Systematic Reviews .................. 165  
with Christopher Marshall  
13.1 Review tools in other disciplines ..................... 166  
13.2 Tools for software engineering reviews ................ 169  

14 Evidence to Practice: Knowledge Translation and Diffusion 173  
14.1 What is knowledge translation? ........................ 175  
14.2 Knowledge translation in the context of software  
engineering ................................................... 177  
14.3 Examples of knowledge translation in software engineering  
14.3.1 Assessing software cost uncertainty ............... 180  
14.3.2 Effectiveness of pair programming ................. 181  
14.3.3 Requirements elicitation techniques ............... 181  
14.3.4 Presenting recommendations ....................... 182  
14.4 Diffusion of software engineering knowledge .......... 183  
14.5 Systematic reviews for software engineering education  
14.5.1 Selecting the studies ............................... 184  
14.5.2 Topic coverage ...................................... 185  

Further Reading for Part I .................................. 187  
II The Systematic Reviewer’s Perspective of  
Primary Studies ............................................. 195  
15 Primary Studies and Their Role in EBSE ................ 197  
15.1 Some characteristics of primary studies ............... 199  
15.2 Forms of primary study used in software engineering  
15.3 Ethical issues ........................................... 203  
15.4 Reporting primary studies .............................. 205  
15.4.1 Meeting the needs of a secondary study .......... 205  
15.4.2 What needs to be reported? ........................ 208  
15.5 Replicated studies ...................................... 208  
Further reading ............................................. 209
16 Controlled Experiments and Quasi-Experiments

16.1 Characteristics of controlled experiments and quasi-experiments ........................................... 212
  16.1.1 Controlled experiments .................................................. 212
  16.1.2 Quasi-experiments ..................................................... 214
  16.1.3 Problems with experiments in software engineering ..................................................... 215

16.2 Conducting experiments and quasi-experiments ................................................................. 217
  16.2.1 Dependent variables, independent variables and confounding factors ........................................ 218
  16.2.2 Hypothesis testing ....................................................... 219
  16.2.3 The design of formal experiments ............................................. 221
  16.2.4 The design of quasi-experiments .......................................... 222
  16.2.5 Threats to validity ...................................................... 223

16.3 Research questions that can be answered by using experiments and quasi-experiments ..................... 225
  16.3.1 Pair designing .......................................................... 226
  16.3.2 Comparison of diagrammatical forms ........................................ 227
  16.3.3 Effort estimation ........................................................ 227

16.4 Examples from the software engineering literature .................................................................... 227
  16.4.1 Randomised experiment: Between subjects .................................................. 228
  16.4.2 Quasi-experiment: Within-subjects before–after study ............................................... 228
  16.4.3 Quasi-experiment: Within-subjects cross-over study ................................................... 228
  16.4.4 Quasi-experiment: Interrupted time series ......................................................... 229

16.5 Reporting experiments and quasi-experiments ................................................................. 229

Further reading ..................................................................................................................... 230

17 Surveys

17.1 Characteristics of surveys .................................................................................................. 234

17.2 Conducting surveys .......................................................................................................... 236

17.3 Research questions that can be answered by using surveys .................................................... 238

17.4 Examples of surveys from the software engineering literature ............................................. 239
  17.4.1 Software development risk ................................................................................... 240
  17.4.2 Software design patterns .................................................................................... 240
  17.4.3 Use of the UML .............................................................................................. 242

17.5 Reporting surveys .......................................................................................................... 242

Further reading ..................................................................................................................... 242

18 Case Studies

18.1 Characteristics of case studies .......................................................................................... 247

18.2 Conducting case study research ...................................................................................... 248
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.2.1</td>
<td>Single-case versus multiple-case</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>18.2.2</td>
<td>Choice of the units of analysis</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>18.2.3</td>
<td>Organising a case study</td>
<td>251</td>
<td></td>
</tr>
<tr>
<td>18.3</td>
<td>Research questions that can be answered by using case studies</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>18.4</td>
<td>Example of a case study from the software engineering literature</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>18.4.1</td>
<td>Why use a case study?</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>18.4.2</td>
<td>Case study parameters</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>18.5</td>
<td>Reporting case studies</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further reading</td>
<td>258</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Qualitative Studies</td>
<td>259</td>
<td></td>
</tr>
<tr>
<td>19.1</td>
<td>Characteristics of a qualitative study</td>
<td>259</td>
<td></td>
</tr>
<tr>
<td>19.2</td>
<td>Conducting qualitative research</td>
<td>260</td>
<td></td>
</tr>
<tr>
<td>19.3</td>
<td>Research questions that can be answered using qualitative studies</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>19.4</td>
<td>Examples of qualitative studies in software engineering</td>
<td>262</td>
<td></td>
</tr>
<tr>
<td>19.4.1</td>
<td>Mixed qualitative and quantitative studies</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>19.4.2</td>
<td>Fully qualitative studies</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>19.5</td>
<td>Reporting qualitative studies</td>
<td>267</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further reading</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Data Mining Studies</td>
<td>271</td>
<td></td>
</tr>
<tr>
<td>20.1</td>
<td>Characteristics of data mining studies</td>
<td>272</td>
<td></td>
</tr>
<tr>
<td>20.2</td>
<td>Conducting data mining research in software engineering</td>
<td>272</td>
<td></td>
</tr>
<tr>
<td>20.3</td>
<td>Research questions that can be answered by data mining</td>
<td>274</td>
<td></td>
</tr>
<tr>
<td>20.4</td>
<td>Examples of data mining studies</td>
<td>275</td>
<td></td>
</tr>
<tr>
<td>20.5</td>
<td>Problems with data mining studies in software engineering</td>
<td>276</td>
<td></td>
</tr>
<tr>
<td>20.6</td>
<td>Reporting data mining studies</td>
<td>277</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further reading</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Replicated and Distributed Studies</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>21.1</td>
<td>What is a replication study?</td>
<td>279</td>
<td></td>
</tr>
<tr>
<td>21.2</td>
<td>Replications in software engineering</td>
<td>282</td>
<td></td>
</tr>
<tr>
<td>21.2.1</td>
<td>Categorising replication forms</td>
<td>282</td>
<td></td>
</tr>
<tr>
<td>21.2.2</td>
<td>How widely are replications performed?</td>
<td>284</td>
<td></td>
</tr>
<tr>
<td>21.2.3</td>
<td>Reporting replicated studies</td>
<td>286</td>
<td></td>
</tr>
<tr>
<td>21.3</td>
<td>Including replications in systematic reviews</td>
<td>286</td>
<td></td>
</tr>
<tr>
<td>21.4</td>
<td>Distributed studies</td>
<td>287</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Further reading</td>
<td>289</td>
<td></td>
</tr>
</tbody>
</table>
III Guidelines for Systematic Reviews

