EFL coursebook innovation: The dynamics of L2 maturity, student engagement, and cognitive development

Robert S. Murphy

Thesis submitted to the University of Nottingham
For the degree of Doctor of Philosophy

May 2019

Abstract

This thesis documents the dynamically changing efficacy of an EFL coursebook series that has been evolving in various ways since 2008, via action research methodology. The *Optimal Levels* EFL coursebook series was specifically written for the Japanese EFL context—but with the strong intent of raising students' cognitive maturity along with their L2 competence. Because of this non-conventional attention to maturity in combination with L2 learning, the ensuing research of *Optimal Levels*' efficacy required a non-conventional evaluation system. This thesis therefore serves to document the evolution of the series (with all of its theoretical underpinnings), and also document the evolution of the various evaluation systems that were developed for this somewhat convoluted project.

To accomplish these goals, the research not only follows the growth of Japanese EFL students' L2 competence, but it also follows the changes of students' feelings of *ownership* of the L2, their changes in engagement with provided L2 tasks, their changes in metacognitive realizations regarding learning/teaching, and also their L2 maturation rate changes under several different conditions.

The efficacy of the original *Optimal Levels* (OL1) was compared to that of *New Optimal Levels* (OL2). Both books have a significant amount of overlap in their pedagogical approach and learning tasks. From a historical perspective, *OL1* was developed as an integral part of my MA thesis while *OL2* was developed as an integral part of this doctoral thesis. The design of *OL2* evolved along with this study's findings. Therefore, in the second half of the thesis I will discuss the developments that led to *OL2 Revised* (an updated version of *OL2* that I also piloted with students for this project). In broad strokes, it can be said that the original design of *OL1* (and then *OL2* and *OL2-R)* stem from developments in *Applied Linguistics*, *Cognitive Psychology* and *Educational Neuroscience*.

Quantitative analysis (SPSS-based statistics and Lectical Analysis) was used to measure the difference in the growth patterns of two groups of university students; the results of the *OL1* group were compared with the results of the *OL2* group. Lectical Analysis data showed that the *OL2* group outpaced the *OL1* group by an 8x growth rate of demonstrable L2 complexity--with a single 15-week semester of teaching. Qualitative analysis (student-based focus group interviews, questionnaires, and evaluation essays from university students) was used to deduce the specifics behind the performance gap between the two groups. The data revealed that *OL2* raised metacognitive awareness and feelings of L2 ownership in the students; participants in the *OL2* group were more willing to discuss the *hows* and *whys* of their L2 growth, while the *OL1* group focused more on *what they did* in class, which is considerably less of a cognitive challenge for students in this age group.

The compilation of data from the variety of evolving studies suggested significant connections across: (a) cyclically designed deep-thinking tasks, (b) metacognitive nurturing in the L2 classroom, and (c) rapid L2 maturation rates. The implication is that in the EFL classroom, a purely "pragmatic" focus on linguistic development can be ironically counterproductive in raising the maturation rates of EFL students' L2 competence.

Two sets of tenets for classroom teachers and material developers were developed as a result of these studies; the tenets are a culmination of all of the major insights that were acquired across this six-year project. For the *finale* of the thesis, the tenets were used to upgrade the design of *OL2-R* into what will become *OL3.* In this way, the ending of the thesis ties the knot between: (1) theoretical understandings of the amassed data and (2) practical applications of the amassed data from this action research project.

Acknowledgements

I would like to thank all the people who helped make this thesis possible. I owe sincere gratitude to my supervisor, Professor Zoltán Dörnyei, for his support and guidance throughout my doctoral studies. I am particularly grateful to him because what I brought to him—in the form of a doctoral proposal—was nothing but an enigma. Rather than turning me away, he took me on and worked with me tirelessly on this project. I could not imagine having a better advisor and mentor for this study.

I would also like to thank Kurt Fischer (Harvard Graduate School of Education) and Theo Dawson (Lectica, Inc.) for their immeasurable support and depth of knowledge regarding cognitive development and growth assessment. This study would not have been possible without their massive contributions to the field, and their kind mentorship.

Next, I would like to thank the long-term *neuroELT* research group leaders, whom are very much like family to me now: Marc Helgesen, Curtis Kelly, Joseph Shaules, Tim Murphey, and Amanda Gillis-Furutaka. Our work and growth as a group is what fueled the engines of this project. Thank you for all the inspiration (and perspiration) over the years.

I would also like to sincerely thank Tracey Tokuhama-Espinosa, Joseph Vitta, Ali Al-Hoorie, and Tom Gorham for their writing advice and invaluable insights. Also, a warm thank you to my DMC research buddies, Zana Ibrahim, Takashi Uemura, and Christopher Stubbings. And, all of the super-dedicated *neuroELT* interns and tutees from the past decade. You guys are the best.

And finally, I would like to thank my wife, Ai, and my son, Kevin, for their tireless moral support and understanding on this long journey. I love you both very much. I could not have completed this journey without you.

TABLE OF CONTENTS

Abstract 3

Acknowledgements 5

1. Introduction 14

2. Literature Review Chapter: Syllabus Design 19

2.1 L2 Syllabus Design ***19***

2.2 The syllabus purpose ***20***

2.3 Is the coursebook a syllabus? ***22***

2.4 Syllabus Analysis ***23***

2.4.1 Product vs. Process 24

2.4.2 Synthetic vs. Analytic 25

2.4.3 Linear vs. Cyclical 26

2.4.4 Synthesis: Syllabus Analysis 27

2.5 Syllabus Design Factors ***28***

2.5.1 Design Factors 28

2.5.2 Toward a Dynamic Design 30

2.6 Instructional blocks ***34***

2.6.1 Motivation in material development 35

2.6.2 Synthesis: Syllabus Design Factors 36

2.7 Some Common Syllabus types ***37***

2.7.1 Grammatical Syllabus 37

2.7.2 Notional-functional Syllabus 46

2.7.3 Lexical Syllabus 50

2.7.4 Task-based Syllabus 57

2.7.5 Content-based syllabus 63

2.7.6 What can be learned from the “Spiral Curriculum”? 67

2.8 Chapter Conclusions ***68***

3. Literature Review Chapter: L2 Materials Design Theory 71

3.1 Introduction to Learning Materials ***72***

3.1.1 What are the pragmatic purposes of the L2 coursebook? 72

3.1.2 A brief history of Methods and Approaches 74

3.1.3 More dichotomies to consider 76

3.1.4 Can coursebook usage be detrimental? 76

3.1.5 Synthesis: Introduction to Learning Materials 83

3.2 L2 Learning Material Evaluation ***86***

3.2.2 Materials Evaluation Procedures 87

3.2.3 In-use and Post-use Evaluation 91

3.2.4 Synthesis 93

3.3 L2 Learning Material Adaptation and Designing ***94***

3.3.1 Adaptation 94

3.3.2 Designing materials 97

3.3.3 Using Authentic Materials 97
3.3.4 Critique 98
3.4 Chapter Conclusion: Looking toward the future ***99***
3.4.1 Toward conclusions 99
3.4.2 What has been uncovered? What can be done? 99
3.4.3 Discussion: Designing cutting-edge materials 101
3.4.4 Motivational factors 102
3.5 Learning lessons from science education literature ***104***
3.5.1 Difficulties in changing teachers' beliefs regarding pedagogical assumptions 104
3.5.2 Science education: teachers' beliefs/views of science as a school subject 108
3.5.3 What other issues are associated with top-down/ for-the-test methods? 110
3.5.4 Teachers' beliefs/views: Top-down vs. bottom-up? 111
3.5.5 Views/beliefs vs. classroom realities 112
3.5.6 Pedagogical content knowledge (PCK) 114
3.5.7 Using Schulman vs. Gardner as a metaphor for a deeper understanding of PCK 116
3.5.8 PCK mindsets: "Top-down vs. bottom-up" mindset in lieu of Shulman's mindset 118
3.5.9 The Knowledge Quartet (KQ) 121
3.6 Materials design in sum:
A new protocol based on Dynamic Equity is required ***128***

4. Foundations of OL1 **129**
4.1 Preamble: The original Optimal Levels! coursebook ***129***
4.1.1 The pedagogical design of OL1 129
4.1.2 The findings of a 2-year study of OL1 133
4.2 OL2 on the surface ***134***
4.2.1 The newer protocol (COOPS) 137
4.3 The details of the OL2 design ***138***
4.3.1 OL2's Table of Contents 138
4.3.2 "Self-assessment sheet (pre)" 140
4.3.3 "About me!" 143
4.3.4 The Module introduction 146
4.3.5 "Cycle 3a. Production: Word categories" 151
4.3.6 "Cycle 3b. Production: Context" 154
4.3.7 "Cycle 4b. Exchange: Grammar Ideas from the teacher" 157
4.3.8 "Cycle 5a. Production – Group Master List!" 161
4.3.9 "Cycle 5b. Class Grammar Templates List" 164
4.3.10 "Cycle 6. Group-owned: Presentation!" 166
4.3.11 "Cycle 7. Presentation Feedback" 169
4.3.12 "Cycle 7b. Presentation Feedback notes" 173
4.3.13 "Cycle 8. Group Work: Design a test!" 176

4.3.14 "Cycle 9. Choice: Best Test Design Award!" 179
4.3.15 "Cycle 10. Testing, Reflection, and New Plans!" 182
4.3.16 "(OPTIONAL) THINK AGAIN" 186
4.4 Concluding remarks ***188***

5. Cognitive assessments: Lectica, Dynamic Skill Theory, Knowledge Quartet, and Microdevelopment 190
5.1 An introduction to Lectica – and, why use it for L2 contexts? ***190***
5.1.1 How does it work? 191
5.1.2 The Lectica "epiphany" and the rationale behind it 195
5.2 Lectical Assessment's validity ***196***
5.2.1 Implications: A better way to look at communicative competence? 197
5.2.2 But, why not linear? 198
5.2.3 In sum 200
5.3 Theoretical background ***201***
5.3.1 The need for the neuroELT niche 201
5.3.2 The "Knowledge Quartet" repurposed 202
5.3.3 How knowledge develops in the brain 205
5.3.4 Dynamic Skill Theory – the basics 209
5.4 In sum: How all of this informed the design of OL2 ***214***

6. Research Design and Methodology 216
6.1 Summary of Research Design ***216***
6.2 Participants ***217***
6.2.1 Semester 1: English Course Participants 217
6.2.2 Semester 1: Lectical Assessment participants 217
6.2.3 Semester 1: Focus group participants 218
6.2.4 Semester 1: Main student questionnaire participants 219
6.2.5 Semester 1: Standardized bi-annual university student questionnaire participants 219
6.2.6 Semester 2: English Course Participants 219
6.2.7 Semester 2: Final Focus Group participants 220
6.2.8 Semester 2: Main student questionnaire participants (Mid and Post) 220
6.2.9 Semester 2: Year-end "My Change" essays 221
6.3 Instruments and Procedures ***222***
6.3.1 Main student questionnaire (SurveyMonkey) 224
6.3.2 Lectical Assessment 226
6.3.3 Focus Group: Interviews and schedules 229
6.3.4 Standardized bi-annual university questionnaire participants 231
6.3.5 Year-end "My Change" essays 233
6.4 Summary of main data types collected ***235***

7. Evaluation of the coursebooks based on student questionnaires and Lectical Analysis 236
7.1 University student questionnaire 237
7.2 Comparing OL2 to world-class coursebooks 239
7.3 My SurveyMonkey student questionnaire 243
7.3.1 Results 245
7.3.2 Current proficiency 245
7.3.3 Discussion of "Current proficiency" 246
7.3.4 English usage in my future 247
7.3.5 Discussion of "English usage in my future" 248
7.3.6 The replication study (2015-2016) 252
7.3.7 The year-end final question 255
7.4 Lectical Analysis results 256
7.4.1 Understanding the data 257
7.4.2 Presentation marks vs. cognitive growth: The need to establish L2 Maturity 259
7.4.3 Discussion 262
7.4 Five years later: the affect of teacher rapport 269
7.5 Conclusion 270

8. Evaluation of the coursebooks based on interviews and student essays 273
8.1 Overview of data sources and data types 273
8.2 How the themes emerged 275
8.2.1 Generation of the frequency scores 277
8.3 Positive and negative themes of OL2 from the focus groups 278
8.4 Student appraisals of the sections of OL2 from the focus group 283
8.4.1 The most popular cycles 285
8.4.2 The least popular cycles 292
8.5 Written student appraisals at the end of the first semester 297
8.5.1 Positive themes 301
8.5.2 Negative themes 304
8.6 Written student appraisals at the end of the year 306
8.6.1 The positive themes 307
8.6.2 The negative themes 310
8.7 A composite of all the data sets from the primary 2014-2015 study 312
8.8 Student appraisal of OL1 315
8.8.1 Student evaluation of the different sections 315
8.8.2 Data from the Focus Group 318
8.8.3 Written data 319
8.9 Comparison of OL1 and OL2 321
8.10 Four years later: Data collection from the 2017-2018 school year 324
8.10.1 Qualitative summary diagrams 324
8.10.2 Relevant quotes for top themes 326

8.10.3 Student generated theoretical diagrams regarding OL2 pedagogical flow 332
8.11 Five years later: The affect of teacher rapport on motivation and course evaluations 338
8.11.1 Quantitative data 339
8.11.2 Qualitative data 342
8.12 Conclusion 350

9. Emerging tenets and implications 358
9.1 Implications from the Literature review for the development of OL3 358
9.1.1 The 19 "classic" tenets based on the literature review 359
9.2 Implications from my own data (toward the development of OL3) 361
9.2.1 The 20 tenets based on EFL student data 362
9.3 Proposed modifications for OL2 365
9.3.1 Main modification tier 366
9.3.2 Other noteworthy modifications: the second tier. 370
9.3.3 Other minor modifications: the third tier 377
9.4 Conclusion 378

10. Conclusion: The implications that go beyond textbook design and assessment 382
10.1 L2 Maturation: Lectical assessment 383
10.1.1 L2 Maturity as a mathematical equation 384
10.1.2 L2 Maturity as an Applied Linguistics construct (non-mathematical approach) 386
10.1.3 Why is this important (the new insights) 389
10.1.4 L2 Maturity in detail 389
10.2 Metacognitive development 394
10.3 Good teaching: Importance of top-down vs bottom-up ratios in the classroom 399
10.4 Good teaching: Revisiting COOPs and "Teaching for the DATC" 401
10.5 Good teaching: Connection to the neuroELT Maxims 405
10.6 Limitations 407
10.7 Future research directions 409
10.7.1 DMCs: Directed motivation currents 410
10.8 Final words 412

References 414

Appendix 428
Appendix A. Tier 3 modification list 428
Appendix B: Lectical analysis test 431
Appendix C: Lectical analysis, collected student responses 435
Appendix D: University questionnaire results 446
Appendix E. Survey Monkey questionnaire 447
Appendix F. OL1 Sample chapter 452
Appendix G. New Optimal Levels (OL2) Revised version, Sample chapter 466

List of Figures

Figure 1: The dynamic effects of the Approach-Design-Procedure framework. 24
Figure 2: The inherent microcosmic nature of the product syllabus and test cycle. 25
Figure 3: The Materials Design Factor Assessment Sheet29
Figure 4: Graves (2000): Framework for Course Development 30
Figure 5: Dynamic syllabus design, five-point and cyclical 31
Figure 6: OUP's levels and number of titles in brackets (accessed 2015) 32
Figure 7: Richards' horizontal and vertical coherence models. 34
Figure 8: Components of motivational teaching practice, adapted from Dörnyei, 2001. 36
Figure 9: A map of HAVE usage, taken from a coursebook lexical activity (Cutting Edge, Longman). 53
Figure 10: Material analysis vs material evaluation 86
Figure 11: Checklist from McGrath (2002:33) 90
Figure 12: McDonough and Shaw's framework for material adaptation. 96
Figure 13: Editing an authentic passage vs. editing a synthetic passage 98
Figure 14: The relationship between Deci & Ryan, Pink, and Murphy's work on motivation. 102
Figure 15: Example of DNA sequencing and color-coded fMRI scans 106
Figure 16: Complex layers of effect surrounding the "nature of science" (Erduran and Dagher, 2014)107
Figure 17: My graphic representation of the sweet spot of teaching (based on Ausubel, 1977) 109
Figure 18: Immediate context changes what you see. From Trevor Shannon (in Eduran and Dagher, 2014)113
Figure 19: Comparison of PCK definitions of Shulman, Grossman, and Loughran et al. (Childs, 2018). 115
Figure 20: Adding PCK to the spectrum 119
Figure 21: Mortimer and Scott's interaction planning framework (2003). 122
Figure 22: Updated version of teaching spectrum 124
Figure 23: The original flowchart that Optimal Levels! (OL1) was based on Murphy (2009). [SIC content] 131
Figure 24: The flow of learning over the course of a single unit of study (Module) of OL1 132
Figure 25: A comparison of the designs of Original (OL1) vs. NewOL (OL2). Stars indicate "climax" activities. Thought bubbles indicate metalinguistic maturation periods.136
Figure 26: The foundation of OL2, the COOPs protocol. 138
Figure 27: Copy of page 4, "Table of Contents" 138
Figure 28: "Self-assessment sheet" from page 5, pre-Module work 140
Figure 29: Copy of page 6, "About me!", pre-Module work 143
Figure 30: Cycle 1 and Cycle 2, Page 11 146
Figure 31: Copy of page 12, "Cycle 3a" 151
Figure 32: Copy of pages 13-14, "Cycle 3b" 154
Figure 33: Copy of page 16, "Cycle 4b" 157
Figure 34: Copy of page 13, "Cycle 5a" 161
Figure 35: Copy of page 14, "Cycle 5b" 164
Figure 36: Copy of page 19, "Cycle 6" 166
Figure 37: Copy of page 22, "Cycle 7" 169
Figure 38: Copy of page 23, "Cycle 7b" 173
Figure 39: Copy of pages 24-25, "Cycle 8" 176
Figure 40: Copy of page 26, "Cycle 9" 179
Figure 41: Copy of page 27, "Cycle 10" 182
Figure 42: Copy of page 28, "(Optional) Think Again" 186
Figure 43: Growth patterns per student are domain specific(Dawson & Stein, 2011). 194
Figure 44: Four out of thirteen levels. Adults typically score somewhere in this zone. 197
Figure 45: The Knowledge Quartet, Equality vs. Equity, and Low support vs. High support 203
Figure 46: Microdevelopment: Natural peaks and dips (Fischer, 2008:143) 206
Figure 47: The three waveforms identified in microdevelopment (Fischer, 2008:144) 207
Figure 48: Microdevelopment-based recommendations for teachers (adapted from Murphy, 2017) 208
Figure 49: A graphic representation of the primary "construction sites" uncovered by EEG scans in the alpha band (adapted from Fischer & Bidell, 2006: 389) 209
Figure 50: Dynamic skill theory: Optimal levels and functional levels (adapted from Murphy, 2009) 211
Figure 51: Four steps in development in dynamic skill theory (Adapted from Murphy, 2015) 212
Figure 52: Using the sport of Baseball to exemplify how individual skills connect and compound to arrive at the goal. Several systems of skills must be created and then meshed by the end to finally achieve a "system of systems" – enabling the child to finally play baseball. (Adapted from Murphy, 2015) 213
Figure 53: Screenshot from one portion of the main student questionnaire 224

Figure 54: Focus group discussion protocol, used for every FG session 229
Figure 55: Main questions, with responses on a 1-5 scale 232
Figure 56: Sample writing of the final reflexive essay. Pseudonyms are used. 234
Figure 57: The nine-part dataset from the four-year longitudinal study of first-year students' attitudes toward the OL2-based course. 239
Figure 58: OL2 vs. Department composite, course ratings per year 242
Figure 59: SurveyMonkey questions that were selected for this study (full questionnaire in Appendix)244
Figure 60: How self-assessment of "Current proficiency" changed across the year. 246
Figure 61: How self-assessment of "English usage in my future" changed across the year. 248
Figure 62: Students' perceived proficiency mean growth, primary study vs. replication study. 252
Figure 63: Students' confidence in future English usage, primary study vs. replication study. 254
Figure 64: The OL2 (NewOL) group showed a roughly 8x faster rate of demonstrated growth over the OL1 (Original) group, on the Lectical scale.258
Figure 65: Presentation evaluation trends OL1 vs. OL2 260
Figure 66: Projected growth rates for two semesters (full year), based on first semester growth 264
Figure 67: Reiterating Dynamic Skill Theory in light of OL1 vs OL2 growth trajectories. 266
Figure 68: Why teach strategies? Example flowchart of how the development of executive function can deepen student understanding (Meltzer, 2007:167)268
Figure 69: The focus group interview question list (meeting protocol)274
Figure 70: The finalized list of paired themes for qualitative coding. 277
Figure 71: Positive vs. negative theme ratios, OL1 (Original) and OL2 (NewOL) focus group meetings 278
Figure 72: List of coded and paired themes from OL1 and OL2 (focus group data) 279
Figure 73: Fractals as an image of cyclical pedagogical flow and development 284
Figure 74: Bar graphs of the "chosen best" cycles (focus groups). 286
Figure 75: Cycle 6: "Group Owned: Presentation!" 287
Figure 76: Cycle 4a: "Exchange: Lexical ideas from the teacher" 289
Figure 77: Popular vs. unpopular cycles 293
Figure 78: The unpopular Cycle 5b 295
Figure 79: The frequency of the positive themes from the assessment paragraphs, in graphic form. 303
Figure 80: Side-by-side of positive themes and negative themes that emerged from the paragraphs. 306
Figure 81: Side-by-side of positive themes and negative themes that emerged from the final essays. 311
Figure 82: Semester one and semester two, positive themes stacked in graphic form. 313
Figure 83: Popular vs. unpopular cycles 317
Figure 84: Graphic representation of OL2, popular vs. unpopular cycles 321
Figure 85: Graphic representation of OL1, popular vs. unpopular cycles 322
Figure 86: Student's theoretical diagram A 332
Figure 87: Student's theoretical diagram B 333
Figure 88: Student's theoretical diagram C 334
Figure 89: Student's theoretical diagram D 335
Figure 90: Student's theoretical diagram E 336
Figure 91: My diagram based on the student's diagrams 337
Figure 92: The reported affect of rapport on motivation, OL1 vs OL2 340
Figure 93: The reported affect of rapport on course evaluation, OL1 vs OL2 341
Figure 94: Comparing two flowcharts, OL2 vs OL3 ("Level up!") 367
Figure 95: Modification set A (right side is the modified version) 371
Figure 96: Modification set B (right side is the modified version)372
Figure 97: Modification set C (right side is the modified version) 373
Figure 98: Cycle 4b modification 374
Figure 99: "Think again" modification 376
Figure 100: Modifications for Self-assessment to Cycle 2 378
Figure 101: Modifications for Cycle 3 to Cycle 4b 379
Figure 102: Modifications for Cycle 5a to Cycle 7 379
Figure 103: Modifications for Cycle 7b to Cycle 10 380
Figure 104: Modifications for Cycle 10 to "Think again" 380
Figure 105: Tri-elemental model from the MA thesis (Murphy, 2008). 390
Figure 106: Illustration of students' understanding of their learning cycles 394
Figure 107: Revisiting the COOPs protocol. 402

List of Tables

Table 1: The tabulated Knowledge Quartet 121
Table 2: Evaluation overview of page 5, “Self-assessment” 142
Table 3: Evaluation overview of page 6, “About me!” 145
Table 4: Evaluation overview of page 11,” Cycles 1 and 2” 150
Table 5: Evaluation overview of page 12, “Cycle 3a” 153
Table 6: Evaluation overview of pages 13-14, “Cycle 3b” 156
Table 7: Evaluation overview of page 16, “Cycle 4b” 160
Table 8: Evaluation overview of page 13, “Cycle 5a” 163
Table 9: Evaluation overview of page 14, “Cycle 5b” 165
Table 10: Evaluation overview of page 19, “Cycle 6” 168
Table 11: Evaluation overview of page 22, “Cycle 7” 172
Table 12: Evaluation overview of page 23, “Cycle 7b” 175
Table 13: Evaluation overview of pages 23-25, “Cycle 8” 178
Table 14: Evaluation overview of page 26, “Cycle 9” 181
Table 15: Evaluation overview of page 27, “Cycle 10” 185
Table 16: Lectical ruler: Grades 5 to Post-doctorate (Lectica website, 2018). Used with permission. 192
Table 17: First semester enrollment, gender ratios, and coursebook choice. 217
Table 18: The two classes chosen for Lectical Analysis 218
Table 19: The seven focus group: numbers, gender ratios, and coursebook choice. 218
Table 20: The main student questionnaire: numbers, gender ratios, and coursebook choice. 219
Table 21: University questionnaire: numbers, gender ratios, and coursebook choice. 219
Table 22: Second semester: Coursebook, enrollment, and gender 220
Table 23: Semester two: Final focus group interview 220
Table 24: Semester 2: Main student questionnaire participants (Mid and Post) 221
Table 25: Second semester: Coursebook, enrollment, and gender 221
Table 26: Comparing OL1 and OL2 teaching schedules 222
Table 27: The schedule from the second semester 223
Table 28: Main topic range coved in the questionnaire 225
Table 29: Focus group interview schedule 230
Table 30: Comparing OL1 and OL2 teaching schedules 235
Table 31: Comparing OL2 class data with a composite of all NESTs in the department 240
Table 32: Results of the “Current proficiency” battery of self-assessment 245
Table 33: The positive themes that emerged and their frequency. 280
Table 34: The negative themes that emerged, their frequencies, and illustrative examples. 282
Table 35: Focus groups’ chosen best cycles’ frequencies. *Highest are highlighted 285
Table 36: Focus groups’ chosen worst cycles’ frequencies *Highest are highlighted 293
Table 37: Assessment paragraphs’ emerging themes 297
Table 38: The positive themes, their frequencies, and illustrative examples (first semester paragraphs) 298
Table 39: The negative themes, their frequencies, and illustrative examples (first semester paragraphs)305
Table 40: Established themes list from end of year final essays 307
Table 41: The positive themes, their frequencies, and illustrative examples (second semester essays) 308
Table 42: The negative themes, their frequencies, and illustrative examples (second semester essays) 310
Table 43: First and second semester, negative themes side-by-side. (*Largest change) 314
Table 44: The “chosen best” Sections in OL1. 316
Table 45: The “chosen worst” sections in OL1. 316
Table 46: Positive themes from the OL1 focus group sessions. 318
Table 47: Negative themes from the OL1 focus group sessions. 319
Table 48: Positive themes from the OL1 class paragraphs 319
Table 49: Negative themes from the OL1 class paragraphs 320
Table 50: 2014 data, University-administered questionnaire, open-ended section. Code on the top row. 325
Table 51: 2018 data, University-administered questionnaire, open-ended section. Codes on the top row. 325
Table 52: Illustrative quotes on rapport and motivation 343
Table 53: Illustrative quotes on rapport and course evaluation 346
Table 54: Justifications for the major modification—the leveling up task 369
Table 55: The seven neuroELT maxims that are fully compatible with the tenets from this thesis 406

1. Introduction

Teaching English in Japan since 1993, I have never been fully satisfied with the coursebooks that were traditionally being used in the language school classrooms, public schools, and even at the universities. Because I have had the freedom to choose my own coursebooks for my classrooms, I have trialed several dozen coursebooks since 1993—yet, dissatisfied with the majority of them, ten years ago I set out to produce my own teaching materials. This thesis documents the past six years of that journey.

I was dissatisfied with what I perceived to be a lack of scientific rationale in the pedagogical designs of the coursebooks that I had studied—and the apparent lack of scientific approaches being used to review those coursebooks. It seemed that many coursebooks were designed to *sell* to specific markets—and so, the priority for the publisher was raising market share rather than how well their books worked as a pedagogical system. What would an EFL coursebook look like if it put pedagogy first, and not cared about sales? With this question in mind, I worked on the design of my first full coursebook, *Optimal Levels!* (OL1), between 2008 and 2009. I then published *OL1* in 2010, with the not-for-profit publishing label *Deeper Understanding Books*. *OL1* is still in print (and is available via online bookstores). Since its debut in 2010, at least 2000 students have studied with *OL1* under my supervision. The worldwide student count is higher than that, but I do not possess that data. I still use *OL1* for some of my own university classes.

In this way, *OL1* was born out my internal need to produce and use something more pedagogically sound. It was like an itch that needed to be scratched. The production of *OL1* coincided with my MA thesis (University of Birmingham, MA TESOL). The thesis involved exploring theoretical backgrounds and my intuitions for better coursebook designs while making explicit connections to cognitive development—especially Fischer's Dynamic Skill Theory (Fischer, 1980).

With a grounded understanding of the issues involved in designing better EFL pedagogy (via my MA thesis involving a pilot study), *OL1* came to life.

In many ways, this Ph.D. study is a continuance of my MA project, albeit, with more elaboration in each aspect of the study. Other than the production of better-designed coursebooks, another primary intention has been to introduce rigorous empirical *evaluation procedures* that can inform the development of course material and syllabus design. This thesis, therefore, discusses a variety of evaluation procedures that were developed to test the outcomes of different versions of my coursebook series: *OL1*, *OL2* (*New Optimal Levels*), and then *OL2-R*, the revised version of *OL2*. Furthermore, *OL1* outcomes were compared to *OL2/OL2-R* outcomes, best practices of pedagogical design were learned and listed, and, by the end of the thesis, strategies for the development of an "*OL3"* came to life from those learned lessons.

Overall, this thesis ended up being a seven-year undertaking. The first year was spent designing what eventually became *OL2* (making *OL2* a direct evolution of *OL1*). During that first year, I also deliberated on how to best study *OL2's* efficacy—especially on how to best compare it with *OL1's* efficacy. Once the action research protocols were established with the help of my supervisor, Zoltán Dörnyei (University of Nottingham), and Theo Dawson (Lectica, Inc.), the second and third years were spent collecting data with Japanese EFL students at the university level.

At the Japanese university (the primary location of the data collection), I had six classes using *OL2* and one class using *OL1* (with two semesters per year). I conducted pre- and post-tests using Lectical Analysis (measuring the complexity of EFL student output) for *OL1* and *OL2* classes. I also collected the results of two different questionnaires, while concurrently conducting focus group interviews with volunteer groups from the *OL1* and *OL2* student bodies. The fourth and fifth years were spent analyzing the varieties of data for patterns, while concurrently writing up significant portions of the thesis. The finale of that period involved the development of tenets to be used as guidlines for the development of "*OL3"*. This turned out to be a showcase of the practical lessons that were learned cumulatively during the course

of this study. In the final months of this project, I focused primarily on cleaning up the writing while revisiting all the collected data (and updating my hypotheses as my understanding of the data became more refined). I also added new data near the end of the study (using questionnaires and student outcomes) to help uncover the importance of rapport building and the potential of a "halo effect" on my students' evaluations.

This Ph.D. project is keenly focused on EFL in the Japanese context; the results are unique to Japan. English teaching in Japan, be it in public schools, private language schools, or universities, has long-standing (and well-known) problems. Japan is a diehard EFL country—not an ESL country. Being half-Japanese myself, I can confidently say that neither the cultural context, nor the (lack of) motivation in Japanese students currently allows Japan to flip over to a naturally localized English form of usage such as can be seen in Singapore and in the Philippines—this is so even in the face of direct competition with South Korea and China (Japan's perceived closest rivals in Asia). Against this tough background, my project had significant challenges to overcome. And, to overcome these challenges, my pedagogical implementations are idiosyncratic and "over-built" to compensate for this challenging context.

What is the *OL* series specifically designed to compensate for? It is meant to compensate for the lack of motivation in the traditionally "apathetic-toward-English" whom are the majority of students living in Japan. The rationale is: the less-responsive population has most likely been under-whelmed thus far with their L2 education in Japan—so, more rigorously attention-grabbing and engaging approaches must be designed into the pedagogical systems, as a tactic to overcome the impeding lingering issues from past EFL courses.

Although the designs that I implemented are meant as a compensation strategy specifically for the context of Japan, I hypothesize that this "over-built" process can also be applied to other less difficult teaching contexts. A fitting adage for this Japan-based design is, *"If I can make it there, I can make it anywhere..."* Some justification for this line of thinking comes in the form of a string of

invitations to report these (Japan-based) findings at venues in Hong Kong, Vietnam, Macao, South Korea, and the Philippines – mostly as plenary presentations or faculty development workshops; I am happy to report that the response from teachers from these other countries, after they had trialed these tactics, has been positive.

Due to the nature of the study, the structure of this thesis is a bit unorthodox. After this introductory chapter, there will be a literature review spanning four chapters (Chapters 2, 3, 4, and 5) – with a small taste of the study results mixed in where it was more logical to do so than to wait until the Results chapter. I have chosen to write it this way under the advisement of my supervisor. This approach was taken to add an extra level of clarity and coherence to the complex arguments being made regarding the outcomes of the various pedagogical designs across multiple timelines within this action research endeavor. Similarly, in the second half of the thesis, where the results and discussions reside, I reiterate a few consequential points from the literature review where it is most logical to circle-back to them to allow the deductions being made in those areas to be more logical and easier to follow.

After the methodology overview (Chapter 6), there is a qualitative results chapter and a quantitative results chapter (Chapters 7 and 8, respectively). And then, Chapter 9 is a material-specific chapter based on the principle, “The proof of the pudding is in the eating”; it houses the lessons learned from the study including the justifications for all the upgrades that were proposed for the development of *OL3* (cyclically linking back to the results of *OL1* and *OL2* evaluations).

In the concluding chapter (Chapter 10), I move away from the keen focus on coursebook talk to view the bigger picture. In Chapter 10, I will focus on the generalizable theoretical lessons learned throughout this project that transcend the realm of coursebook writing theory.

This thesis is, therefore, a multi-faceted action research project that will take what has been learned and provide ideas and proposals for:

(1) coursebook design innovations
(2) coursebook evaluation innovations
(3) understanding the cognitive development of students
(4) understanding students' metacognition regarding their L2 learning
(5) understanding the affect of teacher rapport on motivation and evaluation
(6) EFL course design innovations (in contrast with ESL)
(7) defining L2 Maturity as a construct for Applied Linguistics
(8) two practical tenet lists for teachers and coursebook writers

2. Literature Review Chapter: Syllabus Design

This chapter will provide a critical literature review of L2 syllabus designs. It will cover significant designs from a historical perspective and discuss their significance within the realm of L2 classroom teaching and methodology. While running through the points, it will also be looking toward the future by uncovering candidates for incorporation into a new, dynamically designed syllabus. (Note: Much of Chapter 2 has been previously published in the Kiban Center Kiyo Journal, University of Kitakyushu, and as part of a chapter in *Issues in Syllabus Design*.)

2.1 L2 Syllabus Design

What is the purpose of a syllabus for the L2 classroom? Robinson (2011:294) describes syllabus design as the compilation of decisions regarding *units* and their *sequence* of performance. At the lowest level, the syllabus may merely be a reminder or a list of things to do for the busy teacher who has little classroom planning time. However, a good syllabus does much more than that. Upon examination, a well-written syllabus will likely provide a doorway into the pedagogical beliefs of the course writer. For example, the difference between a grammatical syllabus and a task-based syllabus should be recognizable, even though modern syllabi seem to take more of an eclectic approach than ever before.

In many contexts, the syllabus becomes synonymous with the chosen coursebook (or its Table of Contents). So, in light of this, what could/should ELT coursebooks and materials do to facilitate language acquisition? Below is a modified version of a list from Tomlinson (2008:6).

- Provide rich experience with varying genres and text types
- Provide aesthetically positive experience through graphical design
- Use of multimedia for rich and varied language learning
- Help learners make discoveries for themselves
- Help learners become autonomous language learners
- Providing supplementary materials for extensive listening and reading
- Help learners personalize and localize their language learning

How could/should these points inform the design of syllabi? In this section, I will consider the above points while examining several types of syllabi. I will discuss approaches regarding how materials can be evaluated, and how they can be adapted/ designed to meet the needs of the learners and the teaching context.

2.2 The syllabus purpose

Let us begin this subsection with a brief discussion on *Syllabus* vs. *Curriculum*. Is there a difference? Richards (2001:2) distinguishes a difference in scope. Historically in the UK, syllabus design fits *within* the curriculum design; the curriculum is the bigger picture. A curriculum can incorporate three steps within the bigger picture of L2 teaching: (1) course planning, (2) materials/methods, and (3) course evaluation (Nunan, 1988:4-5).

Traditionally, because grammar teaching was the long-standing norm, syllabus design was just the order of grammar that would be taught within the framework of the bigger picture or the curriculum. However, due to the rise of the more modern syllabi, *needs analysis* and other types of meta-analyses have forced a change of the meaning and importance of the syllabus (Tagg & Woodward, 2011), therefore making the lines between the two blurry. Moreover, this particular discussion is focused on the British usage; in the United States, the two terms are often considered synonymous. This dissertation will implement the British nuance unless stated otherwise.

Armed with a better understanding of the syllabus, what purposes of the

syllabus are conceivable? From Hutchinson and Waters (1987:83-84):

- to break language down into manageable units and provide a practical basis for textbooks and instructional blocks
- to thus provide teachers and learners with moral support
- to reassure students and/or sponsors that a course has been well planned: its cosmetic role
- to give both students and teachers an idea of where the course is going
- to act as an implicit statement of the views held by the course designers regarding language and language learning—telling students not only *what* they are to learn but *why*
- to guide the selection of materials, texts, and exercises
- to ensure an element of uniformity across a school or educational system
- to assess how successful a student has been during a course by providing a basis for testing

Perhaps more importantly, and besides the proposed purposes, what should the syllabus (and/or materials) being used achieve? One point Tomlinson (2011) brings up is *impact*. However, he admits that impact in Brazil may not equate to impact in Austria. But even then, we should embrace his concept. In general strokes, how can impact be achieved? Tomlinson provides a useful list that can help benchmark assessment of syllabus and material design (2011:8):

(1) Novelty
(2) Variety
(3) Attractive presentation
(4) Appealing content
(5) Achievable challenge

2.3 Is the coursebook a syllabus?

The coursebook (or its table of contents) does not have to be, and in some cases, should *not* be, used as a real syllabus design. However, for practical purposes, the coursebook (or the coursebook's table of contents) is often used as a substitute for a specially designed syllabus. Why might this be a problem? Although there are plenty of useful lists published to help in the assessment of syllabus design, as a textbook writer, I have become aware that textbook publishers and their editors often put a priority on designs that will sell the biggest numbers, and not always prioritize what is best for the students. Financial success on the global scale force editors and writers to publish for the global scale and seldom for the needs of specific classrooms, let alone specific students. This is a very inconvenient truth, also acknowledged by Tomlinson (2003, 2008). There are several other obstacles for writers:

> *Materials writers are faced with a number of competing demands, then. They need to make their materials suitable for a wide variety of teachers, who have different amounts of experience, are more or less qualified, and who may have different teaching styles and beliefs. As Johnson et al.'s (2008) study shows, experienced and inexperienced teachers may be looking for different things from materials.* (Harwood, 2011:14)

In the above quote, Harwood discusses the difficulties and eludes to the compromises that are inherent in coursebook writing. With such compromises in place, should the coursebook table of contents be used as the official syllabus of the course? Harwood continues (2011:14-15):

> *Textbook writers need to make their materials easy enough to follow for the inexperienced teacher by, for instance, making activities build upon one another in a transparent and predictable sequence, or by providing detailed teachers' notes, while at the*

> *same time ensuring the materials are flexible enough for the more experienced teacher to adapt in any number of ways. In the words of a textbook editor, then, authors are trying 'to please all the people all the time'* (Young 1990: 77; see also Mares 2003).

From experience, I can say that it is a relatively standard procedure for university EFL instructors in Japan to use the table of contents from their chosen textbook as a quick template for their official syllabus unless their syllabus has already been mandated from above. Perhaps this is a common practice across the globe; Sinclair and Renouf (1988) see apparent reasons for this phenomenon such as: (1) convenience, and (2) lack of supporting documents. If tests are based on the textbooks, perhaps all the more reason to adhere to the textbook's laid down path. Perhaps the best solution then is to base the syllabus on the coursebook's table of contents, but assume that making adaptations should be the norm. That being said, a good syllabus should at least hint at assessment protocols, if not precisely define them. Most mainstream coursebooks do not impose student assessment protocols, nor do they necessarily mandate a particular methodology. For example, the so-called communicative coursebooks often leave the specific methodology up to the teacher (Whong, 2011), presumably because many of them are a mishmash anyway. Because of these points, the coursebook (or its table of contents) should not be equated with the syllabus, although why this is done on the global scale is understandable.

2.4 Syllabus Analysis

Richards and Rodgers (2001:20-34) discuss their tiered framework for evaluating methodologies. The three tiers are *Approach* (the underlying theories), *Design* (selection of content), and *Procedure* (specifics of the activities). This has become a somewhat standard tiering framework for TEFL/TESL courses. It is a method that forces the user to consider the development and the connectedness across the three tiers. For novice teachers and novice course writers, it forces

contemplation regarding theoretical approaches. As some teachers inevitably like to "play it by ear" and don't care to think about pedagogical theory, an adaptation of Richards and Rogers' framework (Figure 1) may help teachers and writers to sharpen their awareness of theory, and force them to examine how theory relates to the design of courses, the procedures within the classroom, and how it ultimately affects theory by either fortifying it or editing it.

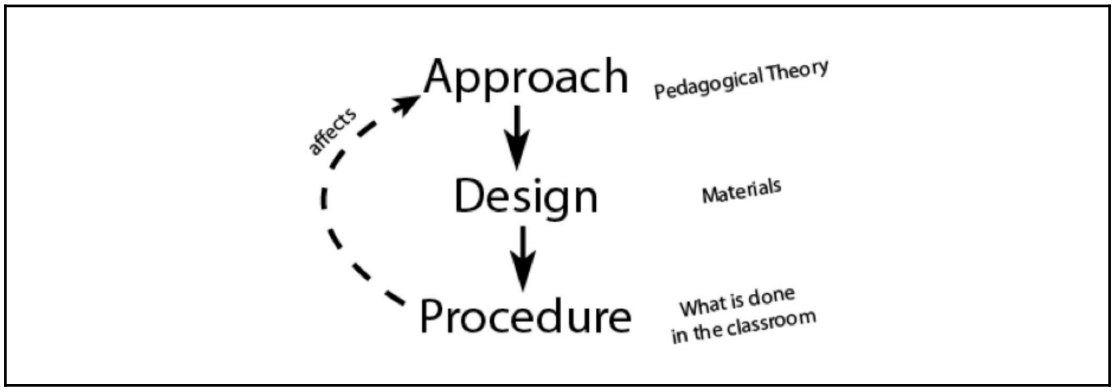


Figure 1: The dynamic effects of the Approach-Design-Procedure framework.

2.4.1 Product vs. Process

There is a significant syllabi dichotomy: *product* vs. *process*. The product syllabus focuses on what linguistic content is to be learned. It is obvious and formal — such as a list of grammar points, or vocabulary words. This may be suited for standardized tests such as the TOEIC where everything is clearly prescribed and relatively transparent. The product syllabus and the standardized test have a *reciprocal* relationship; they tend to inform each other and can create their own microcosm (Figure 2). This is a potential hazard. Since they can (and often do) exist within their own universe, they can exist without real-world applicability. This is not to say that they are not without merit, but the potential of them turning into an existence for their own sake is also very real. This type of syllabus can also be called a content-based syllabus, or colloquially→ *teaching to the test*.

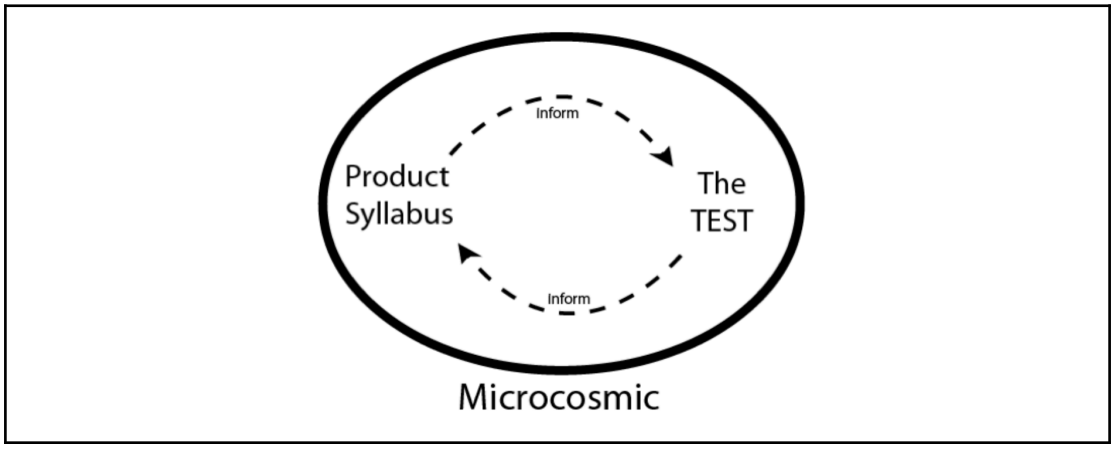


Figure 2: The inherent microcosmic nature of the product syllabus and test cycle.

The opposing type of syllabus is the process syllabus; it does not work off a list of words of bullet points, but a set of learning processes. In short, it defines the skills that are to be acquired, not the content. Nunan (1988:40) discusses how the product syllabus and the process syllabus are incompatible. As defined, the latter creates open-ended learning situations the hopes to build skills for the real world, while the former puts a sharp focus on very particular content—often disregarding the attainment of the skills that will be required for real-world application (White, 1988). However, it must be realized that although this argument is seemingly logical, it is also entirely possible to incorporate both approaches into lesson plans. There is nothing to stop a teacher or course writer from flip-flopping between the two approaches per week, per lesson, or even per activity. Indeed, I am sure some teachers do flip- flop between these two approaches without even realizing it (as I have done so myself).

2.4.2 Synthetic vs. Analytic

Another significant syllabi dichotomy is: *synthetic* vs. *analytic*. Ellis says, (2012: 342), "The former involves a structural approach to teaching that has as its goal the creation of 'form and accuracy contexts,' while the latter involves a task-based approach that seeks to create 'meaning-and-fluency contexts.'"According to Wilkins (1976), the synthetic syllabus assumes that the teaching of discrete bits of the language will cause a gradual accumulation of parts that will eventually create a

whole structural framework for which the L2 will reside. It assumes that language is like a set of building blocks that can be layered down for a foundation and then built into meaningful structures. Such a syllabus may work well with grammar-based instruction.

What about the analytic syllabus? Robinson (2011:306) assures that due to developments in cognitive research and various other factors, "the shift from synthetic to analytic approaches…. can be expected to continue." The analytic syllabus as Wilkins' describes it, is a list of purposes for the L2 learning and the means to meet those purposes. Identify the needs and then satisfy them with the appropriate language usage. It is a more social and perhaps gregariously reciprocal teaching approach. It is real-world and student-centered. This means that grammar teaching often takes second chair at best, but does not mean that grammar should be dropped entirely, so it is different from Krashen's no-grammar mandates (Krashen & Terrell, 1983).

In summation, which is more appropriate for the L2 classroom, synthetic or analytical? To this question, Ellis recently responded in this way:

> *There is now plenty of evidence to show that both approaches can contribute to learning… it will have to be filtered through the teacher's personal understanding of the instructional context, and this, to a large extent, will depend on experience.* (2012: 342-343)

It would seem that although analytical syllabi are in vogue, elements of a synthetic syllabus can, and perhaps should, be incorporated by the syllabus writer, or engaged as an *adaptation* by the classroom teacher.

2.4.3 Linear vs. Cyclical

The third syllabi dichotomy is: *Linear* vs. *Cyclical*. Is language learning more like a truck that is being loaded with parcels, or is it more organic in nature? The linear syllabus assumes the former or at least assumes the former is easier to

understand on paper. Much like with the previous dichotomies discussed, the linear syllabus is a syllabus that attempts to add new blocks of information to a cumulative mass, creating a greater whole. The more blocks added in succession, the more learning. The problem is, language learning is indeed not linear (Dörnyei, 2008:41). We are not computers; we cannot keep stuffing our hard-drives with new linguistic information and expect it all to be retrievable on demand. Research tells that language acquisition is much more of an organic, natural process (Skehan, 1996: 18-19) that would benefit from a cyclical, or recycling, syllabus. Furthermore, *motivation* to learn the L2 is also not linear, but dynamic (Dörnyei, 2011:11) -- meaning that it "organically" changes with contextual and biological changes. . Although this better choice is plainly apparent, this dichotomy is still very relevant because few mainstream coursebooks (and therefore syllabi) are written *cyclically*. It may be assumed that because teachers know that recycling is important, it is overly redundant to incorporate recycling into the syllabus (because the teachers may do the recycling without being prompted to do so).

How much *should* we assume of the teachers? As discussed earlier, some teachers just use their coursebook's table of contents as the template for their syllabus out of convenience. In other words, we know that some teachers are lazy — should we expect teachers to know about—and implement—the recycling of content without explicit prompting from the coursebook? Alternatively, although it may seem redundant and perhaps a waste of paper, should we produce syllabi and coursebooks that are visibly distinguishable as being cyclical in design? From observations made at professional development workshops, my candid response to this is that cyclical features *should* be incorporated into most syllabus designs.

2.4.4 Synthesis: Syllabus Analysis

In this section, I examined the purpose of the syllabus. I also identified several syllabi dichotomies. Of particular interest, among the three dichotomies introduced, I identify most with the latter half of each of the dichotomies, that is, I identify most with the Process, Analytic, and Cyclical syllabus types. When cross-

referenced in this way, I have come to realize that my own teaching and coursebook writing is indeed process-based, analytical, and cyclical *and* that I recommend these styles. Although, I agree there needs to be a *focus on form* and a *focus on the product* at least some of the time. This will be further developed later in this thesis.

2.5 Syllabus Design Factors

This next section will focus on several factors of syllabus design and potential assessment schemes. It will then move on to ideas for the future - dynamic designs and motivational material development.

2.5.1 Design Factors

What factors influence the design of a syllabus? Below is a general list adapted from Tagg and Woodward (2011):

(a) Common practice/trends
(b) Theories of Second Language Acquisition/pedagogy
(c) Wider educational context
(d) Course Objectives
(e) Learner backgrounds

Of these, mainstream coursebooks would connect most directly with A, B, and C. Less mainstream coursebooks may zero in on D, such as coursebooks designed for courses such as *English for Academic Purposes*, or *English for Nursing*. Oddly, this leaves E alone and perhaps left for the teacher to deal with, or to put more bluntly, up to chance. Few teachers, much fewer coursebook writers, know specific details about their students before designing their syllabi, so it is indeed often left up to chance. From a student's point of view, I do not think it is fair to leave what I would feel is one of the most critical factors on this list up to chance, though. What can be done about this?

Below (Figure 3) I have designed an assessment sheet of essential design factors for consideration during syllabus design. I believe this assessment sheet can be useful during the pre-design stages of syllabus and materials writing, and also as an in-use and post- use assessment plan. It can be applied per chapter for single chapter assessment; all chapter's assessments can then be strung together to form the basis of a comprehensive syllabus assessment. Alternatively, it can be used more holistically by applying one assessment sheet to assess the entire syllabus.

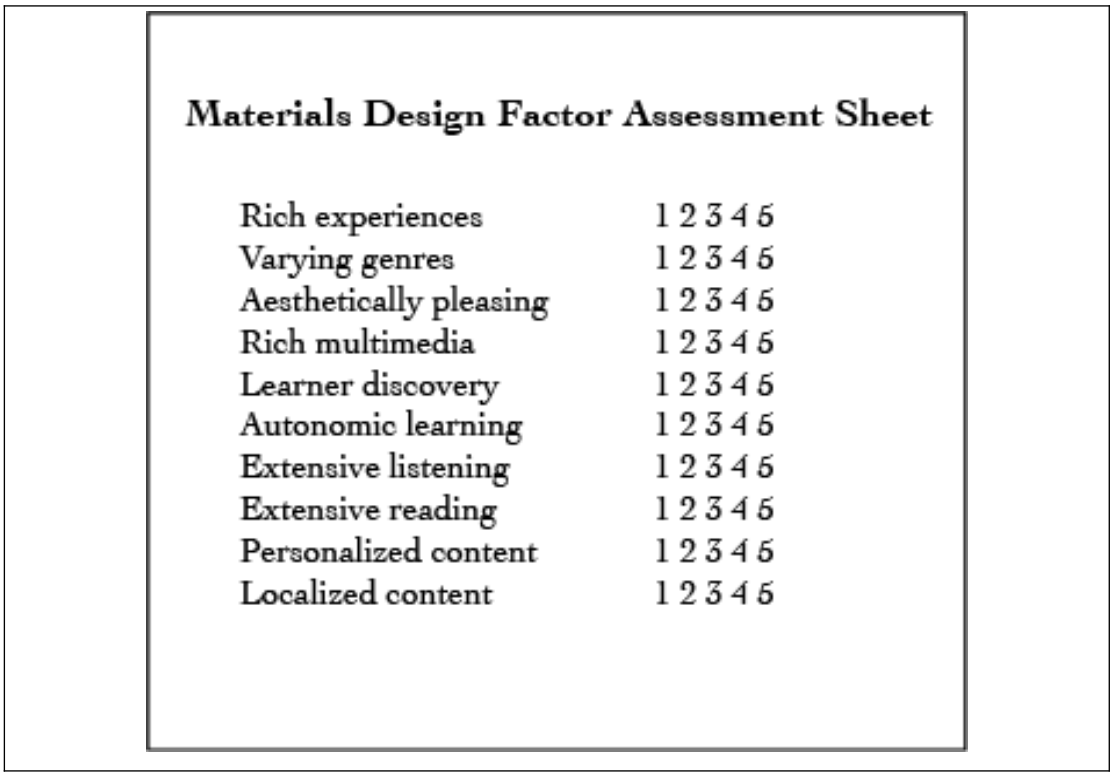


Figure 3: The Materials Design Factor Assessment Sheet

2.5.2 Toward a Dynamic Design

Referring to a more typical approach, Richards (2001:145) provides possible steps that can be taken during the process of syllabus design. They are as follows:

- Developing a course rationale
- Describing entry and exit levels
- Choosing course content
- Sequencing course content
- Planning course content (frameworks and instructional blocks)
- Preparing scope and sequence

These are called steps, but they should probably be viewed holistically, as simplistically going from top to bottom would be irresponsible. In fact, it may be beneficial to view them dynamically. Graves (2000), perhaps with similar intentions proposes the following framework (Figure 4):

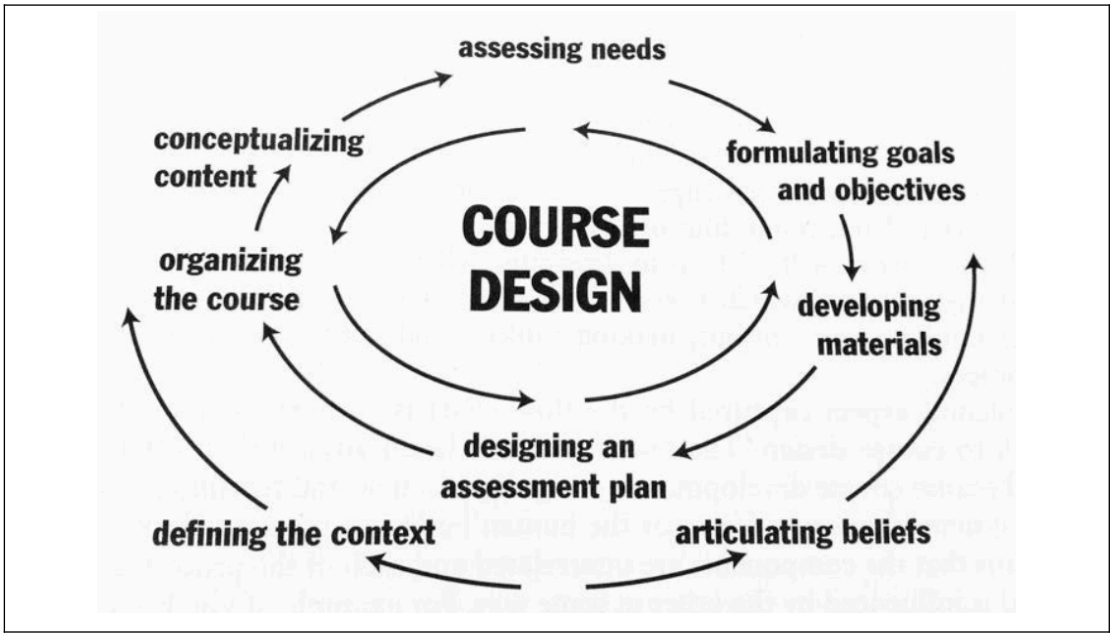


Figure 4: Graves (2000): Framework for Course Development

After consideration of Graves' intricate model, I found it unnecessarily complex, although it may be helpful for some designers. Preferring Richards' list, but

desiring a more dynamic relationship between the points, I propose the following model (Figure 5). All five points must inform each of the other points and form a dynamic relationship, and should, therefore, be viewed holistically. Although it combines elements from both, I find it to be a more elegant solution than Graves' model, and more realistic than Richards' list.

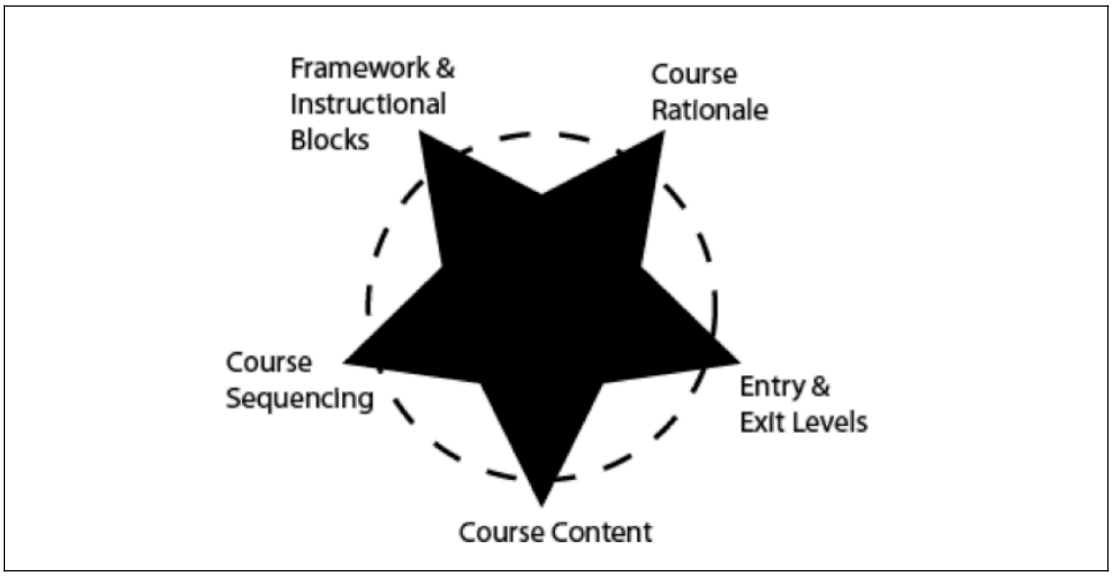


Figure 5: Dynamic syllabus design, five-point and cyclical

A good starting point for writers could be the *course rationale*, but as can be seen in my proposed model above, the course rationale should also be affected and edited in the middle and ultimately even at the very end of the process, influenced by the other four points. Richards suggests that an initializing rationale write up can be a guide during the writing process (2001:146). It can be as little as a few paragraphs discussing the *Who*'s and *What*'s of the course and will ultimately serve as the justification of the course, perhaps touching on the design philosophy as well.

The *entry and exit levels* cap the design nicely when appropriately set. Historically, mainstream coursebooks typically broke student levels into three (beginner, intermediate, high) and then five groups. Oxford University Press (OUP) currently boasts nine levels (Figure 6). Perhaps this diversification, which is not only an OUP trend, is an indication that learner-centric publications are becoming slightly

more popular. It could also mean that publishers are clueless at what they should be doing in the quickly changing digital device age, and are merely throwing out titles to see which ones sell. Personal discussions with several publisher representatives fortify the latter possibility.

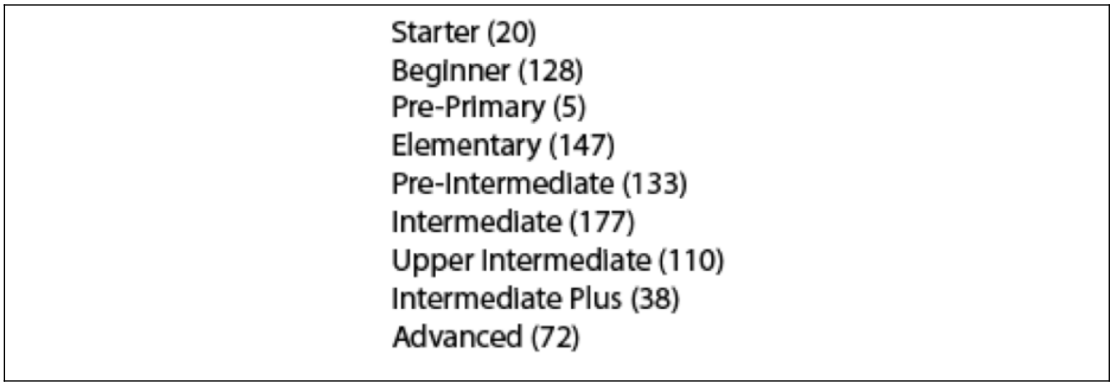


Figure 6: OUP's levels and number of titles in brackets (accessed 2015)

Some teacher-researchers see the above list as being still too arbitrary and prefer to divide the learner population by standardized test scores, probably to assist with their research by making clearer (quantitative) distinctions. Indeed, to fill that gap, several publishers have coursebooks designed for teaching TOEFL, TOEIC, and IELTS, among others. These books line the shelves of bookstores in Japan. While a focus on test preparation books seems fantastically useful (especially for the teacher-researcher), it is also a dangerous step toward *teaching to the test,* as discussed earlier. Also, as discussed in the Syllabus Analysis section, the combination of the test and coursebook have the potential of becoming their own *raison d'être* amalgamation, with less regard to the real world than there should be. The Common European Framework of Reference for Languages(CEFR), used to assess language learning, is also blossoming in this area as a new standard.