22 Systematic Review and Mapping Study Procedures

22.1 Introduction .............................................. 295
22.2 Preliminaries ............................................. 297
22.3 Review management ...................................... 298
22.4 Planning a systematic review ............................ 299
   22.4.1 The need for a systematic review or mapping study ........... 299
   22.4.2 Specifying research questions ......................... 302
      22.4.2.1 Research questions for systematic reviews ............... 302
      22.4.2.2 Research questions for mapping studies .......... 302
   22.4.3 Developing the protocol .................................. 304
   22.4.4 Validating the protocol ............................... 304
22.5 The search process .................................. 306
   22.5.1 The search strategy .................................. 306
      22.5.1.1 Is completeness critical? ......................... 306
      22.5.1.2 Validating the search strategy ............... 307
      22.5.1.3 Deciding which search methods to use ............. 309
   22.5.2 Automated searches .................................. 310
      22.5.2.1 Sources to search for an automated search ......... 310
      22.5.2.2 Constructing search strings ................. 311
   22.5.3 Selecting sources for a manual search ................ 313
   22.5.4 Problems with the search process .................... 314
22.6 Primary study selection process ........................ 315
   22.6.1 A team-based selection process ...................... 315
   22.6.2 Selection processes for lone researchers ............. 318
   22.6.3 Selection process problems ............................ 318
   22.6.4 Papers versus studies ................................ 319
   22.6.5 The interaction between the search and selection processes .......... 321
22.7 Validating the search and selection process ............ 321
22.8 Quality assessment .................................. 322
   22.8.1 Is quality assessment necessary? ..................... 323
   22.8.2 Quality assessment criteria ............................ 323
      22.8.2.1 Primary study quality .............................. 323
      22.8.2.2 Strength of evidence supporting review findings .......... 324
   22.8.3 Using quality assessment results ..................... 328
   22.8.4 Managing the quality assessment process ............. 328
      22.8.4.1 A team-based quality assessment process .......... 329
      22.8.4.2 Quality assessment for lone researchers .......... 330
22.9 Data extraction ........................................ 331
  22.9.1 Data extraction for quantitative systematic
        reviews ........................................... 331
    22.9.1.1 Data extraction planning for quantitative
            systematic reviews ......................... 331
    22.9.1.2 Data extraction team process for quanti-
            tative systematic reviews .................. 334
    22.9.1.3 Quantitative systematic reviews data
            extraction process for lone researchers ... 335
  22.9.2 Data extraction for qualitative systematic reviews ... 336
    22.9.2.1 Planning data extraction for qualitative
            systematic reviews ......................... 337
    22.9.2.2 Data extraction process for qualitative
            systematic reviews .......................... 337
  22.9.3 Data extraction for mapping studies .................. 338
    22.9.3.1 Planning data extraction for mapping
            studies ....................................... 338
    22.9.3.2 Data extraction process for mapping
            studies ....................................... 340
  22.9.4 Validating the data extraction process ............... 342
  22.9.5 General data extraction issues ...................... 342
22.10 Data aggregation and synthesis .......................... 343
  22.10.1 Data synthesis for quantitative systematic reviews . 343
    22.10.1.1 Data synthesis using meta-analysis ........ 344
    22.10.1.2 Reporting meta-analysis results ........... 346
    22.10.1.3 Vote counting for quantitative systematic
            reviews ..................................... 347
  22.10.2 Data synthesis for qualitative systematic reviews .... 348
  22.10.3 Data aggregation for mapping studies ................ 350
    22.10.3.1 Tables versus graphics ................... 351
  22.10.4 Data synthesis validation ........................ 351
22.11 Reporting the systematic review ........................ 353
  22.11.1 Systematic review readership ........................ 353
  22.11.2 Report structure ................................ 353
  22.11.3 Validating the report ................................ 355

Appendix: Catalogue of Systematic Reviews Relevant to
Education and Practice .................................. 357

with Sarah Drummond and Nikki Williams

A.1 Professional Practice (PRF) .......................... 358
A.2 Modelling and Analysis (MAA) .......................... 359
A.3 Software Design (DES) ................................ 361
A.4 Validation and Verification (VAV) ........................ 361
A.5 Software Evolution (EVO) .............................. 362
A.6 Software Process (PRO) ................................ 363