How will CEFR play out in the long term? It is interesting to note that L2 teacher-researchers in Japan have also taken an interest in CEFR (see O'Dwyer et al., 2011) although a deep review of Japan-based CEFR literature may still be premature.

What about the *framework*? As discussed in the Syllabus Analysis section, traditional frameworks used to be not much more than a list of grammar points that

were sequenced "logically," but were arbitrary in the real-life sense, especially from the cognitive point of view. Human cognitive development is much more organic and contextually based (Sousa, 2011). What other types of syllabi are there to choose from? Here is a typical list:

- Grammatical
- Notional-functional
- Lexical
- Situational
- Topical
- Skills-based
- Task-based
- Context-based

The above is a typical list that can be found in TEFL/TESL courses. I shall discuss five of them in depth: *grammatical, notional-functional, lexical, task-based,* and *context-based*. Many coursebooks and syllabi adhere to only one of these designs. However, nothing is stopping a writer/designer from creating an eclectic mixture of ideas for a more diverse framework, as is often the case with 'communicative' coursebooks (Whong, 2011). Such a mixture is sometimes called an *integrated* framework (integrated syllabus). Indeed, an eclectic mixture within a framework would be more like real-life learning which is commonly random anyway. Such an eclectic framework could, therefore, be more in tune with actual cognitive development — although a totally arbitrary mixture would (naturally) be hit and miss. The idea for a new type of coursebook framework that is much less hit and miss will be proposed in Chapter 9.

2.6 Instructional blocks

Once the general design or framework begins to take shape, how are the smaller *instructional blocks* to be designed? An instructional block should be somewhat self-contained, yet contribute to the broader objectives of the syllabus (Richards, 2001). To this, Tagg and Woodward (2011) provide the following list of aims:

- To make the course more teachable and learnable
- To provide progression in level of difficulty
- To create overall coherence

The typical syllabus (perhaps based upon a coursebook) would have individual chapters, units, or modules that would be somewhat self-contained and individually assessable. Within each chapter, unit, or module, there may be several subsections that further break down the learning into smaller chunks with perhaps a variety of activities designed to advance the students toward the ultimate goal of the chapter/unit/module. Richards (2001:166) furthers: the coherence of the chapter/unit/module should ultimately be both *horizontal* and *vertical,* meaning that there should be front to back coherence throughout the chapter/unit/module, but also, if it were a book, top to bottom coherence on each page (Figure 7). Modern EFL textbooks do seem to pay close attention to appealing vertical design for an appealing *sense of closure* on each page.

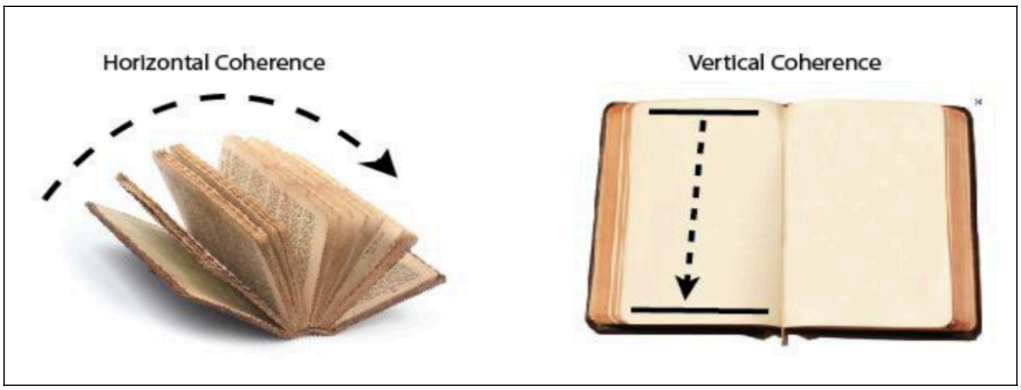


Figure 7: Richards' horizontal and vertical coherence models.

While several mainstream coursebooks certainly adhere to this maxim, there are still coursebooks—especially books based on reading skills, where long passages are the norm— that ignore this maxim pragmatically. It seems that sometimes content characteristics must occasionally override design protocols.

2.6.1 Motivation in material development

One concept that is sometimes tragically overlooked is motivation. Tomlinson (2008, 2011) is one researcher who commonly incorporates motivation into his discussion of material development. Dörnyei has done extensive work in motivational strategies and underlying factors (2001, 2005, 2009, 2011). Surely such a central concept should be applicable and perhaps deserves to be at the core of syllabus design. However, as Dörnyei stresses, "not every strategy works in every context!" (2001:30). Moreover, student motivation is centralized around each student and their own learning contexts, inclusive of their home situations and their own teacher's motivational techniques or lack thereof. Motivation is extremely complex. Therefore, it may be overly ambitious to believe that the incorporation of motivational features into syllabus design will be successful, or fully assessable for that matter. That said, such ideas seem to be perfect for a newer dynamic syllabus design. Below (Figure 8) is an adapted diagram of Dörnyei's components of motivational teaching practice (2001:29).

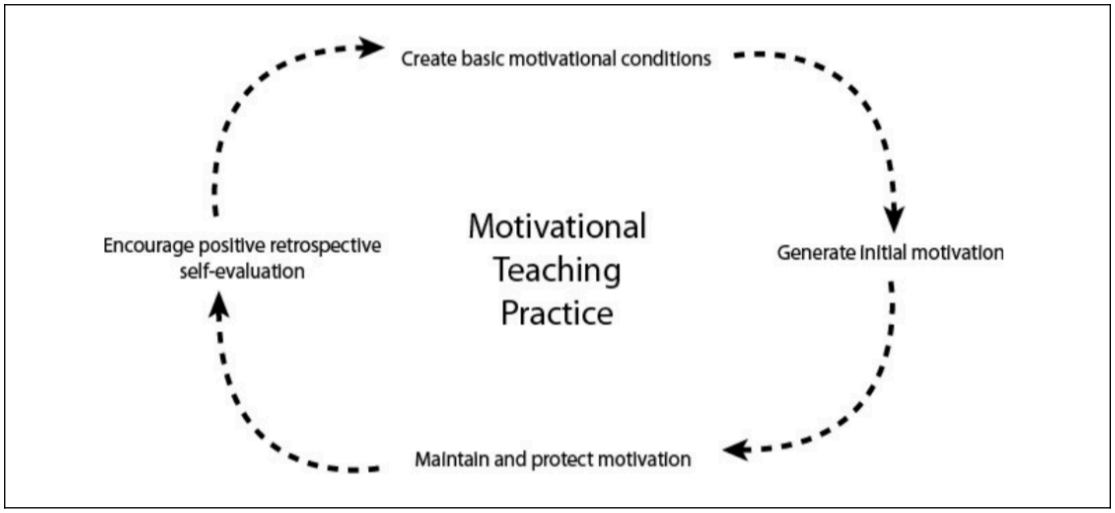


Figure 8: Components of motivational teaching practice, adapted from Dörnyei, 2001.

Being cyclical in nature, the motivational teaching practice model seems to be fit for implementation (with modification) into cyclical syllabi. Perhaps the entire cycle can be implemented within a syllabus on a per-unit-of-study basis. Additionally, Dörnyei (2005) introduced the L2 Motivational Self System:

- Ideal L2 self
- Ought-to L2 self
- L2 Learning Experience

It defined a new way to conceptualize and research motivation. It is introspective, or meta-analytic. Perhaps these reflexive components can also be candidates for a dynamically designed syllabus; they can be incorporated into a syllabus as reflexive questions for learners. It should prove to be useful for low to high-level students.

A prime example of this: I have a few lower level university classes with students who study English (L2) because they are required to do so. They have been required to do so since the age of twelve. Syllabus-level metacognitive activities involving the *ideal L2 self* and the *ought-to L2 self* could be very useful for such students because many of them probably have never had a reason to develop concrete thoughts in these areas, hence a potential lack of motivation in these areas. This may be what is separating them from the higher-level students.

2.6.2 Synthesis: Syllabus Design Factors

In this section, I examined the purpose of the design factors that go into the creation of a syllabus and identified several points or steps that should be taken on the way to designing a syllabus. I also introduced candidates for dynamic design and discussed my own ideas toward dynamic models, inclusive of motivational theory and components. As discussed earlier, these ideas will be carried over to *Conclusion: Looking toward the future.* The next section will critically analyze four mainstream

syllabus types.

2.7 Some Common Syllabus types

This section will take a look at five common language teaching syllabus designs: *grammatical, notional-functional, lexical, task-based,* and *content-based*. Then, I will discuss one syllabus type from Education → the Spiral Curriculum. (Note: here curriculum is used synonymously with syllabus.) I will briefly discuss each of the syllabus histories and their unique significances. I will also critically analyze their designs and recognize their potential.

2.7.1 Grammatical Syllabus

This subsection shall focus on the grammatical syllabus. First, we must define what grammar is because there are different opinions regarding its definition. Cook (2001:20) offers the following list:

- Prescriptive grammar
- Traditional grammar
- Structural grammar
- Grammar as knowledge in the mind

Prescriptive grammar: As the title implies, this prescribes or attempts to police language usage. It prescribes how language should be spoken and written in a very formal way. However, modern cognitive researchers understand how dynamic and evolving language necessarily is, therefore prescriptive grammar can and perhaps should be considered *old school.* That said, it is far from dead, as it is still alive in foreign language teaching (Tagg and Woodward, 2011).

Traditional grammar: This type of grammar examines the usage of and the nature of the various parts of speech. It is commonly taught to L1 learners between upper elementary through to junior high school, usually after gaining native level

communicative competence via several years of real communication in the L1.

Structural grammar: This is the type of grammar associated most with substitution tables. It examines the structural build up of sentences, from word-phrases to the complete sentence.

Grammar as knowledge in the mind: This is the type of grammar that most non-language teacher L1 users implicitly have; they can produce grammatical utterances without conscious effort, nor awareness of the mechanics employed. Before the invention of formal schooling, the only grammar possible would have been type 4, *Grammar as knowledge in the mind*. The advent of formal pedagogy seems to have given birth to the other three types of grammar. It should be of no surprise then that grammatical syllabi commonly only engages the first three types, albeit, deduced from type 4 at some point in history.

The grammatical syllabus, historically, has long been the standard in language teaching. The historical roots of the grammatical syllabus are typically known to have come from the study of *Latin*. However, there is an interesting history regarding its refinement in the early to mid-1900s.

According to White (1988), Palmer, then Hornby and West set out to create more of a rational progression to the syllabus — from simple to difficult, or, known to unknown. Hornby published his *Guide to Patterns and Usage in English* (1954) with pedagogy based on sequenced language and a graded syllabus. Palmer, and also West, took to vocabulary and laid the groundwork for corpus linguistics; West published *A General Service List of English Words* in 1953. Their idea was, by identifying key vocabulary and studying how they were used— reading skills should improve. Such ideas no doubt also lead to the development of the lexical syllabus. Literature from researchers such as Richards (2001) and White (1988) point to the works of Palmer, Hornby, and West as being highly influential and being a standard reference for coursebook design and graded readers for several years.

Although the grammatical syllabus has long been the standard and influential in language teaching, there have been serious reactionary movements against its popularity since the 1970s. The movements became especially strong throughout the

1980s coinciding with the then growing popularity of Krashen's work.

The historical roots of the grammatical syllabus are typically known to have come from the study of *Latin*. However, because Latin was then (and still is a) dead language, real communication in Latin was irrelevant. Typically, only the rote memorization of Latin vocabulary and grammatical points, coupled with translation exercises was sufficient to "learn" the Latin language (Brown, 2001:18). However, according to Stern (1983:453-455), the grammatical syllabus was considered successful in Europe because of the ease of assessment — teachers could present a rule, and then the students would immediately be able to demonstrate usage via practice sentences.

This view now sounds shallow. Why? Contrary to the dead Latin language, English is one of the most "alive" languages on the planet. It surely deserves a lively pedagogy. In the 1970s, a movement began to dethrone grammar-translation from its pedestal. This anti- grammar movement was so strong that its reverberations are still being heard today.

Modern L2 coursebooks that are based on grammar sometimes play down their designs and claim to be *communicative* instead. A prime example of this is the de facto standard in Japanese junior high schools, *New Horizon* (Tokyo Shoseki), being published since 1966, with at least 40% dominance in Japanese junior high schools (FLANG, 2007). The addition of practice dialogues and more listening components of recent years has allowed advertisers to claim their books are now *communicative* in design, while the syllabi remain relatively unchanged — still being based upon grammar and catering to the needs of the grammar-translation teaching style, which is the traditional and predominant teaching style throughout Japan. Pedagogical preferences aside, it is interesting to note that publishers do feel the need to disguise their grammar-based ethnicity, and portray (feign?) a movement away from grammar-based syllabi.

That said, as Nunan puts it (1988:29), a prevailing view is that "language consists of a finite set of rules" and therefore pedagogy must set out to "crack the code" for students. Believers in the view will probably find the grammatical syllabus

to be logical and pragmatic. It is all about: (1) selecting the best items to be acquired and then, (2) helping attach them to preexisting knowledge. Perhaps this also assumes that language learning is linear, which it is not, but theoretically, a grammatical syllabus *could* be conceived within a cyclically-based framework. This is a highly significant point that will be called upon later in the thesis.

One of the typical ways to sum up the grammatical syllabus is that is a *product* based syllabus. That is, there may be very little concern for *how* the learning will happen; the learning process and sometimes even its ultimate usage is often considered irrelevant. This typically makes it a non-learner centered *synthetic* syllabus. The focus would have to be on accuracy, to the detriment of fluency. Ironically, in all its seriousness and typically stringent focus on accuracy, in extreme cases, the language structures need *not* be authentic, nor realistic, as long as they are deemed grammatically correct. One example of this would be the test questions from Japanese public schools that are designed to assess students' knowledge of complex English grammar but to accomplish this task; they use synthetic, unnatural sounding sentences and passages that are recognizably "forced" and clearly non-nativelike. Assessment of English grammar knowledge within this synthetic paradigm would be possible, *per se*, but to what end? What does this assessment really accomplish?

Furthering this discussion, Tagg and Woodward (2011) provide a pragmatic list of criteria, stemming from Richards (2001) and White (1988), for the grading of grammar points to be taught:

- Simplicity and centrality
- Combinability
- Contrast
- Communicative need and immediacy
- Teachability
- Learnability
- Frequency
- Natural order

All of these bullet points end up being subjective and therefore arbitrary, except for perhaps *natural order*, in theory. However, although Krashen (1974) professes natural order, there is little, if any, scientific data that can prove what natural order is. This is due to the contextual nature of real-world learning as discussed earlier; learning is context dependent (Murphy, 2016).

What is the grammar-translation *method*? It is a method of translating text from L2 to L1 and deducing grammatical features (top-down), conducted in the L1. It is very straightforward and needs little (or no) L2 communicative competence from the teacher. In the most significant ways, the grammar-translation method (GTM) is the basis of the grammatical syllabus; the method is indeed derived from the teaching of Latin that "explains why this method was not concerned with developing productive L2 competence in the learners" (Dörnyei, 2009:273). In short, GTM is a simplistic and pragmatic method that is regrettably not "conducive to developing productive language skills."(ibid.) The grammatical syllabus is a systematic extension of this methodology.

Criticisms: The Grammatical syllabus

I have embedded a few of my own criticisms of the grammatical syllabus, above, but there are few more general criticisms about this type of syllabus design from the literature. Skehan (1996) discusses how chosen structures presented by the teacher do not simply lead to language acquisition. Ellis (1997) concurs by saying that teaching may have very little effect on when and how learners may or may not learn a structure because of the internal circumstances per student. Widdowson (1988) argues that because grammar teaching tends to dissociate the structural learning from actual contexts, the true meaning can become lost. This, of course, questions the very purpose of teaching and learning. Indeed, Lightbown and Spada (1999) bring to light studies that question the effectiveness of grammar-based syllabi, stating that high focus on accuracy *did not* necessarily lead to higher proficiency. Moreover, those learners were *unwilling* to take risks in communicating.

The results of this phenomenon are readily visible in daily life in Japan, and therefore this may be one of the most persuasive arguments against using a grammar-based syllabus and teaching with grammar-translation methods in Japan.

Addressing these criticisms, as we will see in the next subsections, several 'anti- grammatical' conceptions surface. However, after pendulum swings an interesting halfway point did emerge (Dörnyei, 2009). There are several incarnations with confusingly similar name structures, but I shall simply discuss them collectively as a *focus on form* (FoF) here. The premise was that although a strict grammatical syllabus was proven to be inadequate for most purposes, the full dropping of grammar teaching was equally inadequate (Doughty and Williams, 1998; Dörnyei, 2009; Whong, 2011).

FoF cleverly hid the word *grammar*, while having learners focus on sub-sentence grammatical issues. Sometimes these issues were based on real-time needs; sometimes they were based on *preemptive* or *proactive* decisions by the teacher (or course writer). In reality, unless the teacher is focusing on form with real-time mistakes of the learners in the classroom, all the preemptive/proactive planning comes dangerously close to being a younger sibling in the grammatical family. It is not my intention to shoot down FoF entirely. In fact, I fully agree with the premise. As long as FoF is grounded in real-time student needs, it is a very welcome concept, but I am hesitant to endorse the variants of FoF because they share the same criticisms as the traditional grammatical syllabus.

Advantages: The Grammatical syllabus

Are there advantages to the grammar-based syllabus? It is easy to follow and easy to assess. Such a syllabus can be very transparent, so by going through the steps, there may be a strong feeling of achievement. Progress can be easy to measure.

In short, such a syllabus is easy to manage because of its pragmatic approach. It is also probably familiar to most teachers from their own L1 learning, so it would be relatively easy to implement for novice teachers. Moreover, while it may not

make every learner proficient, inevitably, some students *will* and *do* have a strong aptitude with this methodology (likely due to a serendipitous combination of neurobiological and contextual facets conducive of grammar- based learning proficiency—that they happen to have at the time) and therefore some of them will build a natural preference for this methodology. These, for lack of a better term, "lucky" grammar-based learners do excel within this framework.

Advantage or disadvantage?

According to Dörnyei (2009:273), the Grammar-Translation method (and/or the grammatical syllabus) may still be preferred by some because it:

- requires little setup time
- requires minimal teacher L2 competence and fluency
- is safe for the teacher
- is easy to assess (with multiple choice tests, for example)
- is easy to explicitly focus on (discreet) points
- can be mostly taught in the L1
- requires no higher-level (meta-cognitive) thinking/assessment

From personal discussions at professional development workshops in Japan, I am aware of several Japanese teachers of English that became English teachers *because* of their aptitude and affinity with the grammar-based syllabus and the grammar-translation methodology. So, from their perspective, the advantages are obvious — *it suited them well.* However, they may be armed with the attitude, "This pedagogy worked well (for me), so why can't you understand this? This should be easy for you." In this way, a *grammar-based teaching preference cycle* seems to be self-perpetuating because the next generation of teachers are self-selected from a pool of the best "grammar-based students." This is how this self-perpetuating cycle lives on.

Is the longevity of this cycle advantageous? It may become a significant

advantage for a minority of the population, while likely being a disadvantage for the majority of the population. For the majority of the Japanese English teachers, however, it is simply viewed as *the traditional* way, and, from their culturally prominent neo-Confucian perspective, it is not something to be opening critically challenged because it was passed down from above. (This is not to say there is no critique occurring in Japan. There is critique, but change has been superficial and slow.)

Synthesis: The Grammatical syllabus

In this subsection, I examined the grammatical syllabus. I discussed its history, rooted in the study of Latin, and how it was refined in the early to mid-1900s. Although it has advantages as being transparent and pragmatic, there are several criticisms of this type of syllabus design. Perhaps the most important is that although it sets out to be a no-nonsense, pragmatic syllabus, it is nevertheless synthetic and moreover, does not often enough result in higher proficiency than other design types. It takes learners "through the motions," and may make learners *feel* like they are studying hard and learning—which may or may not raise their motivation—but by the end of the day, it would seem that much of that work goes to waste.

I find that the biggest faults in the grammatical syllabus lie within the inherent *assumptions* that go with it:

- L2 learners need to explicitly learn grammar
- L2 grammar teaching leads to grammatical L2 usage
- L2 grammar has logical rules that are definable, teachable, and learnable
- all L2 learners can learn L2 grammar at the same pace
- (from proficient teachers) --*"I learned it this way, why can't you!?"*

These inherent assumptions (that may have been nurtured by the long-term historical adherence to grammatical syllabus design), are assumptions that I would

define as *old school.* They are traditional assumptions that have no significant neuroscientific basis either. Ironically, they can still be found in courses that have "moved away from" using a grammatical syllabus. This is presumably because these old-school assumptions have been carried over to other forms of L2 teaching. This phenomenon shall be further discussed in later sections.

The take-home ideas for the new syllabus:

It has become clear that a strict grammar-based syllabus should not be pursued, but the lighter version known as FoF is promising, mostly because it was developed as researchers began to re-realize the need to focus on grammar, but decided to balance the grammar instruction more sensibly. The take-home point from this subsection is: *include a need-based FoF component, preferable zeroing in on real-time issues and mistakes (not preemptive guesswork).*

2.7.2 Notional-functional Syllabus

The notional-functional (NF) syllabus developed in Europe as a reaction toward the grammatical syllabus, although it ultimately, and ironically, gathered many of the same criticisms. What is the NF syllabus? Wilkins' (1976:18) in a definitive statement says:

> *(I)t takes the desired communicative capacity as the starting point. In drawing up a notional syllabus, instead of asking how speakers of the language express themselves or when and where they use the language, we ask what it is they communicate through language. We are then able to organize language teaching in terms of the content rather than the form of the language. For this reason, the resulting syllabus is called the notional syllabus.*

The 1970s showed a shift toward a more communicative approach with more of a focus on context and therefore the social aspects of language usage. Key points from Wilkins (1976) are that: (1) the NF syllabus is a communicative syllabus, (2) it raises motivation because it is communicative, and (3) it covers "all kinds of language functions" (ibid:19).

Let us focus on Wilkins' first point. Is the NF syllabus really a communicative syllabus? Widdowson argues that there is no such thing as a communicative syllabus (1990:130); he categorically disqualifies NF. What *is* the NF syllabus then? To help solve this mystery, let's examine the words separately.

White (1988:75) describes *notions* as categories that describe the intentions of the language usage. Nunan (1988:35) describes them as conceptual meanings via language usages such as objects or relationships. Examples would be time, ownership, direction, frequency, and cause (Nunan, 1988:35). The *functions* are then the communicative purposes of the language in use. Examples would be: approving, persuading, suggesting, and identifying (Nunan, 1988:35). This type of syllabus is a logical step up from the grammatical syllabus, yet it does not attempt to bury

grammar. It merely shifts the priority to more meaningful usage (to notions and functions), with the importance of grammar teaching primarily remaining intact. However, this was undoubtedly a step away from reliance on grammar as the guide for what is to be taught. In this way, it embodied natural language usage more realistically (Whong, 2011).

Although at first glance the NF syllabus may appear to be a process-based syllabus, it still is very much a product-based syllabus, much like the grammar-based syllabus. How can this be? Upon closer examination, although the NF syllabus indeed looks as if it is nuanced to be a process-based syllabus, ultimately it is still a to-do list telling us *what* is to be learned, but, with no defining teaching guidelines (the *way* of the learning) to go with it.

For the above reasons, the NF syllabus should not be assumed to be a process-based syllabus. Additionally, it is also a synthetic syllabus by design. It puts forth a list of to-dos, perhaps feigning a sincere level of catering to learner needs derived from discourse analysis research (Stern, 1992), but ultimately the NF syllabus moves on regardless of the learners' real-world needs that may or not have any significant connection to the discourse that had been analyzed nor the results of some needs analysis program.

Good intentions, but…

The NF design began with excellent intentions/observations: (1) isolated utterances and chunks were not appropriate for study, (2) context must be provided for a better understanding, (3) language learning must include practice with the realization of socio-cultural roles (Tagg and Woodward, 2011). So, although it indeed was a to-do list, it was not immediately focused on using the designated "grammar points of the day," but more focused on the naturally occurring discourse per defined situation.

Whereas the grammatical syllabus would tend to have learners attempt to *avoid grammatical errors at all costs,* the NF syllabus would tend to have the learners focus on what would be *contextually appropriate*. Herein lies one of its most

considerable advantages: the advent of role-playing and social interaction activities with notions and functions, *while limiting grammatical absoluteness*.

Upon implementation, it must have become evident that although the grammatical syllabus was simply expected to be for "everyone" and therefore a needs analysis program was not seriously considered, an NF syllabus, based on notions and functions of real contexts, had to be designed to match the needs of the students or else the activities could be widely inappropriate (White, 1988:83-91) — much more so than the grammatical syllabus. However, being new and still naive with needs analysis as a research tool, researcher-writers tended to only focus on student needs, without regard to the bigger picture (Nunan, 1988).

Criticisms: The Notional-functional Syllabus

What are the criticisms of the NF syllabus? As touched on earlier, the NF syllabus is similar to the grammatical syllabus in many ways, so they share many of the same criticisms. Widdowson (1979) argues that although the shift from a list of grammar points to a list of NF points still produces a *list* — these lists are not automatically compatible with real learning. Brumfit (1981) identifies difficulties in defining these 'notions' and goes on to explain the difficulty in applying the learning to real social contexts, seriously attacking the premise of the NF design. Nunan (1988:37) concurs by saying that any division of language into units misrepresents the nature of communication.

In short, the NF syllabus is as synthetic as the grammatical syllabus. The harsh argument being that only the labels have changed; they are essentially the same syllabus, just dressed up differently. In line with Brumfit's argument, Tagg and Woodward (2011) claim that the NF syllabus may even be *more* challenging to learn from than a grammatical syllabus, pointing at the difficulty in generalizing functions and then creating new utterances from them. Because of the inherent dependency for a needs analysis, it can be argued that the NF syllabus can only be as good as the quality of the needs analysis. However, it is not possible to conceive a perfect student needs analysis (for example: will the student respond truthfully?

Will the student respond comprehensively?). So, the inadequacies of any needs analysis trickle down to the NF selections. This brings up other problems as criteria for assessment and sequencing: Which NF list items should come first and why? How can we decide which functions are more complicated? More appropriate? Hence, *the Guinea Pig Argument* (O'Neil,1977).

Guinea Pig Argument?

O'Neil humorously, yet poignantly, argues that no matter how well the NF syllabus may be construed, it would never be able to foresee the need to—and therefore *never be able to*—help a learner produce an utterance such as, *"My guinea pig died with its legs crossed,"* which was an actual recorded authentic utterance from a classroom. The outcomes of NF teaching are not foreseeable because the syllabus does not control them, therefore making comprehensive design and assessment of the outcomes problematic. Furthermore, the lack of protocols may be one reason for its current lack of popularity, especially compared to the grammatical syllabus with its clear focus on accuracy, making assessment not only easy, but, transparent. In this way, although the NF and grammatical syllabi have several similarities, differences such as a lack of focus on accuracy make it more difficult to implement.

Advantages: The Notional-functional Syllabus

What are the advantages of the NF syllabus? Conceived as an improvement over the grammatical syllabus, there are several advantages to it. It can implement more realistic learning tasks based on real-world communication. It provides a means for contextual understanding and realization before a performance because it grasps onto cultural needs (Finocchiaro and Brumfit, 1983:17). But perhaps most importantly, the lack of assessment protocols discussed as a disadvantage on the previous page allows for more flexible teaching, such as cyclical pedagogic practice, and naturally *allows* for the addition of socio-cultural components. Regardless of the actual learning outcomes, it can also be motivational for the students because the

syllabus, or the to-do list, is transparent; students may gain a healthy feeling of completion after each item is crossed off the list.

Synthesis: The Notional-functional Syllabus

Although the NF syllabus was designed to be a step up from the grammatical syllabus, in reality it is: (a) very similar in significant ways, and (b) actually a step down when it comes to assessment policies. On the plus side, it is more humanistic — it allows for the additions of socio-cultural aspects and the realization of contextual influences. At the end of the day however, it is still synthetic and product-based; it is not scientific nor is it logical from a neurobiological growth perspective.

The take-home ideas for the new syllabus:

The transparency of the NF syllabus is a major plus for the learners as it makes the goals clear and therefore potentially motivational. Moreover, there is a sense of *closure* per unit, much more so than from a grammatical syllabus. Therefore, the take-home from this section is, *functional goal transparency* and for motivational purposes, a *strong sense of closure* per unit of study.

2.7.3 Lexical Syllabus

What is the lexical syllabus? It is a syllabus with a keen focus on lexis (vocabulary or chunk) usage. Historically, and as discussed earlier, West's *A General Service List of English Words*, was a milestone for syllabus and textbook writing, listing 2000 frequently occurring words in the English language. West's list became highly influential, becoming the backbone for lexical selection in coursebooks and graded readers. The trend that West began is still going strong today as can be seen from the catalogs from major ELT publishers such as OUP. In other words, West's list paved the road for corpus linguistics (and derivative pedagogy). A modern proponent of the lexical syllabus is D. Willis with *The Lexical Syllabus* (1990), and

along with Willis: *Collins COBILD English Course* (1990).

Lexis, not Vocabulary

Why is it called lexis and not vocabulary? Is there a significant difference between the two? The more common word is of course vocabulary, and it is often taught separately from grammar; grammar and vocabulary are seen as very separate issues. However, when a "word" is treated as having generative functionality (as in, having its own unique grammatical functionalities and idiosyncrasies), then it is upgraded to the term *lexis* (Tagg and Woodward, 2011). It can also be used for chunks or *formulaic language*, where sets of words are used as a singular utterance – often with a singular meaning. It is important to note that while on paper and within traditional grammar, individual words seem to be their own entity with their own purpose, the lexical concept of formulaic language is that small phrases can act together as a unit and are not registered as individual words within our minds. The set is seen as its own entity with its own purpose.

The concept blurs the premise of the traditional grammar vs vocabulary dichotomy and forces us to think and re-think the psycholinguistic and pedagogical implications set forth by these notions. I shall focus on this progressive form of lexical theory and lexical syllabi, and not the simplistic "vocabulary-like" meaning.

The *Generative Lexicon* theory (Pustejovsky: 1998; Pustejovsky et al., 2012), although perhaps not famous in the mainstream, is slowly becoming its own sub-field of study. Pustejovsky theorizes that lexicon functions generatively by:

- providing words for characterizing lexical information
- developing a framework for manipulating distinctions in word descriptions
- formalizing a set of mechanisms for specialized components that function in context, especially in novel situations

The third part is the distinctive part. Far too many theories of language simply ignore the fact that we can combine words uniquely and use words and

grammar in novel ways based on context, without instruction.

Apparently, even young children have the capacity to create formulaic utterances "from thin air", simply based on the context they happen to be in. Berko-Gleason pioneered work in the area half a century ago with her *Wug Test*, showing how children can create grammatical (appropriate sounding) utterances even with the combination of imaginary objects and words and no formal instruction on how to do so. Yet the results of her test have been used to substantiate both sides of long-standing arguments (Berko-Gleason and Murphy, 2011:2):

> *What followed, and not from me, were claims from some that this 'proved' Chomsky's theory that children have an innate language acquisition device. On the other hand, connectionists used these findings to try to model how a child, or a machine, could learn the system through exposure to appropriate examples. Within the linguistic world, a whole controversy arose about whether children are actually operating with higher order rules (abstractions or algorithms) or if they are doing something a little less complex by looking for analogies. I think the rules vs. analogies controversy is still an unsettled question, and a matter of interpretation.*
>
> *You can also view the Wug test results as evidence that children have the ability to abstract and generalize as part of their all-purpose cognitive armamentarium, or you can think that the results are more evidence that language is a separate faculty, because what children are doing linguistically seems so much more sophisticated than their other cognitive attainments.*

Generative lexicon theory embraces the fact that we can produce specialized compositions in novel situations, in a non-dictionary-like way. It is an important step forward in understanding language learning and pedagogy.

So, what is a lexical *syllabus*? Although some literature discusses a lexical

syllabus simplistically as a vocabulary-based syllabus, I will focus on the progressive form, as discussed above. The lexical syllabus moves away from straight grammar teaching and focuses on word/phrase usage, frequency, and their unique generative powers. Therefore, instead of moving from simple grammatical forms and gradually building up to complex forms, the lexical syllabus focuses on the unique generative properties per word or phrase with the keys being found with the usage and frequency, often based upon corpus studies. In essence, each word/phrase is treated as having its own generative word map; different usage patterns are studied per word/phrase, making the learners experts at identifying the word and its varied usages efficiently. Below (Figure 9) is a map of the usage of *HAVE*.

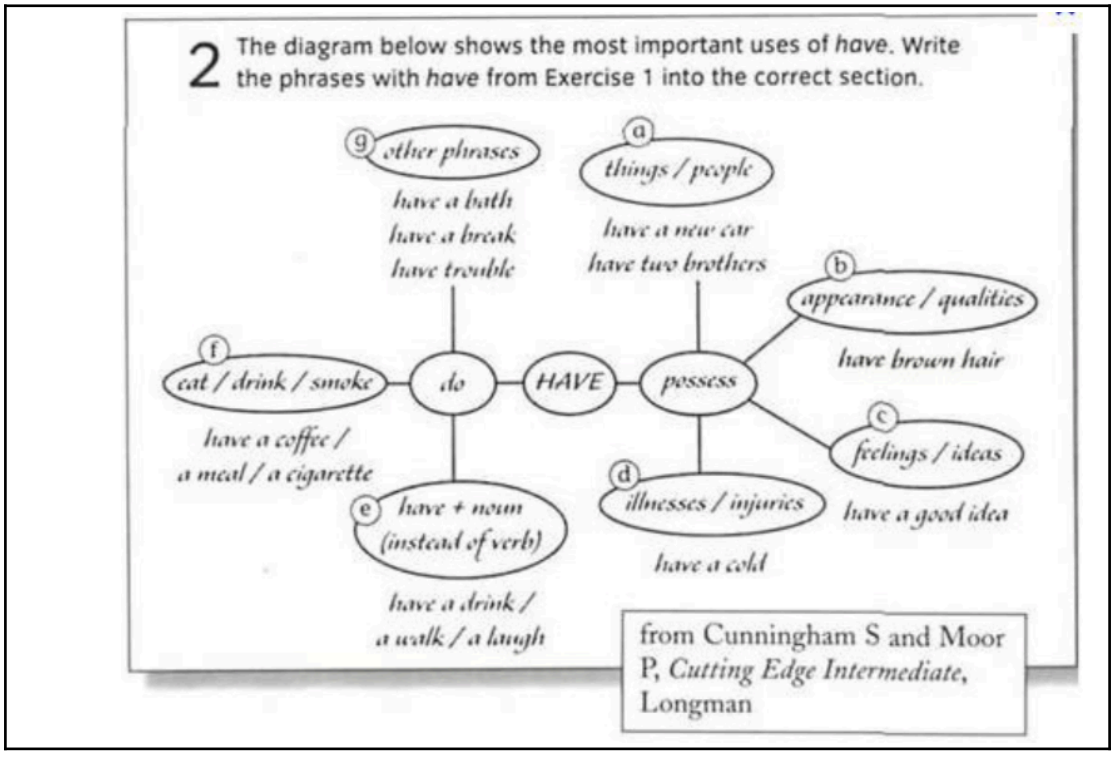


Figure 9: A map of HAVE usage, taken from a coursebook lexical activity (Cutting Edge, Longman).

The lexical syllabus does not (*cannot*) offer an underlying theory of language acquisition (Richards & Rodgers, 2001) for issues explicitly implemented by Berko-Gleason's quote→ this is still an unresolved issue, half a century since Berko-

Gleason's *Wug* experiment. However, the lexical syllabus does assume that the rote memorization of the grammatical rules of a language is not as effective as repeated exposure to naturally occurring usage. Therefore, it agrees with the cyclical syllabus.

The underlying pedagogical theory was, by viewing words/phrases as *lexical items*, the learners would be able to have a better chance of recognizing *and* using the words and phrases correctly in real word contexts because of the cyclically interconnected coverage generated from the fusion of meaning and potential usage of each lexical item. On the other hand, in the real world, a strict *focus on grammar learning* via a grammar syllabus would call for real-time spontaneity, creative thinking, an excellent memory, and the ability to take risks for the sake of communication. This is too much to deal with for the average learner, especially under pressure to communicate.

With the real-world demands put on potential grammar-based learners (as users) being so intense, the lexical syllabus, with the way it primes learners on all typical usages per word, *then providing them with the capacity to generate context appropriate phrases*, seems to be a much more logical and elegant pedagogy. This was an ingenious step away from straight grammar teaching that brought to light a very new way of looking at language teaching and linguistics, but it was not without its critics.

Criticisms: The Lexical Syllabus

Pesky questions that arise with the corpus-derived syllabus such as, which corpus should we choose and how should we use it? Should a beginner course use a corpus created from children's books or from adult's? Either answer (or a mixture of both), could be justifiable but the choice would significantly alter the content of the syllabus; usage and frequency vary greatly per corpus. How can we make that choice? Also, from the Willis' perspective, there is the inevitable heavy reliance of *usage analysis* on the learners' side. So, without comprehensive and cohesive analysis, the learners may gain little from the lessons. It is therefore an ambitious and demanding syllabus design for both the learners and the teachers (Tagg and

Woodward, 2011). As Willis and Willis define it, the lexical syllabus has students be the researchers, while the teacher is only the facilitator, who may not know better than the learners in some cases (the results from the analysis may be just as illuminating to the teacher as it is to the learners). Lewis contradicts this by shifting the focus onto the importance of teacher talk and the teacher's output. Richards and Rodgers (2001:136) discuss how, rather than exist at either extreme, an eclectic mixture would perhaps be more favorable.

Identified problems/questions with the lexical syllabus that have yet to be worked out:

1. No standard or unifying foundational theory of language acquisition
2. Focuses on single words or collocations too much; what about longer phrases?
3. Should the highest frequency lexical items (such as: the, and, but, and a) be taught first, or not? [Logical arguments can be made for both paths]
4. Willis' implementation works in tandem with task-based pedagogy. So, is the syllabus a task-based syllabus, or a lexical syllabus?

The lexical syllabus has yet to gain mainstream popularity, although it remains highly innovative *theoretically*. Perhaps the above points linger over the lexical syllabus and hamper its attractiveness to potential users.

On the other hand, collaborative efforts with the task-based syllabus seem much more promising. Along those collaborative lines, Segalowitz and Gattbonton, who have done extensive work in psycholinguistic issues, provide ACCESS methodology (2005) a form of Communicative Language teaching, with innovative cyclical activities for the automatization of formulaic language. Regarding the lexical syllabus, the emerging patterns seem to be that lexical *theory* can be used for innovative purposes in other syllabi types, but that the lexical syllabus cannot hold its own.

Advantages: The Lexical Syllabus

The lexical syllabus is a unique approach to language teaching—based on research. It is therefore more scientific than other designs. Perhaps the biggest advantage of the lexical syllabus (from the pragmatic point of view) is the lack of guesswork; the syllabus is not based on past writers' assumptions—and observations thereafter. With the focus on lexical items and usage that is based on real-world usage, the learning content tends to be more natural and may even foster deeper processing (Schmitt and McCarthy, 1997:3). But even more importantly, the introduction of generative lexical concept was groundbreaking. Doing away with the grammar vs vocabulary dichotomy, and viewing word/phrases as having their own generative grammatical qualities based on context was a significant step forward.

Synthesis: The Lexical Syllabus

Perhaps the lexical syllabus should be summed up as "having great potential as a hypothetical design, but not developed enough to be a syllabus"; it is a major theoretical syllabus type that is not globally popular, although there are currently some corpora-based coursebooks commercially available; the flagship of such coursebooks seems to be the *Touchstone* series (McCarthy, 2005) from Cambridge University Press. However, without a unifying theory underlying the teaching, the lexical syllabus remains shallow from the researcher's perspective, and it suffers from this. It is therefore often combined with other syllabus types, such as the task-based syllabus, discussed in the next subsection.

The take-home ideas for the new syllabus:

By far, the greatest take away would have to be the generative lexical concept – the viewing of words/phrases as having their own generative grammatical properties and unique nuances based on context. The inspection of this concept into the new syllabus would certainly help to be a much more dynamic syllabus.

2.7.4 Task-based Syllabus

What is a task-based (TB) syllabus? In theory, it is a type of syllabus that is learning- centered (not to be confused with learn*er*-centered). It therefore should be a *procedural* syllabus based upon teacher selected tasks that are assumed to promote the language acquisition process, while downplaying the learning of pre-selected linguistic content (Tagg and Woodward, 2011). Skehan's view (1998:260) is that the students should be given freedom within the framework of the tasks for real communicative engagement in the tasks. Skehan's view can also be construed as being learn*er*-centered, making it a hybrid learn*ing* and learn*er* syllabus. More importantly, in contrast to other types of syllabi discussed earlier that have little or no pedagogical protocols, the TB syllabus, grounded in task-based methodology, is a syllabus that is inseparable from the methodology. In other words, it is a syllabus that is based on a clear methodology and theory. That said, defining what a task actually is or isn't has not been easy, and has become the cause of some controversy, muddling up the clarity of the methodology.

Dörnyei (2009) discusses how TB is little more than a repackaged form of communicative language teaching (or perhaps notional-functional), with the most significant difference being that it may incorporate a bit more *focus on form*. It may be controversial to say this, but since TB is so grounded in methodology, it could/ should probably primarily be viewed as a *method* instead of a syllabus type. In that way, it could be meshed with other syllabi, fortifying its own mission while enhancing the other syllabi, as with the lexis-TB culmination discussed in the previous section.

What is a task?

Anything related to learning can be construed as a task, but this definition is not very helpful. Van den Branden (2006) concurs. He proposes that it should be an activity that has a communicative goal to be met via meaningful usage of language, rather than, but not exclusive of a *focus on form*. In other words, it can be just about any communicative classroom activity that does not have a keen focus on grammatical accuracy. Not everyone agrees. Below are a few typical definitions:

Nunan:

> A piece of classroom work which involves learners in comprehending, manipulating, producing, or interacting in the target language while their attention is principally focused on form. (1989:10)

Van den Branden:

> An activity in which a person engages in order to attain an objective, and which necessitates the use of language. (2006:4)

Edwards and Willis:

- The principal focus of a task is on the negotiation of meaning rather than language form
- Learners are aware of the goal or purpose of the task, which may include creating a list or a map, of solving a problem
- The outcome of the task is to be shared with other students
- Listening, reading, writing and/or speaking may be involved
- Language focused study is incorporated into a TBL lesson, but not generally until after
- the task is completed, so that learners remain focused on the communicative purpose of the task. (2005:3)

Skehan (1998:95) takes this further and emphasizes the need to make the task emulate real-world usage with the goal being the solution to a real-world problem. This seems to be the natural and logical extension of this methodology; however, Van den Branden (2006:6) argues the validity of real-world problem solving within the classroom. Just how realistic can an artificial environment be?

Perhaps the more global question is, compared to the TB syllabus, how realistic are the other types of syllabi and their textbooks? J. Willis (1996a:68) calls

them “impoverished and restricted”; they only render caricatures or over-simplified versions of usage. Communicative significance and socio-cultural nuances are often stripped from, or rendered undetectable from the material because they are presented in packages. D. Willis (1996:51) concurs, "we know that language is a complex system which cannot be ‘presented' to learners in a series of neat packages."

But is the TB syllabus the answer? J. Willis (1996a:88), realized the need to expand; learners needed real-time interaction outside of the classroom for exposure to semi- prefabricated chunks of usage, which could lead to top-down discourse analysis. In this way, we can see how Willis and Willis bring corpus analysis and TBL together.

If linear learning is a fallacy and naturally cyclical learning (relying on each learner’s internal analytical processing that is not directly controlled by the teacher) is closer to reality (as discussed by Dörnyei, 2009), then the TB syllabus should have more potential for elegant language learning and teaching.

Results from TB learning are not linear and therefore may not be immediate (Skehan, 1996:18-19). For this reason, TB does not *directly* teach grammar nor does it expect learners to rote-memorize grammar points. Learners are seen as hypothesizers that learn from implicit and explicit guesswork via task work, or a *conscious* focus on meaning via interaction with an *unconscious* focus on structures (Skehan, 1996:18-19).

The TB syllabus is the first syllabus type introduced here that seriously considers the cognitive processing and development of students. Relevant keywords are *meaningful interaction*, *authentic language*, *analysis*, and *motivation* (Tagg & Woodward, 2011). Long and Crookes (1991), and Prabhu (1987) also argue similarly, with a focus on the importance of *negotiation of meaning* and the reliance of internal analytical processes, while J. Willis (1996b) also stresses the importance of student motivation.

How is the TB syllabus designed?

The typical J. Willis (1996b) style has three parts:

Pre-task: an introduction to the topic, the task, special words/phrases
Task cycle: small group work; the teacher is a monitor; plan and prepare in small groups for a public demonstration of their outcomes.
Language focus: not to be grammar-focused, but an exploration of language usage for consciousness-raising and reduction of fossilization.

Because the term of "task" is so inherently broad, there have been a number of classifications that have been mandated over the years:

Task Classification: Dichotomies

- real-world tasks vs. pedagogical tasks; Nunan (1989a)
- divergent vs. convergent; Long (1989)
- one way vs. two way; Pica, Kanagy, and Falodun (1993)

Willis and Willis (2007:72-78), focusing on cognitive processes list these seven *requirements* for tasks:

- Listing
- Ordering and sorting
- Visual support
- Matching
- Comparing and contrasting
- Problem-solving tasks and puzzles
- Projects and creative tasks
- Sharing personal experiences

No matter what the task, there is a substantial reliance on autonomous discovery along with peer-to-peer affect (inclusive of help *and* pressure). Although this may seem a bit too idealistic, it does emulate real-world interaction, probably more realistic than any of the other syllabi introduced so far. However, it should be noted that the other syllabi were mainly *method-less* (lacking in theory), whereas TB

is the deliberate meshing of theory, method, and syllabus.

Criticisms: The Task-based Syllabus

Are there criticisms of the TB syllabus? Samuda and Bygate (2008) find it to be controversial and to be an attractor of vocal critics, as does Dörnyei (2009). Van Avermaet and Gysen (2006:29) identify three problems: (1) specificity, (2) complexity and (3) extrapolation. Regarding specificity, it is assumed that the teacher/writer bases specifications on some sort of a needs analysis, but is it really possible for the teacher/writer to be able to come up with an appropriate set of tasks for every student need? If not, how useful would approximations be?

Regarding complexity, how is a task to be rated? Keywords are *cognitive familiarity*, *cognitive effort,* and *communicative stress.* Indeed, these are concepts that can be scientifically tested for with fMRI machines, and perhaps EEG machines, but using machines in the classroom or even *bringing the classroom to the machine*, can be problematic—especially the fMRI, due to its size, noise, and cost. Long (1985) and Duran and Ramaut (2006) discuss their views from external observations, which ironically, will not produce accurate results. So, assessing complexity in this context seems to be a difficult, if not a nearly impossible undertaking.

Regarding extrapolation, how much can be expected from the learners? How much should the teacher attempt to induce? If a teacher eventually "caves in" and sums up the learning with explicit grammar instruction, would that negate the previous learning, or add to it? What if the students request explicit grammatical instruction during the task? Would it be wrong to assist them? These are all critical questions that each TB proponent may have different answers to, making the TB syllabus much less transparent and more confusing than it idealistically set out to be.

Moreover, Edwards and Willis (2005:27-8) point out that teachers comfortable with a grammar-based syllabus may be hostile toward TB learning because of the lack of control over language usage and linguistic outcomes. Indeed, a common critique is that a non-native English teacher may not feel competent enough to facilitate English language learning—when there is no answer sheet or rulebook to

rely on. Without a native or native-like competence, how can they assess the learning output *competently*? A competent assessment would be unlikely. Hence the hostility toward TB. This hints at one reason for the popularity of the *grammatical* syllabus among non-native teachers in countries such as Japan – you do not need to have a native-like competence as long as you have an answer sheet or rulebook.

Advantages: The Task-based Syllabus

Are there advantages to the TB syllabus? There are several, some on the theoretical level and some on the pragmatic level. For one, it acknowledges the inadequacies of all product syllabi; it takes into account human cognition and cognitive development by being a process based syllabi. It is also procedural. It builds and relies on learner autonomy — something very real and important when the student leaves the classroom. Also, students will inevitably be at different levels of competence and motivation in any classroom. By consistently using group work in the beginning, (1) learners of different levels may feel less pressure because of the smaller number of eyes on them at any given moment, (2) learners implicitly and explicitly motivate each other and help negotiate meaning with each other at a group level, and (3) no matter how intuitive a teacher/writer may be, their predictions regarding teaching order cannot replace real-time language usage during a realistic task (Cox, 2005). Also, there are more progressive ways to go about TB, for example, Robinson's (2011) simple to complex sequencing of tasks.

The take-home ideas for the new syllabus:

TB is a well-designed and thought-out method. However, it should not be considered to be a stand-alone syllabus. That said, the methodology of TB is progressive and there this much to learn from the design. Ideas from TB that could be used for the new syllabus are:

1. focus on the process
2. raise autonomy by letting the learners make learning/usage choices
3. implement small group work → move to big group work
4. make liberal open-ended tasks
5. provide real-time focus on form sessions

2.7.5 Content-based syllabus

What is content-based instruction (CBI)? Is not all learning "content-based"? According to Brinton, Snow, and Wesche (1989: vii), it is the "integration of content learning with language teaching aims. More specifically, it refers to the concurrent study of language and subject matter, with the form and sequence of language presentation dictated by content material." Historically, CBI has very pragmatic roots, for example, according to Larsen- Freeman et al. (2011), the roots were based in "the 'language across the curriculum' movement for native English speakers in England, which was launched in the 1970s to integrate the teaching of reading and writing into all other subject areas."

What good came of this? Some potent criticisms of standard L2 teaching have been: (1) the separation of "meaningfulness" from the language being taught (Kelly, 2012), and (2) the "dumbing-down" of the content; the learners are typically treated as *cognitively deficient people* (Murphy, 2009). But how else can you teach the basics of an L2? If you are teaching an L2, isn't it inevitable that you end up treating learners as being cognitively deficient? For example, Klein (2001: 13768) says, "Linguists and laymen alike tend to consider children's way to their mother tongue to be the most important type of language acquisition."

CBI does away with this seemingly irrevocable dilemma. With a content-based syllabus, the learners are given a real topic to study that presumably matches their cognitive capabilities, and expects them to make do with the content's learning via their L2. It provides the learners with a cognitive challenge that is realistically motivating in a way that is disconnected from the typical focus on language learning. Although it may sound counterintuitive and/or counterproductive, especially from a grammar-based teaching perspective, this type of learning calls for and expects the onset of natural motivation to use the language to accomplish the activity without ever having to focus on the language usage as an integral part of the learning. In CBI, the L2 usage is not necessarily seen as anything other than what it presents as its inherent instrumental value. In short, the L2 is no longer the *subject of study*; the L2 becomes the *means to reach the goal.* Brown (2001:49-50) seems to concur by saying, "Content-based classrooms may yield an increase in intrinsic motivation and empowerment, since students are focused on subject matter that is important to their lives." The topic of *empowerment* shall play a crucial role in later chapters.

Brown continues, "Students are pointed beyond transient extrinsic factors, like grades and tests, to their own competence and autonomy as intelligent individuals capable of actually doing something with their new language." (p50). Larsen-Freeman et al. (2011:144), state that, "CBI can also be an effective way for students to learn language in the language class, using themes that students find of interest. Such themes provide sustained motivation beyond intermediate levels of proficiency and prepare students." It seems there is a consensus regarding the motivational qualities and the real-world qualities of CBI, at least on a theoretical level.

Criticisms

With such a promising foundation, what could be wrong with CBI? On the theoretical level, it does look promising. However, several drawbacks can be uncovered. For one, finding suitable content may prove to be difficult depending upon the composition of the students' needs. Also, would CBI be appropriate for all

age groups? --Probably not. Therefore, because CBI attempts to *not* "dumb-down" the content, it may not be suitable to both young learners and novice learners; the study of linguistic structures would seem to be a necessary step before a meaningful stab at CBI could be made for novice adult learners. However, these are relatively minor quibbles.

The most significant criticism would be that when and if students fall into the flow of CBI in a meaningful way, they can all-too-easily disregard the *means* (the L2) and "triumphantly" reach their *ends* with their L1, and still be fully satisfied with the outcome. This is because L2 usage, while expected, can become secondary to reaching the goal. Teachers can monitor and police the L2 usage to some extent, but if students are highly determined to reach that cognitively challenging goal, it is within their natural state to lose their self-control (due to mechanisms related to *flow*, [Csikszentmihalyi, 2008]) and slip into their L1—to complete the task.

There is a second problem that is not solely a CBI issue, but a serious issue nevertheless – L2 learners should have at least some explicit L2 instruction. Implicit learning alone is not the best way to learn the L2 (Dörnyei, 2009; Lightbown and Spada, 1999). Admittedly, CBI can include explicit L2 learning, but because the basic premise is to make the L2 learning secondary to a nonlinguistic learning goal, it is natural to assume that learners within a CBI framework may not be able to receive ample *focus on form*.

Advantages

The most prominent two advantages of CBI must be: (1) the motivational aspects that coincide with empowerment and meaningful goals, and (2) the fact that students are no longer only focused on studying *about* the L2, but using it as needed, naturally. Furthermore, CBI is easy to implement. In fact, it does not need a specialized coursebook; any English language book from other school subjects could serve as the centerpiece of a CBI course.

Synthesis

In theory, even considering its drawbacks, CBI is full of potential. In fact, it is popular around the globe -- clothed with different titles. For example, what is the difference between CBI and *English for Specific Purposes* (ESP)? Depending on whom you listen to, they either have considerable overlap, or they are practically identical, with the latter perhaps allowing for more focus on English language learning than the former. Instead of saying that CBI is a synonym of ESP, Whong (2011) goes so far as to say that CBI is a *type* of ESP. From a theoretical SLA perspective, the ESP framing may have the advantage because it would seem to have more inclination to focus on form than CBI would.

The take-home ideas for the new syllabus:

Because CBI forces focus on content-based goals yet can implement open-ended responses, it is similar to TBL. It is much more extensive in scope though, because it is not limited to language task implementation; it is focused on real-world learning, not language learning *per se*. The take-home ideas for the new syllabus are: (1) forget about language tasks at least part of the time and get learners to really learn something new to challenge their intellect, (2) empower the learners by doing something meaningful with the new learning.

2.7.6 What can be learned from the "Spiral Curriculum"?

In *Education* literature, Bruner (1960) is often cited as the instigator of the *spiral curriculum*. There are intriguing features to Bruner's proposed design, yet interestingly the spiral design was never able to bear much fruit in *Applied Linguistics* and *TESOL* groups. I have included it here in this literature review because although it is practically non-existent in second/foreign language literature, it has rightful prominence in *Education* literature. Moreover, this current study borrows from the spiral design but adds a few modern twists and tweaks to make it efficient in the EFL context.

What are the features of the traditional spiral design? Ideally, it is a curriculum that encompasses all major topics that are deemed worthy of study (regardless of genre). Bruner hypothesized that any worthy topic could be taught to young children if included into the spiral (as long as the complexity of the content matched the learners' development during each cycle). Below are four essential aspects to the spiral design:

- Content is revisited. This is not only for review but also for growth.
- The complexity of the content is gradually increased, per return.
- New learning is never *entirely new*; it is always related to previous learning.
- Via the spiral of learning, competence is raised.

Critical analysis and takeaways

Although the spiral syllabus sounds reasonable and is solidly based on the constructivist tradition, it is not without criticism. Johnston (2012), asserts that there is little empirical evidence of its overall effectiveness in the ideal form. However, he rightly contends that since the studies of spiral curriculums entail constructivist teaching approaches by constructivist teachers, it is hard (if not impossible) to separate the efficacy of the *delivery* from the efficacy of the *curriculum*. How can that be accomplished meaningfully? It cannot. So, this is a fair point with significant

implications for this study as well. Also, although the "overall effectiveness" data was somewhat lackluster, Johnston did find notable growth in areas such as reading, writing, and technical skills. So, although Bruner's idea of having *all* worthy domains of knowledge spiraled into a single large over-arching curriculum may not have proven to be as productive as hypothesized, smaller individual areas did show promising growth—I have therefore targeted specific areas for growth with my coursebooks designs, as will be discussed across the second half of this thesis.

2.8 Chapter Conclusions

In this chapter, I discussed five major language teaching syllabus types: *grammatical*, *notional-functional*, *lexical*, *task-based,* and *content based*. I have also commented on the efficacy of the spiral curriculum, and how it influenced my emerging new syllabus design. The oldest of the syllabus types discussed here is the grammatical syllabus and being based on the teaching of a dead language—*Latin*—it has proven to be not well suited for practical language learning such as expected in conversational language classrooms. That said, being the oldest and most naturally a "back to basics" approach for explicit teaching that we have conceived of historically, it seems to have a naturally appealing quality for (uninformed) learners and teachers whom are past the age of puberty or their *critical/sensitive stage* of natural language learning. Because they are past this stage, they are probably less likely to be able to implicitly (unconsciously) pick up their L2. *"*Post-critical period SLA is notorious for its difficulty, high degree of variation, and often very poor outcome. The primary aim of L2 instruction is to ameliorate, if not solve, these problems." (Doughty 2003: 256). This leaves learners and teachers with mainly explicit measures -- grammatical methodologies are well suited for such logically explicit analysis of the L2. So, the appeal is high. The only problem, and this is a *major problem*, is that explicit analysis and memorization of L2 grammatical structures has not been proven to actually improve L2 fluency, and in fact has been shown to be detrimental in some cases (Lightbown and Spada: 1999), even though

EFL students (such as Japanese students) may express a strong preference for it.

That said, focus on form certainly can be helpful for learners. In that vein, Tokuhama-Espinosa asserts, "With the exception of accent, adults can and do learn foreign languages better and faster than children if and when they spend the same amount of time in the process" (2018). The keyword in her assertion is *time*. How many hours does a child or university student spend focused on language learning vs. an adult learner—and—who focuses more on form over content, the child or the adult? When taken together, the implications are huge. How can the ideal balance be achieved?

Finding the correct balance between accuracy and fluency is sometimes described as "an art more than a science." This is due to the inevitable affect of contextual variables and motivational dynamics that *statically* designed syllabi (often top-down) cannot take account of (they do not adjust well). The *NF syllabus* was designed to be an ingenious improvement over the grammatical syllabus, but because it is still product-based and static, it carries over many of the same criticisms of the *grammatical syllabus*.

I would like to applaud the *lexical syllabus* for being scientifically oriented, but without its own theoretical pedagogy, it is vague and somewhat lifeless as a stand-alone syllabus. It is therefore often paired with task-based learning, which is also a limping syllabus type--if it is an actual syllabus type at all. While these two fit together well enough as a mesh of ideas, and they do seem to have good potential as a team, they have never gained substantial global popularity in the classroom (currently a hot topic for researchers on a theoretical level though). This may be because the methodology is complicated to understand and because task-based activities depend upon the teacher as being a monitor and facilitator; non-native teachers may feel inadequate when it comes time to assess without a provided set of "correct" responses that they can rely on—while native English teachers may deem their mandated job as a monitor and a facilitator to be too simplistic, and perhaps a waste of their talents as an instructor. The final language teaching syllabus reviewed was *content-based* (and variant, *English for Specific Purposes*). This syllabus type is

commendable for downplaying the "classroom subject" aspect of English, by doing away with learning *about* English and focusing on simply *engaging* in interesting learning via English as a tool. I would argue that content-based learning potentially lacks a significant focus on form, while *English for Specific Purposes* puts the English language usage back into focus. So, although they are very similar, there is a significant difference with the inherent amount of focus on form. CB is inherently more fluency based while ESP is more accuracy based. Honorable mention goes to Burner's *Spiral Curriculum* because it is sorely missing from *Applied Linguistics* and *TESOL* literature.

All of the summary points I listed above play significant roles in my new syllabus design—the gist being: recycling content *for growth* (not just to help memory), and that *all* new learning should be systematically designed to be *built upon prior learning*, so that it is all given a fighting chance to be systematically linked in the mind of the learner.

Is there one best syllabus type from the bunch? As we have seen, they all do have strengths and weaknesses. So, there is yet to be a *best* design for all contexts. However, one of the goals of this thesis is to design what I find to be the best possible syllabus design solution. It is therefore essential to locate exceptional points from each of the above syllabus designs and attempt to *coherently* synthesize them into a dynamic syllabus—with significant weight put on the idea of including cyclical designs along with the embedding of efficient formative assessment.

The next chapter will discuss material development in light of these points. It will produce a list of best practices in much the same way this chapter did. In this way, these literature review chapters will set the critical issues and the target the best practices so that a master list of best design ideas emerges. This will become the blueprint for the new syllabus design.

3. Literature Review Chapter: L2 Materials Design Theory

This chapter on L2 materials will focus mainly on the *theories* behind language coursebook designs, inclusive of their digital incarnations. In the broadest definition, L2 materials can be anything that helps facilitate the language learning goal (Tomlinson, 2003). It seems safe to presume that language *learning* and language *teaching* occurred naturally via human interactions well before the concept of "school books" arrived. So, we can deduce that humans do not actually *need* coursebooks to teach and learn languages. It would seem that any object or medium could be successfully used as *language learning material* as long as it provides stimulation conducive of language-based interaction. (Note: Much of Chapter 3 content has been previously published in the Kiban Center Kiyo Journal, University of Kitakyushu.)

Why do we have this "infatuation" with coursebooks (Richards, 1998:127), for our teaching endeavours? Perhaps we seek an elegant and transparent teaching/ learning experience. By having the content predefined and pre-developed (and unified across all learners), the teacher can spend more time on the specific teaching mechanics and not so much time on the developmental issues of the course (Richards, 1998:135). However, the most convenient and popular coursebooks tend to systematically avoid controversial content inclusion and end up portraying what is *safe*. They put on display what they consider to be normal language usage as used by an idealized middle-class white person (Richards, 2001:255). In this case, playing it safe sounds silly—but that is how the most popular coursebooks have been designed, traditionally. On the other hand, companies such as Macmillan Language House (Japan), take the niche route and publish books that are explicitly designed for smaller markets such as Japanese schools, focusing on the needs of Japanese students—and the Japanese view of the world. These focused books would not be appropriate for use in most other countries.

In sum, we have a profoundly ironic situation. While literally anything can be used as a prompt for successful language learning, the actual coursebooks that sell the most, do so at the expense of competence. How can that have happened? In this chapter, I will elucidate this awkward state of affairs.

3.1 Introduction to Learning Materials

3.1.1 What are the pragmatic purposes of the L2 coursebook?

Input: Perhaps the most commonly cited purpose would be language input. Coursebooks display language usage of many genres and levels. Some display synthetic constructions to nail in grammatical points, while on the opposite extreme, others display authentic texts that may not fit into standard grammatical forms. Although grammar teaching is still prevalent, the trend seems to be moving toward more authentic and engaging usage examples (Tomlinson, 2003:7), with the internet being an obvious goldmine that is easy for anyone to dig into.

Structure: The structure that coursebooks provide can be helpful for the entire course or even for a single lesson or activity (Hutchinson and Torres, 1994). Richards describes it as a map that provides coherence (1998:130). By returning to the same map over and over again, a teacher may achieve a cyclical learning pattern even with a book that was designed as a linear one. In this way, even with coursebooks that provide a rigid and linear structure, it is often still within the teacher's power to explore tactics that may bring on (ideal) cyclical teaching patterns like the ones discussed in Chapter 2.

Motivation: Coursebooks can be (and probably should be) engaging and motivational (Richards, 1998:135). From the student's point of view, coursebooks make the learning transparent. Without a coursebook, students may feel their classroom time is aimless. Depending on the teacher, they may be correct with that assumption. Motivation is a tricky issue with many conflicting attributes and definitions (Dörnyei, 2011:4). And, not only should books be designed with the right pace and content, but they should also be aesthetically pleasing (Tomlinson, 2011).

Alignment: *Teacher-student alignment* is essential, not only for rapport but also for efficient learning and motivation. By using a coursebook, at the very least, there is potential for teacher-student alignment regarding the course content via the coursebook. However, this can lead to *reification* (Richards, 1998:131) perhaps partially due to too much trust put in the coursebook, the writer, or the publisher (--in other words, it can become a *"thing"* when it may not deserve such a lofty status). Coursebook-teacher alignment and coursebook-student alignment can be highly influential and may be a significant cause for some coursebooks to continue to be bestsellers *despite* their content.

Reformation: As can be seen from point (4), coursebooks can have powerful influence due to reification. One prime example: public school coursebooks' changes can and do cause district-wide educational reform seemingly "overnight." Coursebooks can therefore be used as an agent for change.

Cost-performance ratio: Having each teacher in a school develop their own materials is not unheard of. This is especially true in this age when any teacher can publish their work online. However, the overall quality from a group of professional publishing teams with qualified staff, each assigned to dedicated jobs in areas such as designing, editing, artwork, and print finalization—will very likely produce better results than solo writers; this is obvious. Moreover, the cost to buy these beautifully designed pre-made materials is *significantly* cheaper than self-printing with personal color printers. The tradeoff for all these benefits is that major publishers literally cannot cater to the individual needs of students the way classroom teachers can. The materials cannot adjust to the students' needs, so, the teachers and students much adjust (and adjust to) the materials instead.

How significant a point is this? I hypothesize that our collective devotion to coursebooks (good or bad) over the past several decades *must* have had a significant impact on the *culture of teaching*, teacher beliefs, and what ends up becoming "common and acceptable" teaching practices. Simplistically put, our field has become defined by what the major publishers publish, and the trends they help create. The next section will shed light on the impact of this from a historical perspective.

3.1.2 A brief history of Methods and Approaches

Mainstream coursebooks will often represent at least one primary methodology. Below is an adapted list of major methodologies/approaches and their years of greatest dominance, from Richards (2001:3) and Brown (2001:13-37):

- Grammar-Translation Method (1800-1900)
- Direct Method (1890-1930)
- Structural Method (1930-1960)
- Reading Method (1920-1950)
- Audiolingual Method (1950-1970)
- Situational Method (1950-1970)
- Communicative Approach (1970-present)
- Community Language Learning (1970-1980)
- Total Physical Response (1970- present)
- Natural Approach (1980-1990)

I will discuss some of the highlights from the history of methods and approaches starting from the top; I tend to agree with the list, but Grammar-Translation (1800-1900) is *still* the traditional-standard methodology in Japanese schools. It certainly did not end in 1900(!) and frankly still has a significant influence on Japanese students *and* their outcomes. The Natural Method was a reaction toward Grammar-translation and is its polar opposite in significant ways. It is an inductive method conducted only in the L2, featuring gestures and modeling as vital inductive aspects. In extreme cases, *motherese* is used to make it even more "naturalistic" for the learners. Perhaps it may not be that different from what we now collectively recognize as the Communicative Approach.

The Audiolingual Approach was based on the then-new Behaviorism ideas from psychology. It was (unsuccessfully) designed for high-efficiency in language learning for military purposes. It has learners drill sentences chorally in high-repetition with slight changes made periodically—the drilling was an aptly militaristic design. It was based on structure learning, and it was *not* communicative. In contrast, the Natural Approach (touched on above), seems to be a mixed bag of communicative ideas. It centers around the concept of *comprehensible input.* It denounces grammar stud. Instead it focuses on input activities in the L2 that are done until comprehension is accomplished. The learners are expected to begin speaking when they are ready, but are not forced to speak until that time.

Although the Natural Approach had a huge impact on a global scale, its popularity seems to now be dwindling as the pendulum is moving back toward a *focus on form.* As can be seen from this historical perspective, what can be said is that although we have seen innovations via science, there always seems to be a trend pendulum that shifts around the importance of accuracy and fluency.

3.1.3 More dichotomies to consider

Chapter 2 discussed several dichotomies that needed to be considered in syllabus design. Is there anything *practical* for the contemporary materials writer to latch on to? Thornbury (2011:192) explains how he updated Richards and Schmidt's (2002) list of six domains that distinguish L2 methods. That is a good list that seems to be more relevant for the researcher or student of applied linguistics. So, instead, here I present a more practical, question-based style list for material writers based on Richards and Schmidt's, and Thornbury's work:

Which is more important for your materials' content?

- structural memorization, or (b) fluency?
- formal/explicit learning, or (b) informal/experiential (holistic) learning?
- confined target language usage, or (b) liberated (but less accurate) usage
- one-way learning, or (b) interactive learning?
- deductive (unambiguous), or (b) inductive (mysterious)?
- multi-lingual, or (b) mono-lingual?

This list is meant to provide a sense of clarity for material writers—inclusive of the designs for my current study: Q1 is based on the formal vs. functional dichotomy. Q2 is based on the analytic vs. experiential dichotomy. Q3 is based on two similar dichotomies: product vs. process, and accuracy vs. fluency. Q4 is based on transmissive vs. dialogic. Q5 and Q6 are self-explanatory.

3.1.4 Can coursebook usage be detrimental?

Coursebooks may not be as useful as they are depicted by publishers and other media. In fact, they can be detrimental. Here are four major points of consideration:

(1) The Humpty-Dumpty postulate: Dismembered parts cannot be put together again. In support of this postulate, Apple and Jungck (1990) have a "separation of conception from execution" argument. That is, when a whole is broken down into smaller parts, it is all too natural for us to lose sight of the bigger picture, and it is erroneous to expect that learners will be able to make all the disjointed pieces coherent by will. This seems to be true for both the teachers and the students. Both the teachers and the students may get so wrapped up in their "micro-studies" of language usage that the macro-level becomes irrelevant; they may forget to question how working on a small piece of the language affects the bigger picture—such as language in the real world.

(2) Coursebooks lowball. Individual differences are inevitable. Tomlinson (2011) argues that coursebooks tend to dumb-down learning by *expecting* short attention spans and a lack of intelligence. A better design would allow learners to follow their passions within the given framework—following their gut feelings regarding the learning tasks. This is "bottom-up teaching" (Murphy, 2009; 2012). The efficacy of "teaching to learning styles" is infamous as a neuro-myth; that concept has been debunked. However, allowing students to follow their passions is not something that can be debunked—it is literally one of the most natural things a student can do in the fabricated classroom that is imposed on them. And so, I remain a strong proponent of this. I trust that the *following of passions* by students naturally leads to further learner autonomy, which reciprocally nurtures students' feelings of ownership toward the target content. (I mention this here with conviction because these notions are the backbone of the coursebook designs in the present study.)

(3) Books do not always do what they advertise (Tomlinson, 2008; Tagg and Woodward, 2011). Publishers are businesses. Businesses have business interests. These business interests influence their advertising campaigns—and their books. It is naive to believe otherwise. "(T)he financial motive of these groups does have an affect on the way in which textbooks are produced and marketed and ultimately on teaching and learning" (McGrath, 2002:9). It is highly unfortunate that, perhaps unbeknownst to them, so many learners may be putting faith in learning materials that prioritize money over good teaching design.

(4) Coursebooks are books of compromise. Although Bell and Gower (1998:129) argue that compromise is probably beneficial, authors tend to disagree (Tomlinson, 2003;2008) with reasons naturally stemming from point (3) above.

To further these sentiments, Tomlinson (2008:8) published a remarkable list of how ELT materials can and do *inhibit* language acquisition. I have modified his list here:

- Underestimate learner language level
- Underestimate cognitive ability
- Treat linguistically low-level learners as intellectually low-level learners
- Impoverish learning experience in a misguided attempt to make learning more accessible by simplifying the presentation of language
- Create an illusion of language learning by using a Presentation/ Practice/Production approach which simplifies language use and results in shallow processing
- Create an illusion of language learning by ensuring that most activities are easily accomplished as a result of involving little more than memorization, repetition of a script or simple substitution or transformation
- Confuse language learning and skills development by teaching language features during listening and reading activities
- Present them with bland, safe, harmonious texts and require participation in activities which don't have learners think and feel

- Provide learners with far too much de-contextualized experience of language exemplification and not nearly enough experience of language in fully contextualized use.
- Focus on activities which require efferent listening or reading for detailed and literal comprehension
- Provide very little aesthetic listening and reading which stimulates the total engagement so useful for promoting both enjoyment of the language and acquisition of it
- Fail to help learners to make full use of the language experience available to them outside the classroom
- Focus on uni-dimensional processing of language through activities requiring only the decoding and/or encoding of language rather than on multi-dimensional representation of language through activities involving the use of the full resources of the brain

Coming back to the initial question as to whether coursebooks *can* be detrimental, the answer would be *yes*; there is always the risk that the decision-makers may be incompetent, too business-oriented, malicious, or any combination of the three. Tomlinson came to the following conclusions in 2008 (219-220), below.

Writers and publishers may:

- use too much PPP
- underestimate learners
- not recycle enough
- be too exam-oriented
- not provide enough out-of-the-classroom connections

Unpacking Tomlinson's list:

- **PPP:** *PPP* stands for activities that "Present, Practice, Perform" prefabricated dialogues for pair- or small-group work. It is ubiquitous in popular coursebooks and TEFL certification programmes, although it is less pedagogically sound than it is sold to be. It is popular because it is utterly transparent and easy to assess, making "going through the motions" potentially motivational for the students even if they are learning close to nothing from the exercises. Interestingly, it is now being included in grammar-translation based books in niche markets such as in Japan. This combination is ingenious (but perhaps inevitable) because it: (1) makes the grammar-based texts appear to be more communicative, but more importantly, (2) the inherent transparency of PPP assessment (*just do as the book says*) allows for incompetent L2 speakers to still be L2 language teachers, therefore allowing traditional grammar-translation teachers a viable escape route into (quasi-)communicative language teaching without ever really leaving their comfort zone.

- **Underestimating learners:** Learners are treated as being cognitively deprived by the mush in many mainstream textbooks. The mush is there for several reasons, but commonly cited ones are: (1) editors not wanting to risk controversial issues, and (2) catering to the lowest common denominator to maximize potential sales.

- **Not recycling enough:** As covered in Chapter 2, it is neuroscientifically clear that "linear learning" is a fallacy; learning is a much more organic and yet compounding ordeal. And so, cyclical teaching designs are more readily compatible with natural brain function. Still, all too often, writers and publishers either do not know how to implement this on paper, do not know how to implement this for commercial success, or perhaps they just cannot be bothered to change their long-standing linear approaches to teaching designs (that most consumers and teachers do not seem to have an argument with anyway). That said, the situation is changing with the advent of smartphone-based, tablet-based, and CALL classroom-based learning—the natural segue into more *active learning* classrooms. How do these devices help the learning become more *active*? These devices provide instant feedback with naturally recycling elements; they *can* incorporate a cyclical process that is entirely dynamic (but this should not be simplistically assumed to be the case with all study apps and devices). The irony is that these apps and devices come with their own set of inherent problems—such as a lack of real socialization for the students using them. Still, the essence of this type of design gained from gamified lesson plans should provide material designers with hints for ways to implement cyclical learning in the text-based classrooms, too. In this way, digital media and gamification have serendipitously brought active learning into classrooms that were not originally planned to be active by design.

- **Too exam oriented:** Self-explanatory, but not understood enough by the general public—especially in Asian markets, where *testing for licensing* can be an unquestioned way of life. Tomlinson's notion of *too exam-oriented* was likely meant to mean that such a position is "not oriented for the real world." Ironically, the opposite opinion is the popular opinion in many Asian countries – it is often thought that exams offer safe passage to the real world.

- **Not provide enough out-of-the-classroom connections:** This is a further and natural extension of the above point -- *being too exam oriented*. Classrooms are fabricated microcosms; they tend to become their own little worlds, sometimes with little too input from, or view of, the outside world. Theoretically, this is detrimental to language learning, which is fundamentally and naturally *social* at its core. In this age of digital products, perhaps the outside world is not as closed off as it used to be, but Tomlinson's point should be well taken. As a rule of thumb, language learning activities should have meaningful connections to the world in a *real* way.

On top of this, Richards and Renandya (2002:67) found that coursebooks can "deskill teachers…, the teacher's role can become reduced to that of a technician whose primary function is to present materials prepared by others." The concluding section of this chapter will discuss tactics to reduce the detriments addressed here.

3.1.5 Synthesis: Introduction to Learning Materials

There is little doubt that coursebooks *can* be useful in L2 classrooms, but clearly not all of them are. Their existence, design, usage, and assessments are complex matters—sometimes deceptively so. So, instead of adhering to a strict by-the-book policy, teachers should learn to pick-and-choose from the coursebook's content to dynamically match the needs of their own students. The concept of *dynamically matching students' needs* plays a key role in the studies and conclusions of this PhD thesis.

In that frame of mind, the question of *which* coursebook to use may not be as important as the question of *how* the selected coursebook should be used and modified. For that, we need suitable and reliable evaluations of coursebooks. Ideas for such material evaluation shall be discussed in the next section.

To help bring gravitas and closure to this current section, below is a discussion focused on a personal communication with recognized coursebook author, Marc Helgesen (2012). After decades of materials writing, Helgesen came to the following conclusions as to what is essential in materials writing:

> ***Authoring team*** *(and most good course books have two or maybe three authors). It adds to the creative mix. In my experience, I would avoid anything over three authors. Did that once and it was a disaster. The series lacked focus. Too many creative (and otherwise) forces were going in different directions.*
> ***Editorial team****. People think it's about spelling. It's not. Editorial is really important in helping you see beyond your own classroom/ experience.*
>
> ***Design****. These people are magicians. They can make your eyes move in any direction they want.*

> ***Marketing team***. *In addition to putting your books in front of people, they are one of the best sources of feedback from the market. Much better, for example, than editors based in NY or London who fly in, meet with a few teachers and assume those teachers represent the market. I remember one editorial meeting where an editor said, "Taiwan doesn't want X." One of my co-authors countered, "Would that be the whole country or 3 teachers in a Taipei focus-group?"*

Helgesen identifies four major points in the material writing process. He makes it clear that it is a team effort. The team makes or breaks the coursebooks. Helgesen continues,

> *Something else I find interesting: Books by "super-star" academics are usually crap. The reason is those people usually know everything about a narrow slice of the field. To be a good coursebook author, you have to be a good generalist. Very broad.*

Here Helgesen compares the quality of coursebooks written by stars in the field of Applied Linguistics/TESOL *vs.* the generalist writer. He strongly suggests that the generalist writer has a broader vision and can, therefore, write better coursebooks than the stars who may be too focused on minute topics within the field—and may have some petty grudge or point to prove. And so, generalist writers in general seem to make better coursebooks writers. An example of this is McCarthy's *Touchstone* series.

As discussed in Chapter 2, *Touchstone* is based on corpus linguistics principles. It was found to be massively innovative within the eyes of corpus linguists, but this means very little to the average user who has no idea was corpus linguistics is. (I would like to add that the latest version of *Touchstone* has rectified many of these issues and now has a more generalist design implementation. These changes seem to have helped it gain market share.) Helgesen furthers the discussion:

> *Something else interesting (maybe), and this is just a personal observation, nothing I have "evidence" for -- I've never found books on ELT publishing to be particularly helpful or accurate. Again, they are usually written by an academic about what theoretically ought to be, rather than what is.*

Helgesen points out that he sees a serious problem within ELT publishing (academic, professional development books for teachers and researchers). That is to say that it has become an academic writers' microcosm, being informed too much from wishes and views emerging *from within the circle*, but not necessarily learning from the realities of the classroom.

Why does he feel this is a problem? Helgesen feels this is a problem because virtually all MA TESOL programmes around the world are founded by research based on this singular sphere of ELT literature. This means that teachers, students, and coursebook writers have all been—and will continue to be—influenced by this sphere of literature that is likely misrepresenting many of the realities of our EFL language classrooms.

3.2 L2 Learning Material Evaluation

What is material *evaluation* and how does it differ from material *analysis*? According to McGrath (2002:22), evaluation comes from determining if what you were looking for in the material was indeed there or not — it tends to be subjective in that way. On the other hand, analysis tends to be objective — merely finding out what is there and simply listing them as an observation without subjective critique (Figure 10).

Material Analysis → *objective and scientific*

Material Evaluation → *subjective and personal*

Figure 10: Material analysis vs material evaluation

What is Ellis's (1998) *micro* vs. *macro* evaluation? They are two different types of evaluation with very different purposes. The macro-evaluation, being at the higher level, tends to evaluate the entire course, perhaps looking at accountability and/or course-wide coherence. This is the type of evaluation that is commonly published and read. It provides valuable general information about the course. The micro-evaluation tends to focus on the evaluation of individual tasks. McDonough and Shaw find micro-evaluation to be increasingly essential for teachers (2003). There can be three stages of evaluation:

1. pre-use stage
2. in-use stage
3. post-use stage

The three stages are self-explanatory. The *pre-stage* use is the period before the materials are actually used in the classroom. This is (obviously) not going to be as useful as in-use or post-use evaluations, but it *is* the easiest and quickest type of evaluation. The *in-use stage* is the period when the evaluator observes the material in classroom use and makes an evaluation based on real-time usage. It can be important to make journal entries during classroom time because thoughts and observations can be fleeting during a classroom session. The *post-use* stage is the period after the material has been used. This is also an important stage because the observer will have time to synthesize and realize ideas that may not have come to him/her during the teaching session.

Ellis (1998), Tomlinson (1998), and Rubdy (2003) agree that most focus is usually on point (1), the pre-use stage. This may sound counterintuitive, especially from a scientific perspective, but there are reasons for this. As Ellis (1998:221) rightly points out, teachers will not find much use in carrying out daunting analysis and evaluation of activities that to them obviously either worked well or didn't work well. However, micro-analysis at a play-by-play level may not be as daunting as assumed.

3.2.2 Materials Evaluation Procedures

Evaluation by humans is a subjective and arbitrary process. Therefore, it is no surprise that there is no consensus on how something as complex as a coursebook should or shouldn't be evaluated (McGrath, 2002; Ellis, 1998). What is the bottom line? At the very least, a teacher's teaching philosophy and teaching needs would need to match the course's design at some level for a favorable evaluation. McDonough and Shaw (2003) make an interesting observation: it is possible to subdivide design into the macro level common interests and micro level local interests. This is what McGrath may distinguish as *general criteria* vs. *specific criteria*. The more "hits" per level would gain a more favorable evaluation. Here are Richards' (2001:256-7) criteria for coursebook selection:

- Is there a well-developed curriculum that describes the objectives, syllabus, and content of the program or will this be determined by the textbook?
- Will the book or textbook series provide the core of the program, or is it one of several different books that will be used?
- Will it be used with small classes or large ones?
- Will learners be expected to buy a workbook as well or should the textbook provide all the practice students need?

Richards' criteria list seems to be stemming much more from the publisher/ writer's side than then from the end user's side; they are all design issues that come up during pre-writing stages for a coursebook. However, I noticed that all of those points could be rectified by the teacher at some point during their usage. Therefore, although all of those points may be important at some level, I cannot view any of them as being of top-priority; Richards' list is displayed here to show how divided and mismatched coursebook evaluation is. Although the evaluation process may be divided and subjective, the *analysis* process must be more concrete. For example, McGrath (2002:22-25) offers the following two-level strategy:

- **Initial analysis:** made from the back cover and skimming through the materials for details (target learners, teaching context, publication date, cost)
- **Deeper analysis:** the examination of units and activities (aims, methodology, roles, syllabus order)

Ellis (1998:227-8) provides a template to describe a task that would be useful in analyzing potential materials:

- Input
- Procedures
- Language activity
- Outcomes

Ellis' template is useful for mapping out and comparing different types of activities/ tasks objectively. After the analysis is complete, an evaluation may begin. Cunningsworth (1995) provides four criteria for coursebook evaluation:

- They should correspond to learner's needs.
- They should reflect the uses that learners will make of the language.
- They should take account of students' needs as learners and should facilitate their learning processes, without dogmatically imposing a rigid 'method.'
- They should have a clear role as a support for learning.

This seems to be a much more useful and helpful list than Richards' list for textbook selection discussed earlier. It is much more grounded in the end-user's needs. Also, quick checklists have proven to be popular with evaluators. Here is an example (Figure 11) checklist from McGrath (2002:33):

Context relevance

suitable for course:

length of course Y/N

aims of course? Y/N

syllabus? Y/N

exam? Y/N

suitable for learners:

age? Y/N

level? Y/N

cultural background? Y/N

suitable for teachers:

-required resources (e.g., cassette recorder) available? Y/N

-evidence of suitability (e.g., piloted in a local context)? Y/N

Likely appeal to learners

layout Y/N

visuals Y/N

topics Y/N

suitable for medium-term Y/N

Figure 11: Checklist from McGrath (2002:33)

The next step after the quick checklist could be the more in-depth evaluation of two or more units in the coursebook (McDonough and Shaw, 2003) with these points to cover:

- The presentation of the skills in the materials
- The grading and sequencing of the materials
- Where reading and discourse skills are involved, is there much in the way of appropriate text beyond the sentence?
- When listening skills are included, are recordings 'authentic' or artificial?
- Do speaking materials incorporate what we know about the nature of real interaction or are artificial dialogues offered instead?
- The relationship of tests and exercises to a) learner needs, and b) what is taught by the course material

- Do you feel that the content is suitable for different learning styles? Is a claim (and provision) made for self-study, and are such claims justified?
- Are the materials sufficiently transparent to motivate both students and teachers alike, or would you foresee a student/teacher mismatch?

This is the most comprehensive list so far. Extensive lists can be found under other authors as well. They are all thoroughly planned out and well developed, but somewhat ironically, they seem like overkill and are asking too much of the busy teacher. Their overall practicality, therefore, comes into question. That is to say that dedicated researchers may find relevance in McDonough and Shaw's list, but I question how many classroom teachers would share the same enthusiasm?

3.2.3 In-use and Post-use Evaluation

As a somewhat easy compromise for teachers' evaluation plans, teachers can make logs (written or recorded) for in-use evaluation. Ellis and McGrath are proponents of this. Log entry ideas from Ellis (1998), focused on the learners and outcomes:

- How the task was performed
- What learning took place as a result of performing the task
- The teacher' s and the learner' s opinions about the task

McGrath (2002:181) takes a different approach, focusing more on the material design and functionality:

- What proportion of the materials was I able to use unchanged, that is, without needing to 'shape' (i.e., adapt) them in any way?
- Did the unchanged materials appear to work well? What evidence do I have for this?
- What spontaneous changes did I make as I taught with the materials? Did these improvisations work well? If not, what do I need to do differently?

Ellis also suggests post-evaluation log points (1998:230 in Tagg and Woodward, 2011). The points should:

- relate to the outcome of the task
- relate to whether the learners can perform the task (or a similar task) without pedagogic support (so suggesting whether any general learning has taken place).

Other ideas from McGrath include evaluation inclusive of learner feedback, learning gains, changes in self-confidence, changes in interest in the learning, and the teacher's natural enthusiasm for the book.

How can this data be collected? They can be obtained via questionnaires, discussions (interviews and conferences), and through various types of assessment tests. However, data collection such as in the form of surveys have serious drawbacks (Dörnyei, 2010:6-9). There is no perfect way to collect this information due to their subjective nature; results will change dynamically on a daily basis — all cognition, inclusive of questionnaire responses, is context dependent (Murphy, 2016).

Although this is not covered in the literature with much detail, another essential aspect is coursebook/syllabus *piloting*:

> *The syllabus needs to have tasks that students see as worth doing that lead to students being able to talk (read/write/listen) about their own ideas. A key to making this work is piloting. Good books involve activities and activities need to be tested – ideally by other teachers. It is great if you not have your knowledge of the material, it isn't the authors' right to expect it to be taught in the way they intended it.*
> (Helgesen, personal communication, 2012)

Taking this issue to heart, this current study fully documents the assessment of the piloting of the new syllabus design vs. the old syllabus design, incorporating a wide range of assessment types. The specific designs and methods used for the piloting assessments will be discussed in detail in Chapter 6.

3.2.4 Synthesis

Humans tend to evaluate what is placed before them. Similarly, teachers seem to evaluate their coursebooks spontaneously—as part and parcel of their job. However, there is a broad spectrum within which these coursebooks evaluations exist — from arbitrary and spontaneous evaluation to rigorously designed, officially funded scientific research. Types of evaluation can be a dichotomy, such as micro-evaluation vs. macro-evaluation. Alternatively, evaluation can be paced within stages such as pre-use, in-use, and post-usage stages. Teachers can make logs, or implement data collection schemes such as administering questionnaires, or by recording interviews and panel discussions. How much would constitute over-kill? This is a difficult question, especially considering that no data collection and analysis can be perfect. What can be said though is that the magnitude of the undertaking should be well-matched with the intended use of the data.

3.3 L2 Learning Material Adaptation and Designing

This section will critically analyze L2 material adaptation as found in the literature. It will also look at what the literature says about material design and the usage of authentic materials. Finally, it will offer suggestions for the future.

3.3.1 Adaptation

What is the adaptation of materials? It is the changing of material content to better match the needs of the students and/or the teacher. McDonough and Shaw (2003:73-76) stress that evaluation must take place before adaptation can happen. While this is unequivocally true, especially from a neuroscientific point of view, I am not sure how noteworthy their position is, considering this is the only way that it can occur. Nevertheless, the intended point must be that adaptation should be directed by the needs resulting from the evaluation of the materials. Reasons for adaptation of materials (an adaptation of Tagg & Woodward, 2011):

- Personalization needs
- Learning style differentiation needs
- Localization needs
- Modernization needs
- Simplification needs
- Needs to increase/decrease the challenge

How are these adaptations made? Here is a list from McDonough and Shaw (2003:78-84):

- Adding
- Deleting/omitting
- Modifying
- Simplifying
- Reordering

For most seasoned teachers, small to medium adaptations may be commonplace. Perhaps skipping over a section, or adding a personal side-story here and there may not seem like much deviation, but it still constitutes an adaptation. At the other extreme, McDonough and Shaw (2002:85) provide an elaborate pathway to material adaptation (Figure 12). While some teachers may find it to be overdone, understanding this flowchart and using it during planning stages, could be helpful for many teachers, especially novice teachers.

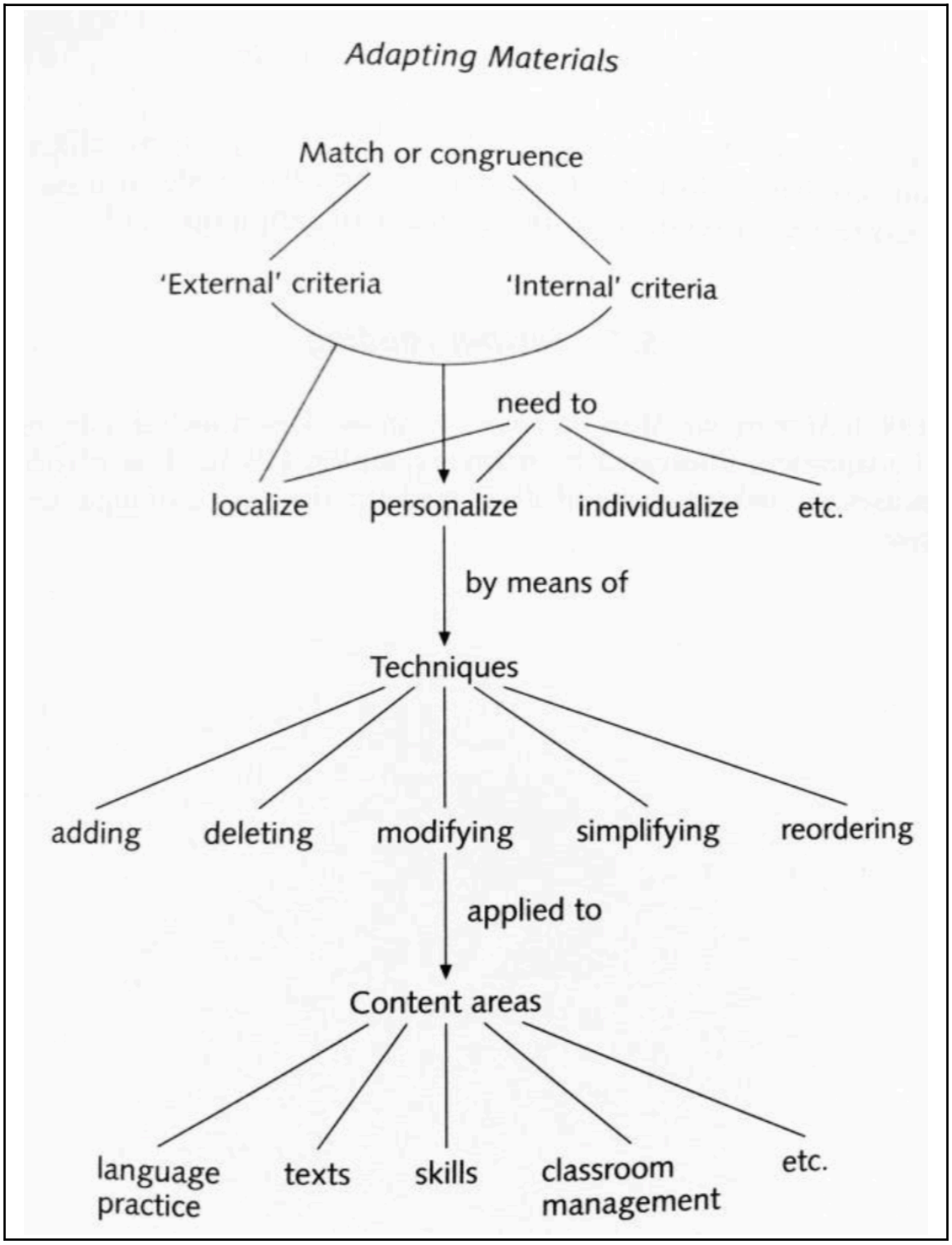


Figure 12: McDonough and Shaw's framework for material adaptation.

3.3.2 Designing materials

When material adaption is not sufficient, teachers may step into the realm of material design. What is the accepted definition of material design in the L2 literature? Researchers in this area such as Richards and McGrath take a rather narrow view of what this entails. For example, Richards (2001:252) defines it as "texts, photographs, video selections, and other teaching resources that were not specially prepared for pedagogical purposes", while McGrath (2002:103) defines it as "the selection of authentic texts and the preparation of tasks based on these." Suitable texts are readily available on the internet. According to Richards and McGrath, copying a passage from a magazine, or printing out a story from the internet, then having a task built around it, is a typical procedure for material design. I find this to be slightly narrow-minded if it implies that teachers cannot (or will not) muster the creativity to design suitable original content. But, I will stick to their definition for this review. Essentially, the definition is: the combination of *authentic material + task*.

3.3.3 Using Authentic Materials

As consensus in second language acquisition research tells us (such as in Ellis, 2012), studying non-authentic dialogues *does not* readily transfer to real-world applicability. There is also a consensus that the study of authentic language *does* help in the acquisition of natural aspects of language usage — if this were not the case, children would never learn their L1 before schooling, so, this is plainly true. However, L1 acquisition and L2 acquisition can have significant differences (Whong, 2011), mostly having to do with age, the context of the L2 exposure, and various motivational aspects (Dörnyei, 2009). Therefore Tomlinson (1998:13) notes that although authentic exposure is necessary, authentic exposure *alone* is not sufficient.

McGrath (2002) and Tomlinson (1998) discuss how then, the authentic text could be modified for L2 teaching. The modifications would entail stylistic changes and editing to match the target content.

3.3.4 Critique

How does modifying an authentic text to match target content differ from taking a synthetic passage and editing it to be more realistic? The diagram below (Figure 13) depicts how an authentic passage modified to meet the target language may produce nearly the same results as a synthetic passage that is modified to be authentic. This poses a dilemma and a solution. If catered to well enough, an authentic passage can be modified into a usable pedagogic passage. On the other hand, if a synthetic passage is edited well enough, it *can* become "authentic enough." This means that, theoretically, going about it from either side can produce good enough results (perhaps deemed mediocre results, depending upon how highly you would rate authenticity).

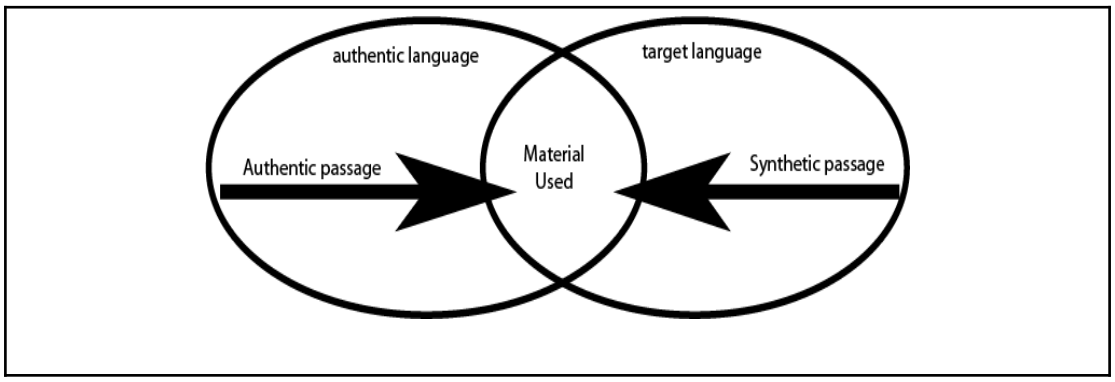


Figure 13: Editing an authentic passage vs. editing a synthetic passage

What criteria are there for selecting passages? Adapted from McGrath (2002:106):

- Relevance of topic
- Intrinsic interest of topic/theme
- Cultural appropriateness
- Linguistic demands
- Cognitive demands
- Logistics (length, etc.)
- Quality
- Exploitability

All of the above criteria should match the teaching goal well, with the final point, *Exploitability* being the most relevant. That said, rather than having exploitability as one of the criteria, perhaps the sum of the other seven should be calculated into an *Exploitability Quotient.*

3.4 Chapter Conclusion: Looking toward the future

3.4.1 Toward conclusions

In the *L2 Learning Materials and Evaluation* section, I examined and discussed six pragmatic purposes for L2 coursebooks. The discussed purposes were: *input, structure, motivation, alignment, reformation,* and *cost-performance ratio*. Although these purposes are undoubtedly pragmatic, I uncovered four major negative points regarding coursebooks and their usage:

- Humpty-dumpty postulate (dismembered parts not becoming a whole)
- low-balling
- false advertising
- inappropriate compromises

I then went on to discuss published coursebook evaluation procedures, many of which I found to be overly detailed and impractical for busy classroom teachers—although I agree that some form of evaluation is necessary. To help fill that gap, more "user-friendly" evaluation processes are sorely needed.

3.4.2 What has been uncovered? What can be done?

Helgesen has this to say about progression in this field:

> *One of the Catch 22's is that, of course, the publishers need to make money. And it is expensive. It might cost a half-million dollars to launch a multi-level coursebook series. So that makes them risk-averse. But risks are the only way you make progress. (2012)*

In continuance with the pursuit of best practices for a new syllabus, what has been uncovered in the current chapter?

- Linear language learning is a fallacy → learning materials need to implement cyclical learning structures
- Focus on grammar alone does not provide favorable results → place more emphasis on *social* learning, but do not remove all grammar exercises
- Static syllabi tend to disregard individual learner needs → use a dynamic (flexible) design
- Similarly, top-down syllabi disregard individual learner needs → use a dynamic bottom-up (flexible) design
- Coursebooks falsely advertise → don't be naïve; learn to analyze; learn to improvise

Gathering from the literature, it seems the future of material design must become more flexible yet cyclical, and become more social in the classroom (but not forgetting grammar exercises altogether).

Coursebooks should *not* be general; they should be specific—yet, they often do the opposite. Considering this, is it a coincidence that some teachers and researchers say we are now at a *beyond methods* stage (Bardovi-Harlig et al., 1997) and/or that we no longer need coursebooks? Moreover, as Allright discusses (1984:3), there is not a clear relationship between what is *taught* and what is being *learned*.

Material designing also must consider how the material fits or does not fit into the bigger picture. The materials may be designed well enough to work as small learning chunks, but how well do they fit into the school's syllabus or curriculum? Schools need to plan out *syllabi* for a plethora of reasons ranging from academic reasons to purely bureaucratic reasons. Considering Chapter 2 and this current chapter, what recommendations surface regarding the syllabus? The ideal syllabus would have to be a *dynamic* syllabus that could eclectically string together all the best ideas uncovered thus far—into a *thematically cohesive* and *theoretically congruent* package that somehow also caters to the students' needs. Ideally, all material designs must be able to competently support that dynamic syllabus.

3.4.3 Discussion: Designing cutting-edge materials

Tomlinson (2008:321) closes his critical review on materials with this request list for material writers and publishers:

Requests for Future Developments in ELT Materials

- pay more attention to how to facilitate intake and less attention to the presentation of language points
- aim for more affective and cognitive engagement
- provide more opportunities for outcome-focused communication
- be much more flexible in providing opportunities for choice of texts
- help teachers to achieve class specific adaptation

These requests have a common goal — to somehow raise and sustain student motivation. Therefore, this is a perfect segue to the following discussion on motivational factors.

3.4.4 Motivational factors

Three major factors in the motivation literature stemming from Deci & Ryan are *autonomy*, *competence*, and *relatedness* (such as in Dörnyei & Ushioda, 2011:33). The concepts, dressed in a different suit, have reached pop status. An example is the work of Pink (2009), as expressed in his popular TEDtalk video, with roughly 27 million views as of April 15, 2018 (TED + YouTube combined). For general public consumption, the terms have now been slightly modified to *autonomy, mastery*. The term modifications may have mixed reviews from academics in the field, but the change likely makes the list better suited for the general public.

From a different vantage point—my research in neuroELT—I arrived separately at five points of advice for teachers (Figure 14). Although I arrived at the five points separately, there is a perceptible overlap with the ideas. I believe my five points (adapted from Murphy, 2012) are more readily applicable to the EFL classroom. This is because they were developed with the EFL classroom in mind, whereas Deci & Ryan, and later Pink, were focused on more general psychology, and general consumption.

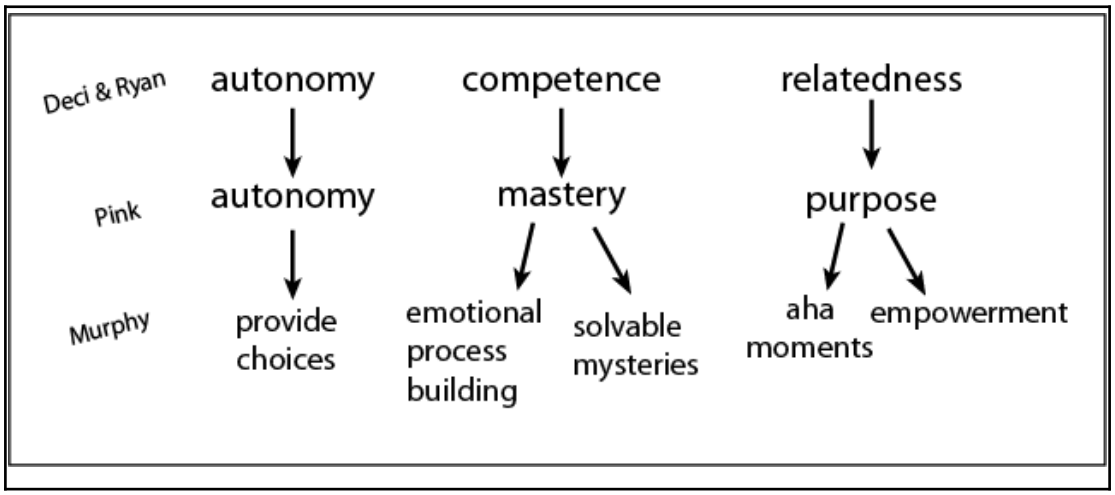


Figure 14: The relationship between Deci & Ryan, Pink, and Murphy's work on motivation.

Raising motivation can equate to raising learnability as long as the content is appropriate for the learners *and* the learner is ready for the learning. From a concerted effort in *neuroELT* research (Murphy, 2016), it has been uncovered that:

- ***Providing choices*** (even **false choices)*, raises motivation through a feeling of ownership for the chosen activity. Teachers can provide choices per activity to raise motivation and direct ownership of the learning to come. (**False choice example:* The provision of three choices, while you know that two of the three choices are not going to be appealing to the students; students still benefit from deliberating and making the "right" choice out of the three.)
- ***Emotions*** are central to all learning. Learning is a *build-up process*. Boring information goes nowhere in the brain. Teachers much embrace the idea that *emotional process building* is at the core of all effective learning and teaching.
- ***Solvable mysteries.*** As in countless accounts in real life and in literature such as Vygotsky's ZDP work (1978) and Krashen's i+1, all learners feel good when they can solve a mystery that is just beyond their immediate understanding. This is why puzzles and games provide hours of entertainment for suitably challenged people. This concept should also be embraced by teachers.
- ***Aha moments*** are the natural dopamine and/or serotonin rushes when we "see the light." That is, when separate ideas come together to form a new understanding, usually a higher-level understanding, we naturally feel good about it. *Aha moments* are at the center of all higher-level learning and understanding
- ***Empowerment***. My usage here connects to all three of Deci and Ryan's points. That is, my meaning of empowerment calls for peer-to-peer assessment, and peer-to-peer teaching. This gives the learner a sense of autonomy, competence, and relatedness all at once. It makes the learning significantly more meaningful (and naturally more dynamic).

The five points should be logically integrated into new material designs, or in the very least, be instrumental in the development of new material.

3.5 Learning lessons from science education literature

3.5.1 Difficulties in changing teachers' beliefs regarding pedagogical assumptions

When reading articles from the huge domain of *Education*, it becomes obvious that at times the field of *Applied Linguistics* pursues new issues that have already been settled in *Education* research (which has a much larger body of knowledge and much longer history as a discipline). The opposite is often the case too → *Applied Linguistics* finds solutions to issues that *Education* has yet to notice. With that in mind, what can we learn from the domain of *Science Education?*

Firstly, how do teachers view science as a discipline? The conclusions on this are noticeably divided. Watson claims that, *"the self-conscious effort to define science education as a discipline is a fruitless search for the Holy Grail"* (1983:264). He gives the example of engineers—and how they would not waste time debating whether engineering was a discipline or not, saying that society *"accepts them as professionals who apply their disciplined skills to a variety of social decisions"* (1983:264).

Watson's claims correspond well with issues that many EFL teachers have. Accordingly, Watson argues that it is the context and society that decides whether science is a discipline or not. He is saying that science education should be perceived as a social science. This would account for why views have been so divided—we are only human; our responses are naturally contextual. Therein lies the issue. Teachers' beliefs are contextually nurtured over long periods of time and are therefore often diehard beliefs (due to the long history that nurtured those thought processes). This stubborn-due-to-biological-pathways issue can be highly problematic if you want veteran teachers to adopt, or even seriously consider, new pedagogical approaches. In such cases, it may be helpful if teacher training courses take the time to firmly discuss the key issues within this chapter (with the understanding that resistance to change is natural), and the two tenet lists that are provided in Chapter 9.

What have other science education researchers said that will resonate well

with applied linguistics issues? Though similar to Watson's views, Fahnestock frames this issue from a "*rhetoric of science"* perspective. From her vantage point, the goal of science is, *"…to understand the means, the how of texts, movements, or argument fields that recruit adherents for their claims about the natural world"* and therefore, *"…a rhetoric of science is most closely allied to the history of science"* (2005:278). Fahnestock continues,

> *…But it also requires knowledge of the formal resources of argument, language, and other modes of presentation or construction available to the scientist. This intersection of the historical and the formal is the rhetorical… the rhetoric of science involves applying rhetoric, either in a contemporary or perennial form, to science. But in the process, the terms on both sides of the preposition in the phrase "rhetoric of science" come out somewhat changed. A rhetoric of science puts pressure on both science and rhetoric.* (Fahnestock, 2005:278)

With her alluring argument, Fahnestock stresses a particular need to emphasize the visual rhetoric of science. For example, "observational and experimental methods are frequently designed to produce the least mediation between a source and its sensible sign" (2005:284) — such as biologists with DNA sequence tagging, and neuroscientists with color-coded fMRI scans (Figure 15). Although her argument may be appealing, it is not perfect; Fahnestock admits, "studies in the rhetoric of science tend only to demonstrate how cleverly rhetoricians can read texts rather than how scientists read them" (2005:284). These are pertinent issues for the language teachers that I work with, and the reason for highlighting them here; it is similar to the dichotomy of how students see the L2 vs. how the teachers sees the L2. Moreover, it also calls to mind the differences between the coursebook author's intent (visual rhetoric), the teacher's perceived intent of this, and the students' personal interpretations of both. The core question for coursebook authors is → what is the relationship between the rhetoric of the textbook's L2 and real-world usage occurrences of the L2?

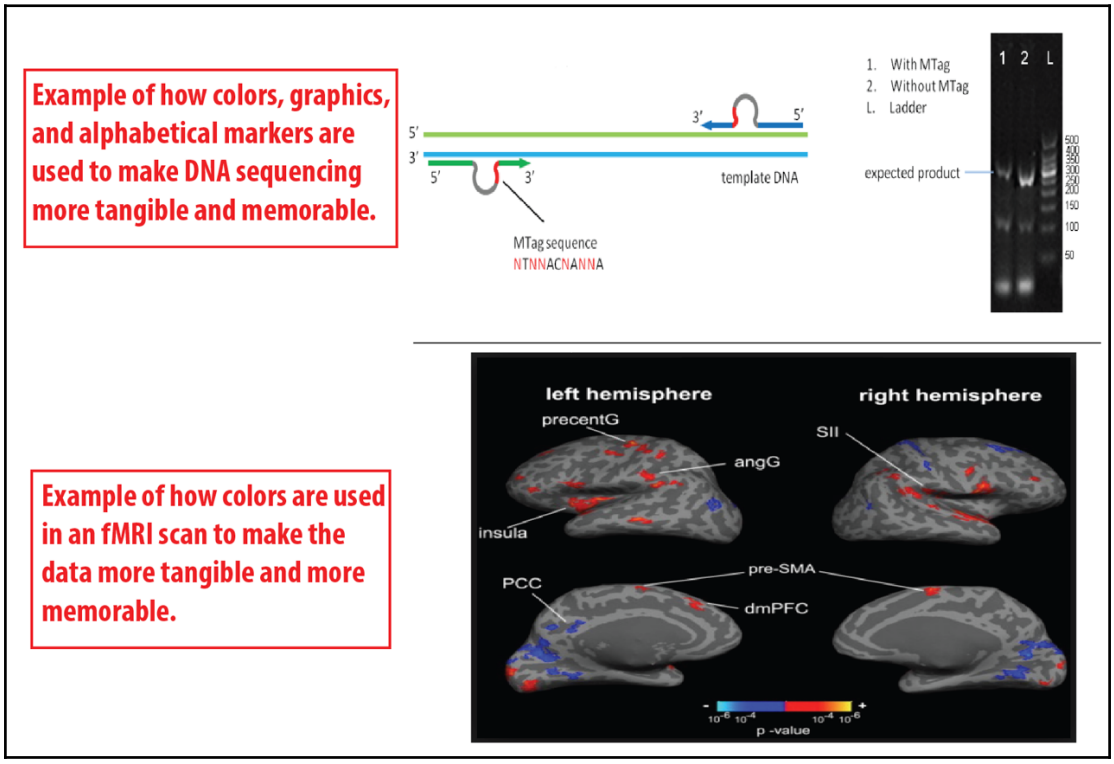


Figure 15: Example of DNA sequencing and color-coded fMRI scans

Whom else might agree with Watson's and Fahnestock's positions? Lederman may agree, considering the emphasis he puts on epistemology (1992), cultural aspects, and creativity (2002). And if so, should these areas be put into EFL coursebooks to compensate? Furthermore, Kimball (who found little difference in the effect of academic backgrounds/experience in the level of naivetés regarding their understanding of "the nature of science") (1968), may likely agree with Fahnestock and Watson. Erduran and Dagher (2014) pursue the complexity of this issue by heavily emphasizing the effects of social organizations, along with politics and financial systems (Figure 16).

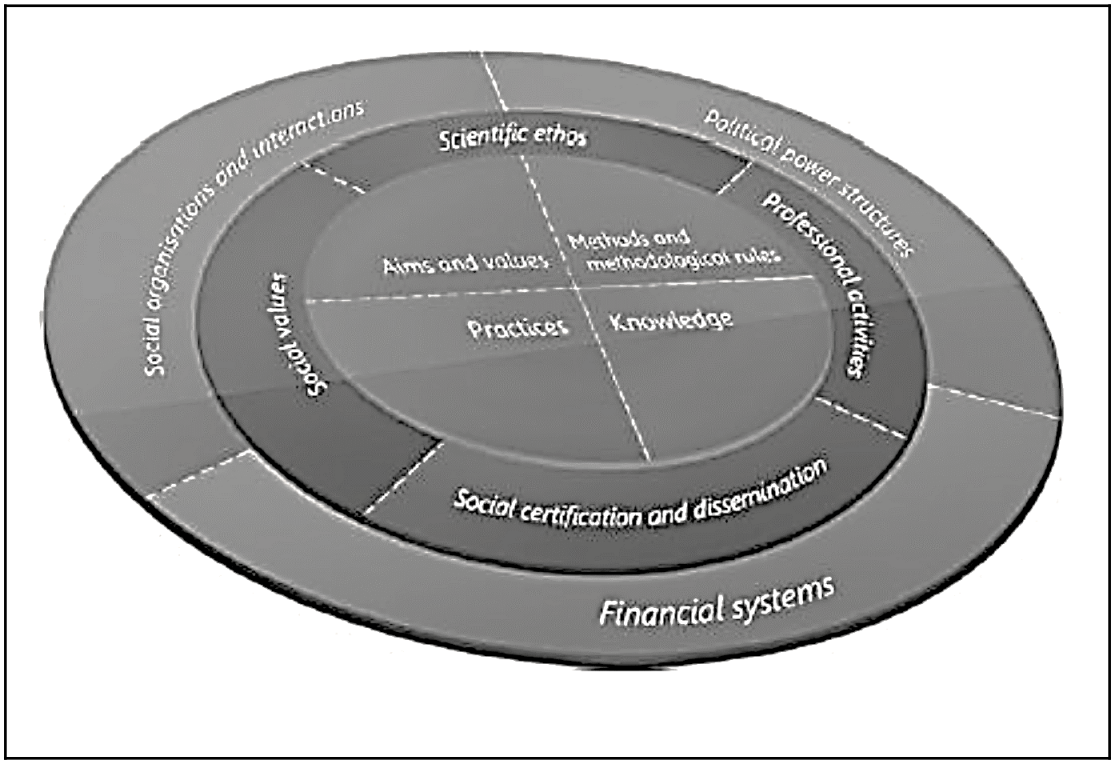


Figure 16: Complex layers of effect surrounding the "nature of science" (Erduran and Dagher, 2014)

From the literature in this section thus far, it seems logical to assume that culture and context may have a significant influence over *science as a discipline*. This argument surely holds true for Applied Linguistics as well. In addition, Abd-El-Khalick (2005) found that explicit teaching for change will only elicit a temporary solution, whereas, Nott and Wellington (1998) argue that what they call *critical incidents* (triggers) are a means to a change within teachers. They explain how this can be linked to professional development—and how it is via professional experience and practice that teachers can develop further. The affect of these critical incidents (triggers) and of *critical breaks* (breaks that impede the flow of learning) will be explored throughout this thesis.

The takeaway

It would seem that many teachers' views and beliefs are indeed contextually and culturally-biased. This makes it hard for teachers to change. Purpose-designed implicit interventions may not work as well as (unplanned) critical incidents and/or cultural and contextual paradigm shifts described by Kuhn (1996) that presumably affect the teacher over a significantly longer period of time, and/or in a significantly different manner. When this occurs, it has the potential to negate pendulum effects and outright negativity that may ensue when sudden changes are expected of people, as discussed in Evans (2001), *The Human Side of School Change: Reform, Resistance, and the Real-Life Problems of Innovation*.

In sum, authors of coursebooks that are radically different from the norm (and radically different from EFL teachers' norms) need to be wary of two things: (a) there is a strong possibility that the coursebook will be rejected outright, and (b) the more training a teacher receives regarding the new approach [at a deliberate pace], the lower the chances are of outright rejection, while raising the likelihood of a paradigm shift → which helps to raise the potential of long-term adoption of the new pedagogy.

3.5.2 Science education: teachers' beliefs/views of science as a school subject

Should science be rote-learned, or more "hands-on"? Ausubel (1977) makes the distinction between "reception learning vs. discovery learning" and "meaningful learning vs. rote learning". Unless the goal is to teach to the test, the literature gathered below suggests that "hands-on" and "discovery" methods are the better choices. I have created a graphic to clarify this point further (Figure 17). The upper-right corner would seem to be the "sweet spot" of good teaching.

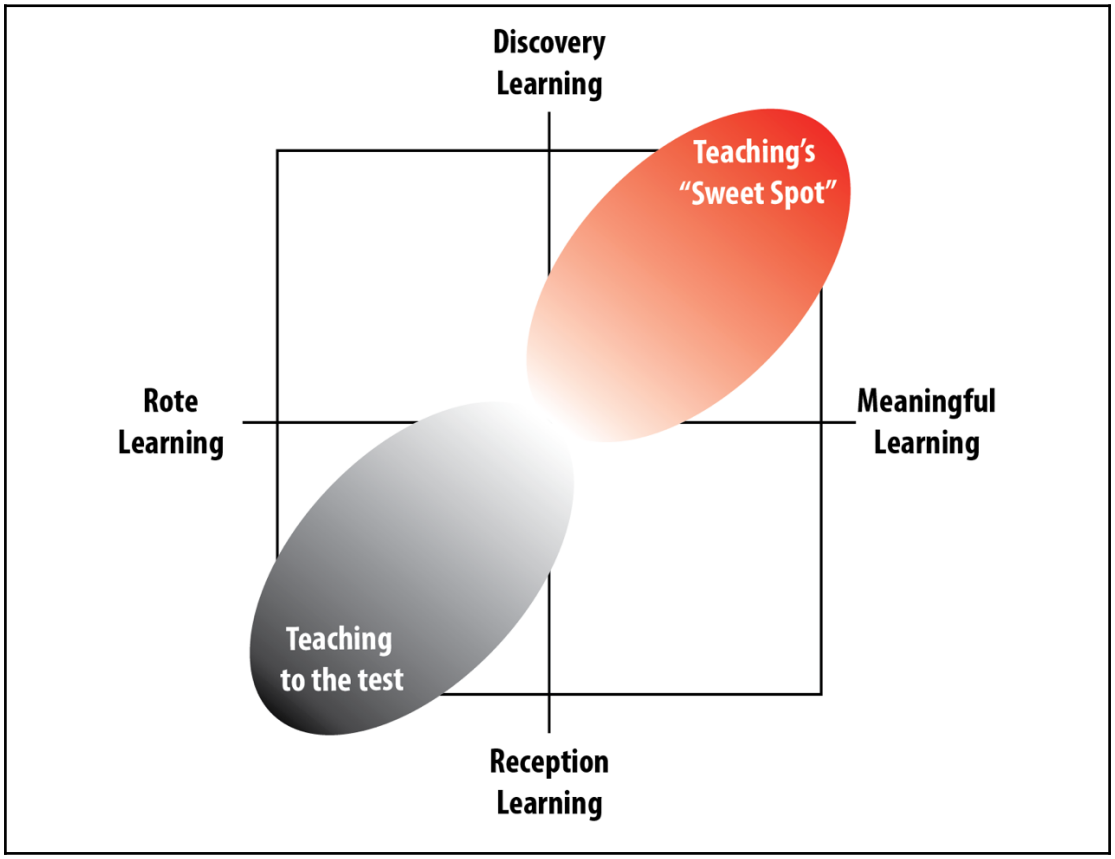


Figure 17: My graphic representation of the sweet spot of teaching (based on Ausubel, 1977)

UDL (Universal Design for Learning) guidelines put forth by CAST (in Ralabate & Nelson, 2017), discuss *Engagement*, *Representation*, and *Action & Expression* as three key principles of learning that should be incorporated into any classroom. The guidelines set forth by UDL compel teachers to abort simplistic lecture-styled teaching. It shows how to cater to student needs dynamically, and gives teachers the confidence to move away from traditional methods. Methods that may be detracting from deeper learning potential (Posner, 2004). This is fully compatible with my graphic representation in Figure 17.

3.5.3 What other issues are associated with top-down/for-the-test methods?

The brain is a *social organ* (Tokuhama-Espinosa, 2014:33)—it thrives on real life interactions, and gets bored when not engaged. Most importantly, *"Students can, and often so, achieve learning without a teacher"* (ibid.). So, it might be better if the teacher 'steps down from the podium' and allows autonomous learning to happen within their students. Moreover, *"The brain remembers best when facts are embedded in contexts that are authentic for the learner"* (Tokuhama-Espinosa, 2014:34); she argues that student-centered, problem-based learning activities are essential to achieve optimal learning experiences. Yet, to add a historic perspective, these suggestions firmly go *against* what the social norms were for science teaching until a few generations ago (Donnelly, 2005). Fast forward to today – we are now 18 years into the 21st century, and yet we are still discussing the implementation of 21st century skills, as asserted by Jenkins (2007), and more recently, Liebenberg whom—*in 2017*—questions whether we have lost sight of the meaning of education (Liebenberg, 2017).

Watson's earlier *"science education as a social science"* statement is ever coherent in this context: What good is rote learning if you can easily *Google it* on your phone? Why take a science class if you can find the facts online? Tsai, being a proponent of *active learning* and student *engagement* (Tsai, 2002), would likely agree.

Here I have reviewed literature that, when taken together, suggest that classroom teaching, now more than ever, should attempt to minimize rote-based teaching. It should encourage teacher-facilitated, autonomous learning that leads to critical skill development for the real world—a world with a superabundance of information that is causing what Schmitt et al., call IO [information overload] (2016). So, although there may be a divide between teachers, moving toward "active" or bottom-up teaching seems to be more desirable for long-term benefits—*especially in our age of IO*.

How can this transformation be successfully realized? Murphy (2016) suggests requiring the teachers to acknowledge that there *is* a balance to be had between bottom-up and top-down teaching—so that they can make a conscious decision regarding optimization of top-down vs. bottom-up ratios in their classrooms. Furthermore, my Cartesian chart (Figure 17) may be a suitable catalyst for furthering comprehensive discussion during transformative exercises with coursebook writers and the users of the to-be-produced novel coursebooks.

3.5.4 Teachers' beliefs/views: Top-down vs. bottom-up?

To add further critical thinking and external views into the mix, I shall compare teacher beliefs to what researchers say about the *neurobiology* of learning. In lay terms, we say we "learned" something if we can demonstrate that we can come to the same conclusion as before, relatively quickly. If we require a long time to come up with solutions, we say are still "in the learning process", or, we are "trying to remember". There are four main factors in the biology of learning: (1) the amount of emotion attached to the stimuli, (2) how often the task is called upon again, (3) general health, and (4) sleep (Cosolino, 2013; Sousa, 2016; Kelly and Murphy, 2017; Murphy, Stubbings, and Uemura, 2017).

What do science education researchers say about *top-down vs. bottom-up teaching*? Scaife (2008) may be the strongest proponent of more bottom-up teaching; he positively discusses Vygotsky and Bruner, while even differentiating constructivism with *radical* constructivism—the latter of which he admits to being a *"strong supporter"* of (2008:74). Tsai reported on the popularity of traditional *top-down* transmission of information to students by experienced science teachers in Taiwan (2002), but goes on to discuss the phenomenon of *nested epistemologies*—meaning that teachers can hold, and are able to differentiate, microcosms of distinctive epistemologies, based on the subject matter. Kagan (1992) reported a virtual smorgasbord of inconsistencies in the studies within her meta-analysis, hinting that teachers are indeed humans too.

My interpretation is that teachers are:

A. affected by their past
B. affected by their real-time context
C. prone to say one thing, but do another

3.5.5 Views/beliefs vs. classroom realities

Rather than re-treading the overlapping issues from the above sections, here I propose to accept that teachers *are* "human" (not computer-like), meaning that they *in fact* do not always do what they set out to do, nor do they consistently adhere to their self-proclaimed beliefs. With that in mind, I will use this subsection to bring to light a wider range of studies that help elucidate the phenomena.

Bandura (1982, 1997) discusses the self-efficacy mechanism in human agency. He puts forth the notion that *perceived self-efficacy* helps to account for coping behaviors produced by various modes of influence. Moreover, perceived *collective efficacy* is what brings on social change (and when people do *not* perceive self-efficacy, it reflects negatively in their behaviors). So, rather than limit this section's scope to a simplistic realization that teachers' results in the classroom are incongruent with their self-proclaimed views, I believe the discussion will be more captivating by taking a few steps further and examining the inherent mechanisms at play—particularly, the ones that are not habitually noticed. In this case, the influence of *perceived* self-efficacy (rather than: [1] an assumed level of efficacy from the administrators' perspective, and [2] the actual amount of efficacy that the teachers have, but leave untapped) seems to be a key influence on *teachers'* compensation strategies, and therefore what they end up doing in the classroom. I suspect that these aspects—inclusive of *on-the-fly compensation strategies*—are not always well-remembered by teachers, and have less of a chance of emerging in post-lesson interviews.

Below is an example from Erduran and Dagher (2014) on how context and perspective radically affects deduction; even the most valiant attempts from teachers can *easily* be foiled by contextual shifts (Figure 18).

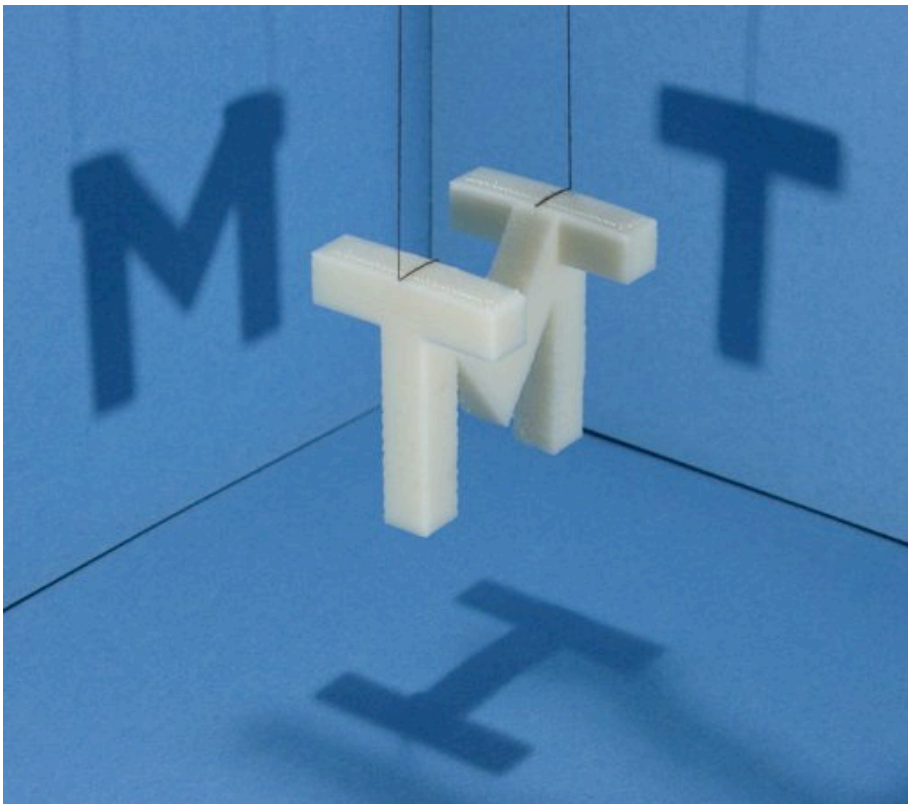


Figure 18: Immediate context changes what you see. From Trevor Shannon (in Eduran and Dagher, 2014)

In this light, it becomes difficult to cast doubt on the understanding that context and biology will invariably affect outcomes. However, the implementation of *high support* contexts can elicit higher-quality responses, with more congruence (Kennedy, 1994; Cheng, 1998; Fischer, 2008) [See Chapter 5]. The phenomenon of *dynamic reliability* can be used to explain the incongruences between the studies of Brickhouse (1989) and in Lederman and Zeidler (1987), with the former finding consistencies between teacher beliefs vs practice, while the latter found the opposite to be true.

The takeaways for materials design

- teachers' views and beliefs are contextually-biased and culturally-biased
- bottom-up (discovery-based and meaning-based) teaching/facilitating is likely better tuned for 21st century skill development
- teachers may say one thing, but do another
- we should expect inconsistencies across studies because of methodological errors, incompatibilities, and *dynamic reliability*
- high support contexts can help cut though these inconsistencies and help achieve optimal results in interviews

The next section shall shift to the discussion to the much-debated realm of Pedagogical Content Knowledge and its implications for teaching.

3.5.6 Pedagogical content knowledge (PCK)

PCK (Pedagogical Content Knowledge) was initially put forth by Shulman (1986) as a way to reconcile the commonly occurring divorce of pedagogy from content. Lee and Luft (2008) later found inconsistency within the field regarding PCK's definition. When Shulman revisited PCK in 2004, and again in 2015, he found it amusing in how "his child" had inevitably grown into something new (Shulman, 2004). Initially, PCK was defined by Shulman (1986) as part of this trio:

- Subject Matter Knowledge
- Pedagogical Content Knowledge
- Curricular Knowledge

Shulman and others later made adaptations and took liberties in the definition and the significance of these points. Figure 19 is a chart from this course by Childs (2018) illustrating a sampling of the wide variance across researchers in the field:

Reference	a = a distinct category in the knowledge base for teaching; b = not explicitly discussed, PCK = pedagogical content knowledge							
	Subject matter	Representation and strategies	Student learning and conceptions	General pedagogy	Curriculum and media	Context	Purpose	Assessment
Shulman (1987)	a	PCK	PCK	a	a	a	a	b
Grossman (1990)	a	PCK	PCK	a	PCK	a	PCK	b
Loughran et al (2001)	b	PCK	PCK	b	PCK	b	PCK	PCK

Figure 19: Comparison of PCK definitions of Shulman, Grossman, and Loughran et al. (Childs, 2018).

What is relevant to this review section is the notion that there is a long-standing acknowledgment that teaching content and pedagogical designs should be considered in tandem, regardless of that fact that there is no consensus in the field regarding definition specifics. The over-arching question is—how well do coursebooks (their pedagogical designs) link with the teaching content? It seems this is a question that is not given enough weight in the EFL materials writing literature, so as with the above sections, I have turned to the broader field of *Education* for insight.

3.5.7 Using Schulman vs. Gardner as a metaphor for a deeper understanding of PCK

Schulman, as the "father" of PCK returned to the concept in 2015 to write the introductory chapter for a compendium on PCK. In his return, he did not write as a champion nor a savior; he was slightly apologetic in tone as he systematically updated and tweaked the original construct that, as he stressed, was borne from a specific point in time, to address a specific issue at the time. Among other things, he noted that he was pleased to see that PCK seems to have taken a life of its own:

> *Now that is has survived adolescence and reached emergent adulthood, most of you know far more about PCK than I possibly could… I have been caught up with a number of other topics in the last 30 years… PCK certainly has its weaknesses, and I trust that many of you have been shoring up those deficiencies…* (Schulman, 2015: 3-4)

This attitude toward a seminal work that has been cherished for decades is not unlike Gardner's position toward his fatherhood of *Multiple Intelligences* theory (MI) from 1980. However, there are two significant differences: (1) While Gardner fully notes that MI was born out of a need from a specific point in time [to rectify how IQ testing was "criminally" reducing human cognition to a single number], and although he also made updates to the original publication, he is *un*apologetic. In fact, he strongly defends MI in the face of strong criticism (Murphy, 2011). And, (2) rather than allow his "child" to be nurtured by foster parents as Schulman puts it, Gardner never lets MI venture far from home no matter how hard people try to nurture their own understanding of it.

Both Schulman's and Gardner's work have left a strong and ever-growing impression on *Education* over the past few decades. Yet, their approaches toward their gift to society have been vastly different; this phenomenon and difference serves as a good metaphor for PCK (as Schulman describes it). That is to say that PCK must be "domain specific and contextualized" (Schulman, 2015:7).

Continuing on with my metaphorical usage, let us now view Schulman and Gardner as two teachers with the world as their respective classrooms. For PCK to make sense for the world (Schulman's classroom), PCK must be flexible enough to adapt to contextual factors, as it is a proposal for the integration of knowledge and teaching methods—such needs are unknown. On the other hand (in Gardner's classroom), MI's mission has nothing to do with pedagogy, but all to do with changing the way the world views *human intelligence*; priority is on the way IQ test scores are (still being) misused and misinterpreted—Gardner's proposed need is highly specific and very clear. In this "classroom", Gardner is forced to take a rigid stance toward MI teaching because his teacher-students all-too-quickly take MI as a way to refine and define pedagogical standards (I.E. designing their pedagogy with MI in mind)—and go so far as to mesh MI theory with the now thoroughly de-bunked *learning strategies* approach. This confabulation occurs even though Gardner maintains that MI is not a prescription for pedagogy, nor even a template for the classroom. So, in this way, the "teaching strategies" employed by Schulman and Gardner are significantly different—based on both the *domain* and the *context*—even though both Schulman and Gardner have expert knowledge of their respective areas.

In this way, this topic serves as a metaphor for the PCK construct. The domain and context of PCK has allowed Schulman to move on to other research areas while concurrently allowing PCK to mature into adulthood. The domain and context of MI (an all-too-easily misconstrued concept) has, conversely, required Gardner to keep up a line of defense for the ever-misconstrued understanding and usage of MI.

This metaphor can be concluded with two images: one of a teacher saying "Ok, now I want you to take the knowledge and make it useful for yourself", and another of a teacher saying, "No. That's not what I said. Please have a closer look – then, maybe we can talk about it next time." This strongly reinforces Schulman's position regarding how we must realize how strongly context and the domain affect teaching methods—often regardless of the teacher's intent. It also calls to mind the issue of readiness of the student (as in Piaget's stages), and makes me wonder why MI is so often construed as a sort of template for teaching (often as a misconstrued version of Gardner's intent). The answer to this is likely embedded in the theory of MI itself—regardless of the general difficulty of any concept assessed under a Piagetian scale, some people have higher or lower aptitudes for specific domains of knowledge. This is a significant issue that ties the various points of this section together; it is difficult to control what a student learns from the provided knowledge unless it is being semi-controlled by both parties (both the teacher and the students). The implications of this will be widely explored in the second half of this thesis.

3.5.8 PCK mindsets: "Top-down vs. bottom-up" mindset in lieu of Shulman's mindset

In reading Schulman's revisiting of PCK (2015), it became clear that one of the driving factors behind the inception of PCK was the discovery of, and his fascination with, how different contexts and domains give rise to different teaching strategies. While I fully agree that contexts and domains significantly affect teaching, I believe Shuman's explanation of this is overly complex.

It is true that there is a mind-boggling plethora of domains and sub-domains to consider within the realm of education—if one were to attempt to define all pedagogical differences from that perspective. But, that approach has a noteworthy downside; you can never map out an infinite number of domains.

Is there not a way to grasp the entirety of PCK without taking this complicated route? In Figure 20, I have graphically organized what I see as the spectrum of approaches that are possible in teaching. The triangular graph displays the spectrum that spans from: (a) an apex that depicts the most rigid teacher-centered, lecture-styled teaching possible, to (b) the widest-ranging dynamically-changing teaching style possible. Within this gamut there are three broadly defined zones: top-down teaching zone, "bag of tricks" zone, and bottom-up teaching zone. PCK resides in all three zones—this is the most poignant aspect of my graph.

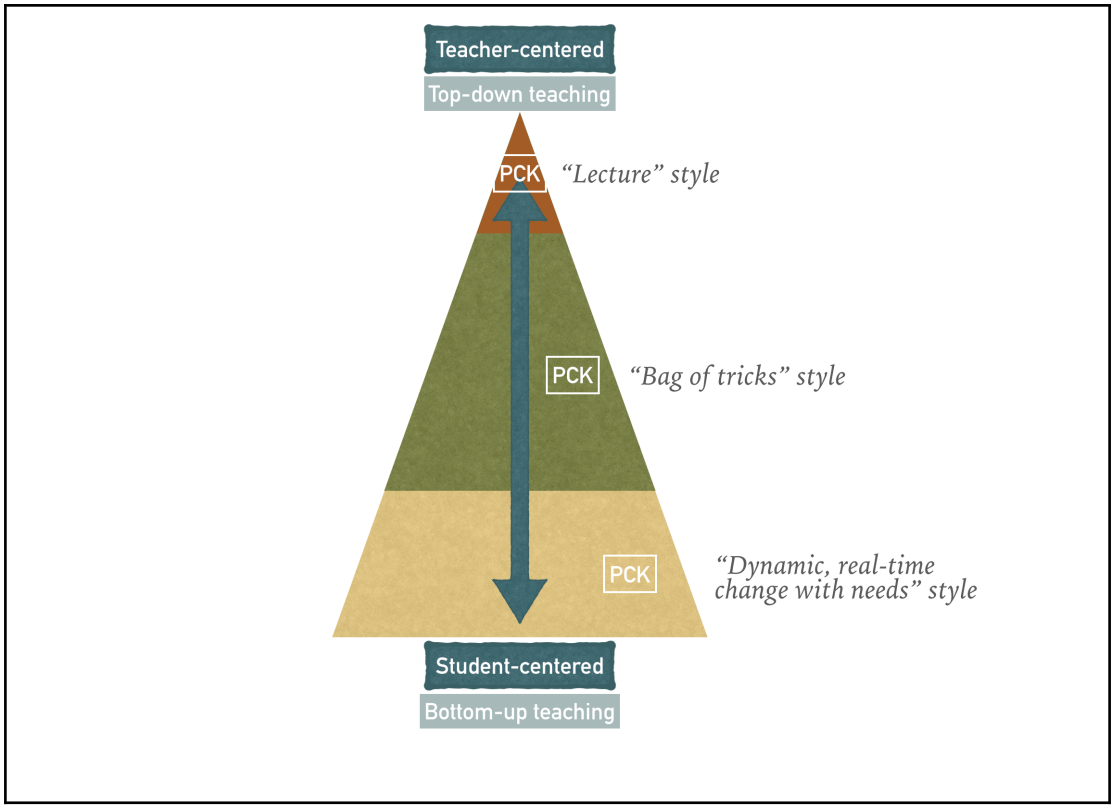


Figure 20: Adding PCK to the spectrum

What is the significance of noting that PCK resides along the entire spectrum? I am not being critical of Schulman for the sake of critique, but because after reading Schulman (2015), I sensed Schulman's discomfort as he discusses what he calls *Signature Pedagogies* — "…profession specific modes of teaching that we associate with that profession" (p 11). There is discomfort in the fact he was (understandably) not able to define all of the distinctive features that every domain possesses via their signature pedagogies. He concludes, "What's even more fascinating (and disturbing) for us and teacher educators is that most of those deep structure principles of the signature pedagogies of the professions are absent in most teacher education programs." (Schulman, 2015: 12).

However, regardless of domain, all forms of teaching fit within the range that I have defined in Figure 20. This is true even considering the implications of MI theory (having a strong aptitude in one domain does not necessarily correlate with aptitudes in other domains, nuancing the need for flexibility). Therefore, PCK's signature pedagogies need not be shockingly different per domain as Schulman alludes to. While all teaching belongs at some place on this singular spectrum, specific domains do require domain-specific content for their respective lessons (giving rise to those *Signature Pedagogies)*. Teachers of different domains learn to nurture their *signature pedagogies* along this spectrum, and dynamically move up and down within that gradient as needed (whether they realize this or not). In contrast, Schulman painted a picture that conveys an infinite number of domain-specific pedagogies, conjuring up an *unmanageable* impression—but I strongly doubt that *signature pedagogies* vary as much as Schulman worries. The simple proof being, that all forms of teaching are recognizable as teaching; there are no domain-specific forms of teaching that are unrecognizable as teaching to teachers of other domains. So, putting all teaching on this single recognizable spectrum sums up the reality of PCK more practically and more realistically.

Moreover, Shulman found it "disturbing (that signature pedagogies were) absent in most teacher education programs", but as long as all teaching is recognized as being on this spectrum, it would be far less disturbing; the challenges of professional development courses seem much less challenging if we only need to focus on this singular spectrum.

In sum, the mindset that is required to see signature pedagogies as vastly different constructs *per domain* is overly complicated. Schulman likely finds it to be disturbing because of its unbounded loom within his mindset. On the other hand, the mindset for a singular spectrum of teaching, with signature pedagogies *residing within the spectrum*, simplifies the discussion to "where the teaching is" within that singular construct; the recognition here is that there is a smaller number of archetypical teaching approaches unto which content specifics are naturally attached to.

3.5.9 The Knowledge Quartet (KQ)

Rowland et al.'s (2005) grounded theory study of maths lessons led to the development of the *Knowledge Quartet* (KQ). KQ is commonly used as a framework for the analysis and development of teaching even outside of maths. Being a quartet, there are four quadrants in the theoretical framework (Table 1) that were developed by coding data from maths teaching contexts:

Table 1: The tabulated Knowledge Quartet

Foundation	*Transformation*
Connection	*Contingency*

Each of the KQ quadrant categories is typically used to focus attention on that particular element of teaching—*to better understand that element*. Being developed in the context of maths teaching by Rowland et al., it is not readily applicable to other subject matters—Rowland concurs, as he suggests modifications if attempting to use the KQ for other subject matters. This framework would have broader utility if it is: (1) trialed extensively in fundamentally different teaching contexts, and then (2) more dynamically accepting of additional elements—based on data from those fundamentally different contexts. The current study may help shed some light on this.

Mortimer and Scott: Talk in the classroom

Mortimer and Scott (2003) published work for the analysis of language usage in the science classroom. Because interaction with students can be tricky, Mortimer and Scott (2003) proposed a framework for analyzing and planning science teaching interactions (Figure 21). The framework consists of focus, approach, and action sections, with proposed aspects of analysis per section. In this section, I will briefly discuss what they mean by Content, Communicative approach, Patterns of discourse, and Teacher interventions.

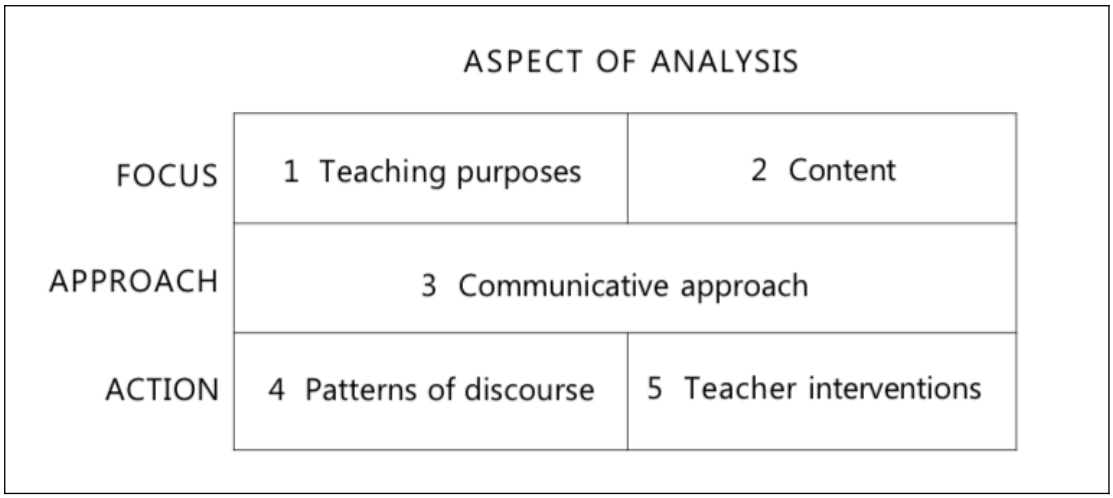


Figure 21: Mortimer and Scott's interaction planning framework (2003).

To zero-in on the *content* section of their framework, the following three categories can be considered:

- everyday ←→ scientific
- description ←→ explanation ←→generalization
- empirical ←→ theoretical

The over-arching notion here is that teachers need to realize the difference between everyday language vs. scientific language, and that the differences can be difficult to perceive from the teacher's perceptive because each student has a different linguistic background. And, even if most of the linguistic difficulties have been accounted for, there are still "internal" differences such as empirical vs. theoretical. For the EFL teacher, the dichotomy is again textbook English vs. real-word English -- or, English learning before a perceived need vs. English learning due to a viscerally perceived need.

Regarding *Communicative approach in Education* (not to be confused with TEFL's *Communicative approach*), there are four classes proposed. I have lined them up here from the most "conservative" to the most "liberal" forms of communication in the classroom:

- Non-interactive and authoritative
- Interactive and authoritative
- Non-interactive and dialogic
- Interactive and dialogic

Mortimer and Scott's work here is fully compatible with my teaching spectrum (above), and adds credibility to the model. Lined up in the order of the bullet points, "non-interactive and authoritative" is at the apex of my teaching spectrum (most conservative). And, "interactive and dialogic" is at the bottom of the spectrum (most liberal). Figure 22 is an updated version of the teaching spectrum inclusive of Mortimer and Scott's communicative approaches.

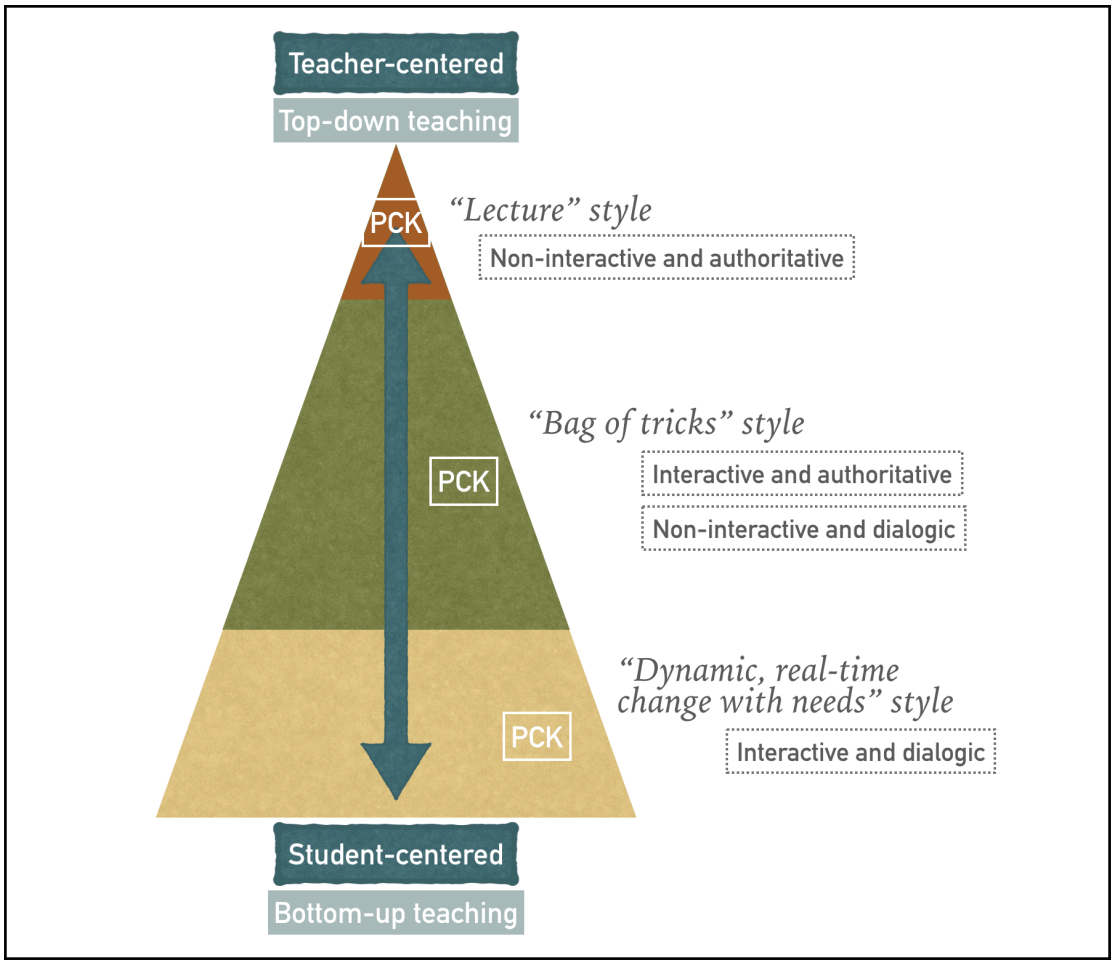


Figure 22: Updated version of teaching spectrum

Regarding *Patterns of discourse*, Mortimer and Scott discuss the I-R-E pattern and the I-R-F-R-F patterns. Below is the I-R-E pattern (Mehan, 1979), which is the variant just under that.

- I = initiation from the teacher (question)
- R = response from the student.
- E = Evaluation by teacher

- I = initiation from the teacher (question)
- R = response from the student
- F = feedback from teacher
- R = response from the student.
- F = feedback from teacher
- ...continues

The truncated and quick format would be I-R-E. The more free-flowing, and student-centered approach would likely have more of the involved variant come into play. This discussion is a natural segue into *Teaching interventions*.

Regarding *Teaching interventions*, Mortimer and Scott discuss six intervention types. The intervention types are: Shaping ideas, Selecting ideas, Marking key ideas, Sharing ideas, Checking student understanding, and Reviewing. *Patterns of discourse* and *Teaching interventions* naturally go hand-in-hand into the "Actions" section of Mortimer and Scott's analytical framework—from an *applied linguistics* point of view, they would both be considered to be *pragmatics*.

What are the implications regarding the literature review content?

Earlier I discussed how Schulman's worries regarding signature pedagogies may be excessive. This is because I felt that all teaching, regardless of the context, can be put someplace on that singular spectrum and because I believe that there are not a large number of archetypes to teaching. This means that only the content-specific teaching points need to be attached to the most appropriate teaching archetype, dynamically. It is a matter of the "how to" part, or the teacher's pedagogical maturity, that is required to make the connections to the appropriate locations on the spectrum.

What are the implications for bottom-up teaching?

Considering the popularity of lecture-styled teaching in Japanese EFL classrooms, and the oft seen adversity toward task-based teaching by non-native EFL teachers, it becomes apparent that content that a teacher is less familiar with seems to naturally pull such teachers to lecture-based teaching; as long as a teacher has the teaching content well-prepared, they can more-or-less relax while hitting the points in sequence—the need for ad-libbing is minimal, and so it is less worrisome. And, regarding bottom-up teaching—even well-maturing teachers may forget to implement this into their lesson plans, so it may be helpful for teachers to view the three zones in the teaching spectrum as a template for better lesson designing.

The takeaways for materials design

In this section, I focused on Schulman, Gardner, Rowland, and then Mortimer and Scott's work. These are the implications:

- comfort with I-R-F-R-F patterns may be the sign of growing maturity in the teacher.
- all teaching, regardless of the context, is on the teaching spectrum
- bottom-up teaching comes more naturally with familiar teaching content
- the three zones in the teaching spectrum are a good template for lesson designing
- there is a limited number of archetypical teaching approaches unto which content specifics are merely attached to

Although this may be a controversial conception, I do feel that I have made a valuable argument regarding the existence of a singular teaching spectrum (especially considering how it parallels with key elements of Mortimer and Scott). The gist of my own argument is that the mechanics of teaching naturally exist within the spectrum. *Signature pedagogies* are not complicated; what changes per teaching context is the context-specific content—while archetypes of the underlying pedagogy already exist on the spectrum. Put in this way, PCK as a construct is much less convoluted; it is a matter of sliding up and down within the spectrum for the different activities. I would like to continue with this line of thought, making attempts to better define PCK, the KQ, and my own teaching spectrum model.

3.6 Materials design in sum: A new protocol based on Dynamic Equity is required

To create truly useful and "cutting-edge" materials (that are motivational and learner-efficient), the above points should converge in a unifying design protocol. The protocol should be *process-based*, *analytical*, and *cyclical*. But it must not forget to include *focus on form*. It should also take into account *course objectives*, *learner objectives*, *learning context* (socio-cultural issues), and *learner creativity*. From the teacher's perspective, it should also consider that teachers do not always do what they say, context affects teacher cognition, teachers apply more bottom-up approaches when they grow comfortable with the content, and there are only a few archetypical pedagogies that exist on a single spectrum—and classroom teaching moves up and down this spectrum dynamically, so for optimal teaching practices, teachers should pay attention to where they are on the spectrum as they teach. The only way to successfully combine all of these points is to have a dynamically flexible protocol that is cyclical and is:

- semi-controlled by the teacher
- semi-controlled by the learner
- accepts regular updates based upon outcomes and contextual variables.

How can this *Dynamic Equity* be implemented? Chapter 4 is my solution to this complicated question.

4. Foundations of OL1

This chapter will examine the theoretical groundwork that went into the designs of the books in the *OL1* coursebooks. Not all of the books in the series were created equally— the most different being the very first incarnation—the original *Optimal Levels!* (OL1). To put those differences into perspective, for this chapter I will open with a preamble section that concentrates on the objectives that went into the original *Optimal Levels!* books. In that preamble, I will discuss: (a) the pedagogical design of the original *Optimal Levels!* coursebook, and (b) the findings of a 2-year study of the original by Murphy and Uemura (2013) to see if those original objectives were met.

In the next few sections of this chapter, I will provide specifics of *OL2's design*: (a) how the design was transformed, and (b) the theoretical justifications behind the transformations. Much of this chapter is a page-by-page walkthrough evaluation of a full chapter of *OL2*, focusing on theoretical elements and innovations, grounded in and drawing upon insights established earlier in this thesis. (Note: Content from Chapter 4 has been previously published via professional development workshops, and conference proceedings.)

4.1 Preamble: The original Optimal Levels! coursebook

4.1.1 The pedagogical design of OL1

The original *Optimal Levels!* coursebook (OL1) was designed as the centerpiece of my MA dissertation from the University of Birmingham (Murphy, 2009). My goal at the time was to pioneer a pedagogy that integrated the best practices of TESOL with the-then untapped research in *Mind, Brain, and Education* from the Harvard Graduate School of Education. From a historical perspective, the marriage of these two separate fields marked the birth of what is now referred to as *neuroELT*—and in retrospect, *Optimal Levels!* came to be the first coursebook series

to be explicitly designed with what concurrently emerged as neuroELT maxims.

Figure 23 is the flowchart from my 2009 dissertation depicting the systematic design that each chapter in *OL1* followed. Unlike mainstream English coursebooks, it does not follow many of the de facto standard coursebook obligations (such as seen in the PPP and Grammar-Translation coursebooks). Being a neuroELT-based design, *OL1* begins with student choices regarding the pre-set theme in the book. Without typical "top-down lecturing", it closely follows the implications of cognitive development as seen in Fischer's Dynamic Skill Theory (1980), which is based on Piagetian concepts of development. *OL1* therefore incorporates constructivist-based tasks that *build-up* the theme. An example flow is: transitioning from student chosen L2 lexis, to L2 sentence building, to L2 paragraph writing (regarding that designated theme that is being "constructed"). It then wraps up with what is called a *Performance of Understanding*—this is jargon used at the Harvard Graduate School of Education (see Murphy, 2009) for a presentation of some sort that cannot typically be rote memorized, but is designed to genuinely display (and make tangible) how deeply each student has learned the content and skills over the course of the unit of study.

Visually, the pedagogical design of the *OL* series is more akin to a workbook's design by clearly displaying how student learning is compounding each week, rather than a typical English coursebook that is full of dense reading passages and grammar notes (that students may end up glossing over). In short, it is what some people like to call an "active learning" design. See the Appendix (F) for a full example chapter of *OL1,* plus a full Teacher's Manual that details how to use each section of the coursebook.

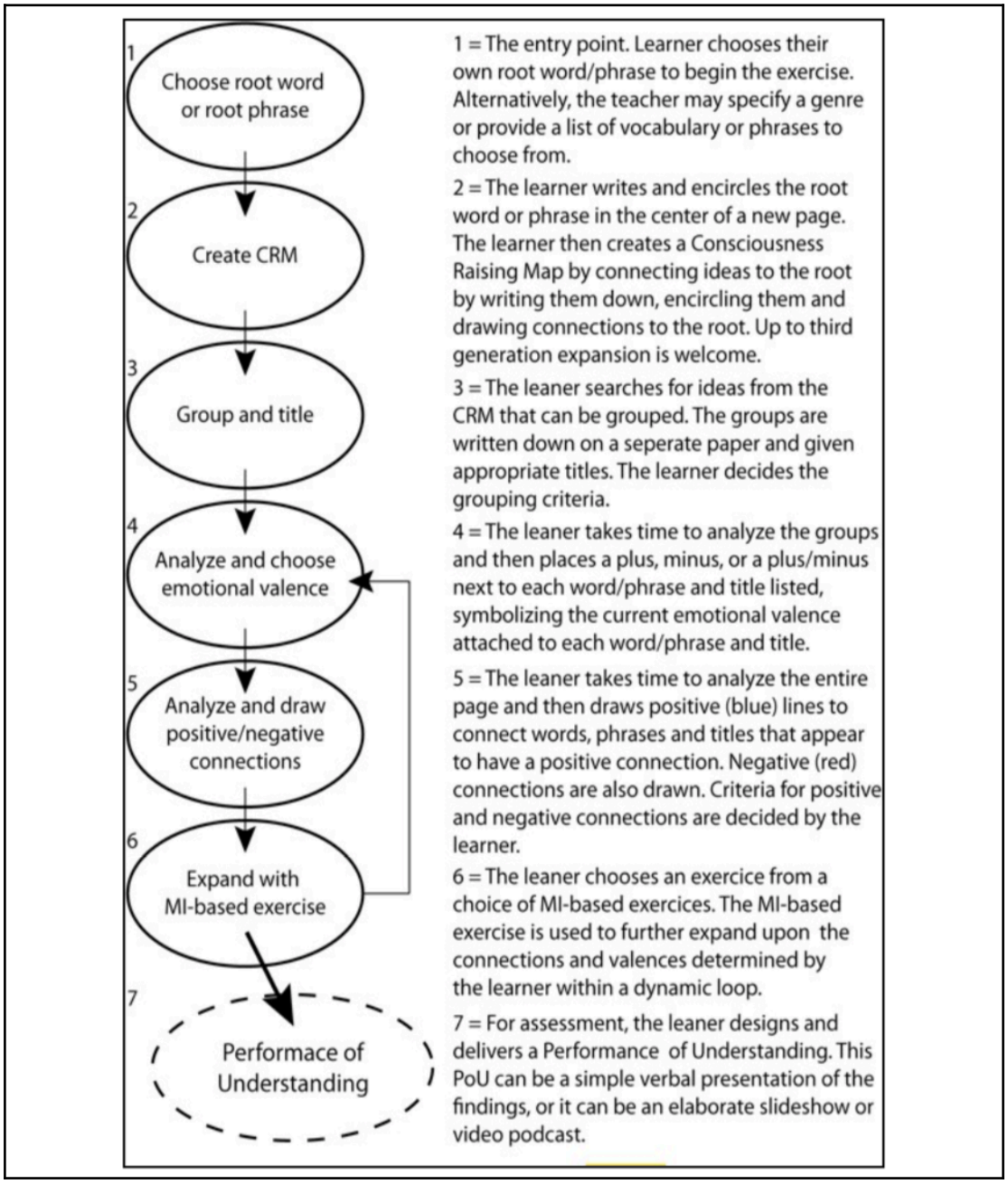


Figure 23: The original flowchart that Optimal Levels! (OL1) was based on Murphy (2009). [SIC content]

Figure 24 lists the 10 activities per unit. The specifics of each activity are listed in the right-hand column. It fits in well with the flow in Figure 23. And, (in retrospect) it follows a linear progression, which was later discovered to play a key role in *negatively* affecting outcomes—when compared with cyclical designs. This

comparison and general affect will be discussed in this chapter (and shall be compounded upon with more intricacies throughout this thesis).

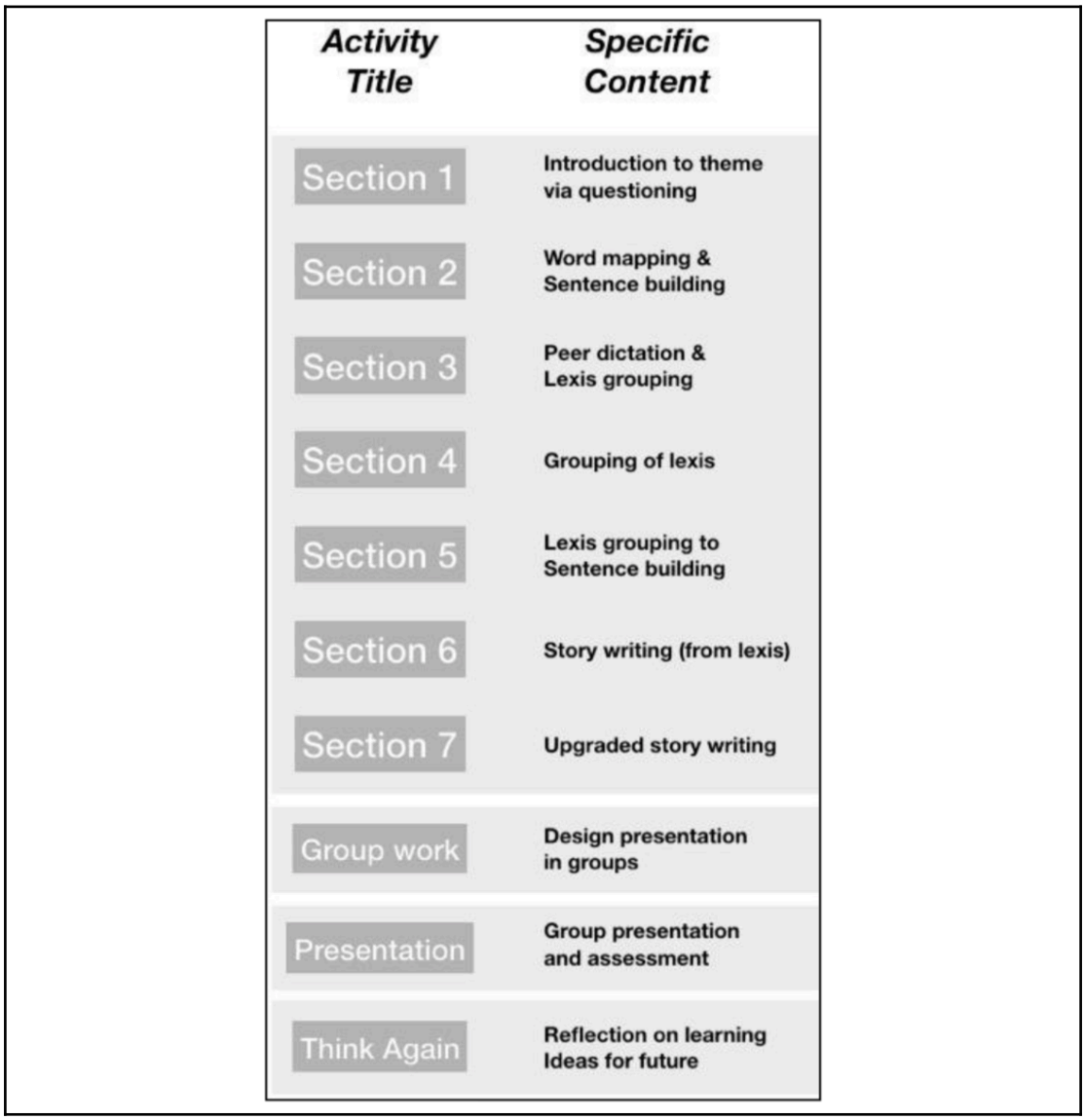


Figure 24: The flow of learning over the course of a single unit of study (Module) of OL1

How well did this original design fair? Did the learning compound within the students as intended? Also, what did the *students* think about this unique coursebook style? The next sub-section briefs the findings of a 2-year study of *OL1* (Murphy & Uemura, 2013), conducted at a Japanese university.

4.1.2 The findings of a 2-year study of *OL1*

A colleague and I conducted a longitudinal study on the efficacy of the original *OL1* coursebook (Murphy and Uemura, 2013). We collected data ranging from 2010 to 2012 (four semesters) from students using *OL1* at a university in southern Japan. A total of 1037 student responses (quantitative and qualitative responses) were collected and analyzed.

Over this 2-year research period, several conclusions surfaced. Firstly, we questioned whether the *OL1* pedagogy was genuinely process based, analytical, and cyclical; we concluded that these three points had been covered well enough by *OL1,* but with due reservations regarding their efficacy; deep discussions regarding this appear throughout the reminder of this thesis.

We found that the implementation of this pedagogical approach yielded positive results at the university level; student enjoyment factors were high—as well as their motivation levels. Students voiced appreciation for the unique-to-them peer-to-peer assessment that was implemented. This was a positive result, yet, at the time it was an unanticipated result for this Japanese university teaching context. From the data, we concluded that we should, therefore, continue to implement and perhaps even encourage other teachers to implement more peer-to-peer assessment for language students at Japanese universities. Students also made it clear that they desired to keep studying with the approach taken by *OL1*. To further investigate this quality of the design, regarding the dynamics of the taught courses, we questioned if the classes being taught with *OL1* were indeed:

(1) semi-controlled by the teacher
(2) semi-controlled by the learner
(3) accepting of regular updates based on outcomes and contextual variables.

We concluded positively with points (1) and (2) but realized that the teaching may have been lacking in point (3). *OL1* was not designed to seek for and integrate into itself regular updates based upon immediate student outcomes and contextual

variables. This means that the *self-attribution of knowledge* (compounding internal growth per student) was less effective than it could be. We therefore concluded that the future syllabus designs should be fine-tuned to allow for more controlled real-time changes based on prior student outcomes—this cyclical and self-amending design can now be deemed vital, but it was completely missing from the *OL1*.

The most striking results of that study came from our final quantitative question. It asked the participants to compare "traditional" Japanese teaching methods to this class's approach. Sixty-seven percent indicated that the methods used in this current class were perceived to be more effective than the traditional EFL methods that they were accustomed to (namely Grammar-Translation).

What areas of *OL1* were found to be in need of improvement? Most prominently, there was a call for more *top-down teaching* and for more *correction*. The students noticed a lack in *focus on form* and even requested this. This insight from the students provided significant gravitas to the idea of adding more *focus on form*. So, rather than just paying lip service to the idea of adding more grammar practice, a *focus on form* strategy was implemented centerstage in *OL2*.

4.2 *OL2* on the surface

The newer coursebook design (*OL2*) was to be an evolutionary step up from the original coursebook, *OL1*; see Appendix (G) for a full sample chapter of *OL2*. Figure 25 illustrates the flow of *OL2* compared to the original (also, see the Appendix for an example full chapter of *OL2*). There are several noteworthy changes; the most significant changes are marked in the figure. The star marks the timing of the "peak" per the unit of study. As can be seen in Figure 25, there is a long build-up toward the peak in the original book's design. There is then a rushed reflection right after the climax. This flow of events seems to be quite traditional for the language classroom. However, from a neuroELT perspective, such an ending is all-too-rushed while the build-up is much too drawn out—potentially killing a positive flow of motivation and learning. So, for *OL2* the peak is a timelier event.

Moreover, there are now two of these climaxes, intended to "double" the

excitement of the unit and keep motivational flow higher than it was for *OL1*. There is an active in-depth reflection period after the two climaxes—having students seriously look back at their work with the goal set at making two realistic new *plans of study* to raise their current skill level. The students are explicitly made aware of the goals of these reflective planning sessions—to ensure that the outcomes are more deeply insightful.

These changes were instrumental in making *OL2* successful as a dynamic and equity-based (catering to individual needs) cyclical syllabus. The shortening of the build-up period, and then the lengthening of the reflection period nurtures the deeper metalinguistic properties of the design, allowing time for the students to uncover what is necessary to fill "between the lines" on their own (thus, providing appropriate equity per student), solidifying the learning in a much more evocative and consequential way. This is specifically the type of learning that was presumed to be happening in, but was found to be missing from, the original book. Therefore, these meta-linguistic equity-based maturation periods were carefully included in *OL2* units as a part of a very controlled process.

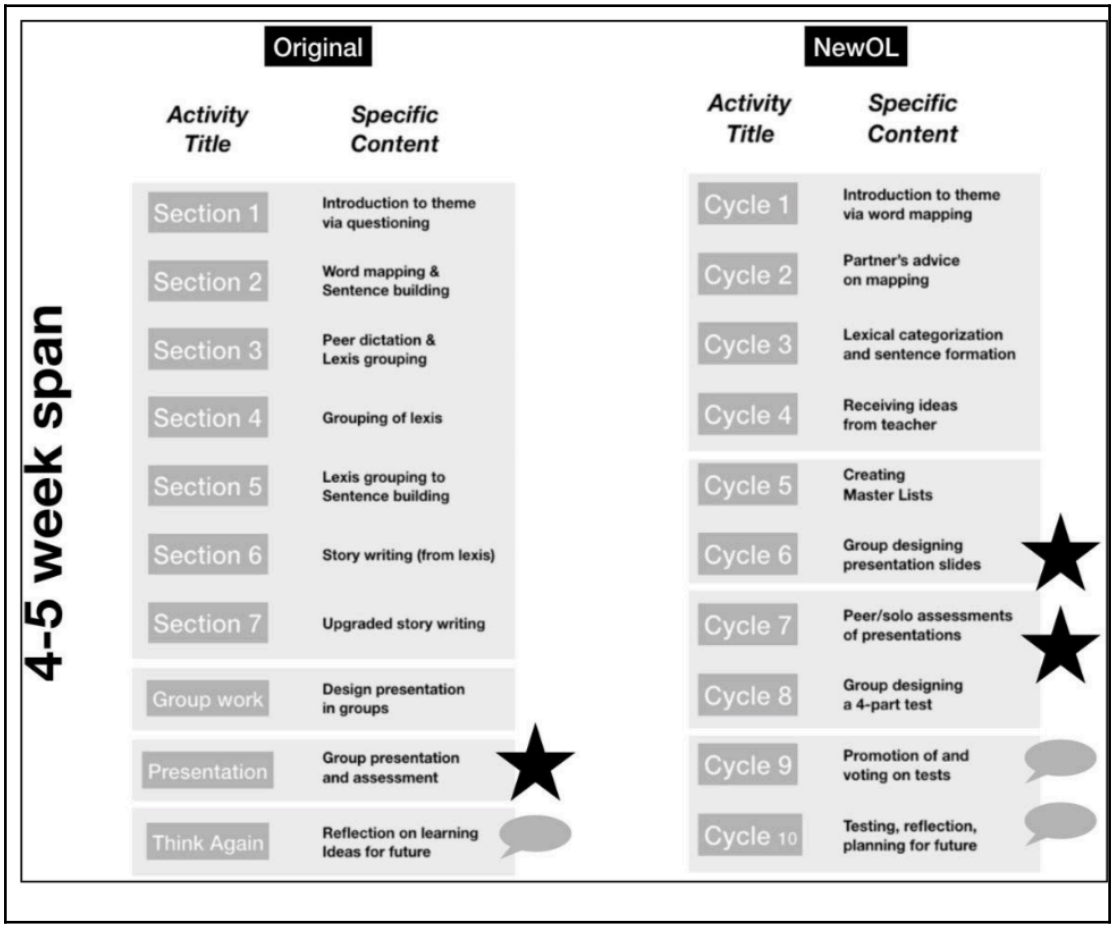


Figure 25: A comparison of the designs of Original (OL1) vs. NewOL (OL2). Stars indicate "climax" activities. Thought bubbles indicate metalinguistic maturation periods.

Much like financial investments, small but well-designed changes can compound over time into significant wealth. In practice, the original book's design had self-contained units. And, in retrospect, *OL1* did not incorporate ample metalinguistic maturation time, nor was it truly what can be regarded as *equity-based*—students were not provided with the opportunity to realize and later fill in their own skill and knowledge gaps. *OL1's* maturation time was therefore too rushed—and then the learning progression came to a full stop as the students moved on to the next unit of study. In short, student learning was not given the opportunity to compound into something significant. To make matters worse, the next unit of study hardly recycled any of the prior learning points. Frankly, much like many other coursebooks, from the perspective of the students, "all could be forgotten" after the

unit's assessment was over; in retrospect, there was clearly a lack of a compounding continuum within the pedagogical design of the coursebook. *OL2* was therefore designed to rectify these issues.

The original book required a "self-attribution of knowledge" –but it expected the students to do so without an explicitly designed continuum space to grow into. On the other hand, *OL2* has been designed to be cyclical and "orderly organic" –this provides the continuum space for students to organically grow into in an orderly manner. I cannot overstress the significance of this addition to the pedagogical design. This particular addition cultivates a metalinguistic maturation continuum for the students that is equity-based (not teacher-centered) —this particular design and procedure culminates to more significant transformations in student learning, real-world skill building, and content retention in the long term.

4.2.1 The newer protocol (COOPS)

The move from *OL1* to *OL2* required a radical re-thinking of pedagogical strategies. Learning from the best parts of *OL1*, it was decided that the fostering of "feelings of ownership" within the students was still imperative. I found the need to establish an elaborate but easy to follow protocol to ensure this incorporation would always take place—and not be something that is easily forgotten. So, to nurture and sustain that feeling of such ownership in the students, I devised the Cyclical Optimized Ownership Protocol (COOPs; Figure 26). The protocol mandates what must be provided to the students. Those mandated provisions are: thematic choices, procedural choices, and linguistic choices. These student-based choices are tethered to student-based responsibilities—and with this protocol the cycle of choice and accountability is set in motion. The cycle is meant to be sustained for the entirety of course while an explicit compounding of learning is also expected to occur, concurrently. Because COOPS played a fundamental part in the design and assessment of *OL2*—and in the larger discussion of general pedagogical theory—this protocol and its significance will be mulled over several times within this thesis.

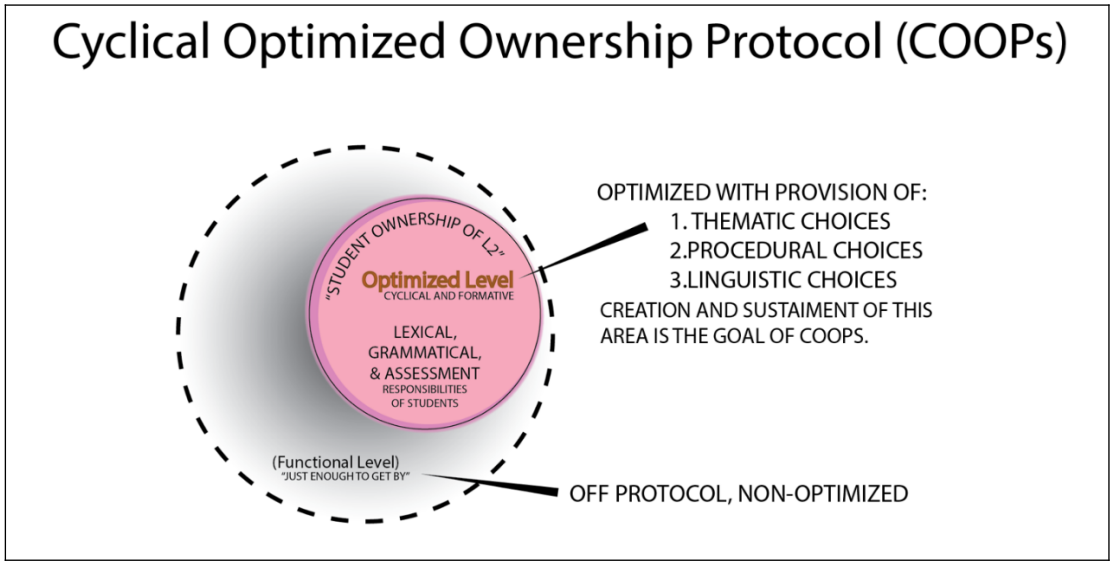


Figure 26: The foundation of OL2, the COOPs protocol.

4.3 The details of the *OL2* design

4.3.1 *OL2*'s Table of Contents

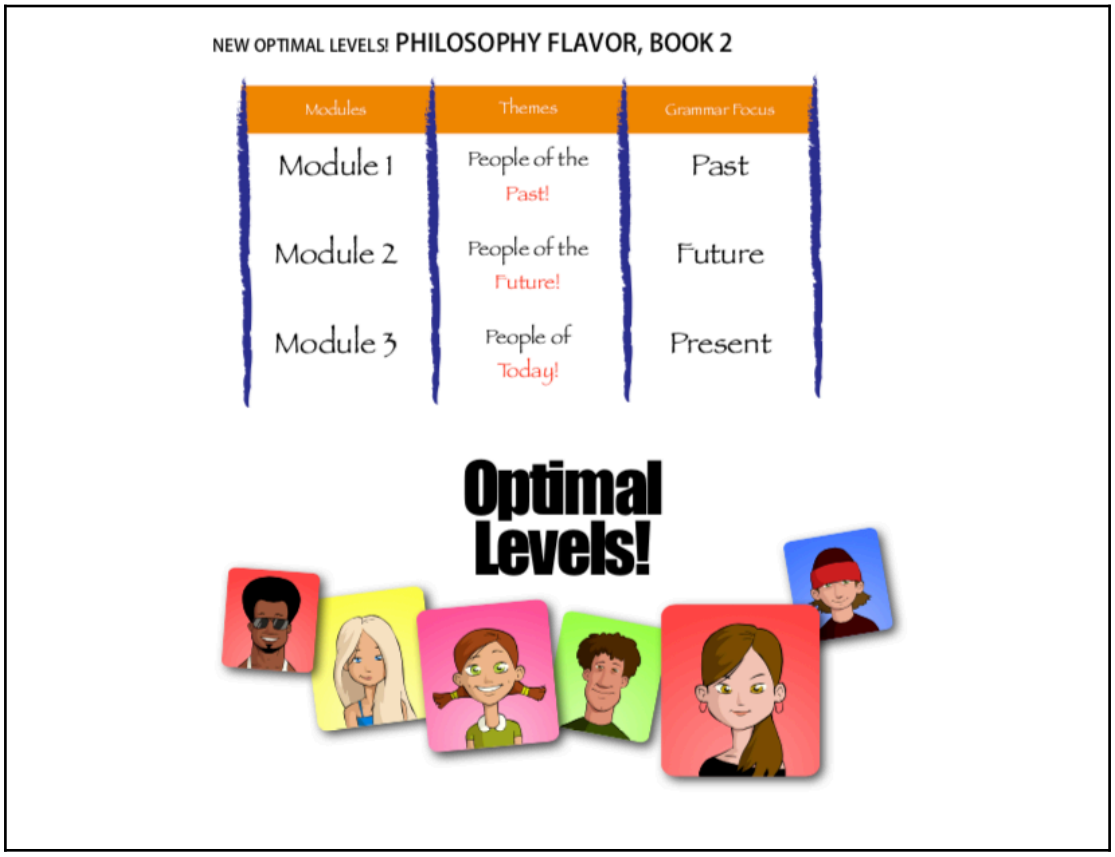
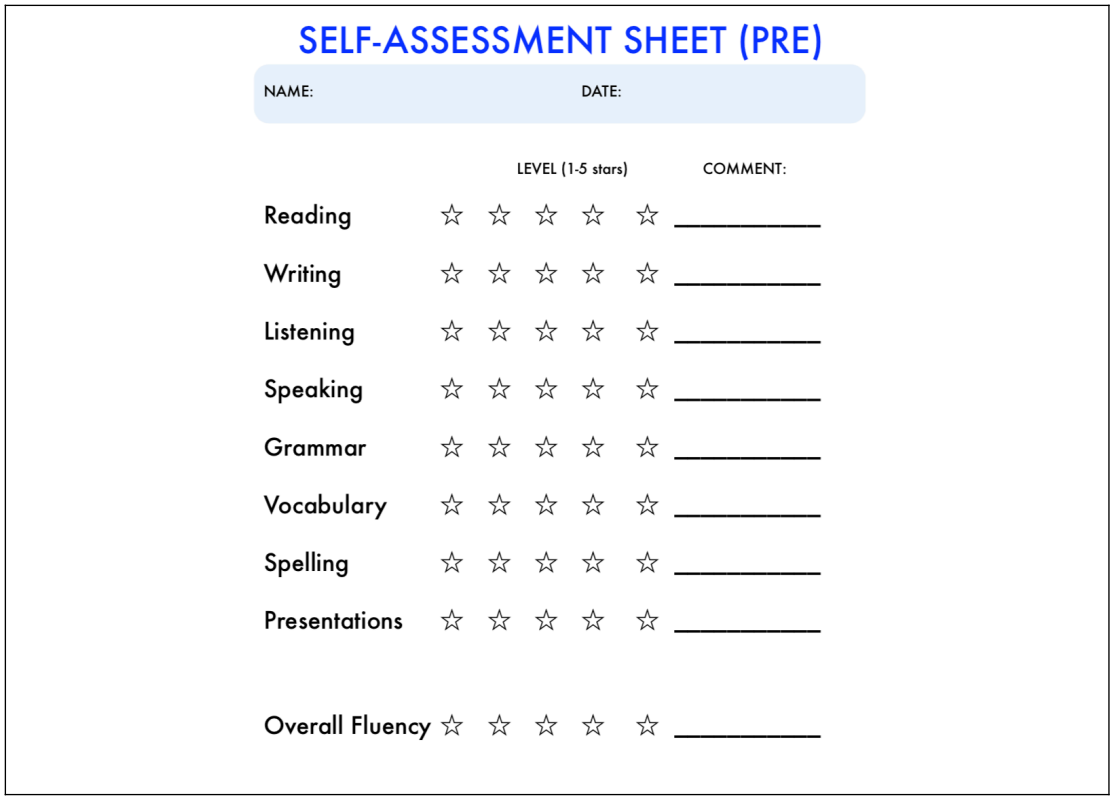


Figure 27: Copy of page 4, "Table of Contents"

The units of study (chapters) in *OL2* are referred to as *Modules* to help nuance the cyclical nature of the design. Each term, three modules covering three different themes (Figure 27) are offered to the students. It is the *students'* job to create content for and present on, those themes. There are a variety of *OL2* coursebooks (referred to as "flavors" within the series). For this page-by-page review, I have chosen *Philosophy Flavor, Book 2*.

The three themes presented in this particular book are: *People of the past, People of the future,* and *People of the today*. Because the themes are *past, future,* and *present*, respectively, this "cleverly" forces the students to use the past tense, the future tense, and the present tense repeatedly, while speaking, while writing, and while listening, for the duration of each Module. This gamut of varying grammar practice is accomplished by the students out of a stealthy necessity, so it goes on in the background—unnoticed by the students. In this way, the students are *focusing on form* in "stealth mode", which is a practical way to subdue potential negative baggage that some students will be carrying regarding explicit grammar studies.

4.3.2 “Self-assessment sheet (pre)”



SELF-ASSESSMENT SHEET (PRE)

NAME: DATE:

	LEVEL (1-5 stars)	COMMENT:
Reading	☆ ☆ ☆ ☆ ☆	_____
Writing	☆ ☆ ☆ ☆ ☆	_____
Listening	☆ ☆ ☆ ☆ ☆	_____
Speaking	☆ ☆ ☆ ☆ ☆	_____
Grammar	☆ ☆ ☆ ☆ ☆	_____
Vocabulary	☆ ☆ ☆ ☆ ☆	_____
Spelling	☆ ☆ ☆ ☆ ☆	_____
Presentations	☆ ☆ ☆ ☆ ☆	_____
Overall Fluency	☆ ☆ ☆ ☆ ☆	_____

Figure 28: “Self-assessment sheet” from page 5, pre-Module work

On the first day of class, the students are asked to evaluate themselves in eight categories (Figure 28). The evaluation criteria specifics are left up to the students—giving them a taste of autonomy from day one. The assessment scale is from 1-5, and they are asked to add a brief justification and/or comment next to each self-assessment. “Overall Fluency” is their averaged self-assessment. The students are asked to fill this section in after they are satisfied with the eight self-evaluations.

One major through-point is the honing of metacognitive skills. It starts with this page (Figure 28), which *precedes* the first module. Students are asked to self-evaluate eight skill areas + overall fluency and take short notes regarding their evaluations. They are told that they will come back to this page at the end of the year, so taking good notes now will be beneficial and fun for them (to compare with) at the end of the semester. In this way, we start off the semester by providing:

- Metacognition (via self-analysis and self-assessment)
- Personalization (this coursebook starts off with "me")
- Goal-setting (implicitly activated)

Note: an identical assessment sheet comes at the end of the coursebook; it is titled "SELF-ASSESSMENT SHEET (POST)." At the end of the semester, the students compare the results of these two evaluation sheets to reflect upon their personal growth and on their personal achievements from this semester of study.

Page Evaluation (see Table 2)

This is a repeating cycle, but the second cycle only appears at the end of the term. Perhaps completing a self-evaluation sheet after each Module would be more beneficial for the growth of students' metacognitive skills, and as a way for the students to more rapidly realize their own gaps of knowledge—as well as noticing the gaps of knowledge that their peers display. The evaluation is 5-star based (a typical format for students who are used to online shopping), with a sidebar notation area. *Ownership* of the coursebook is being built in this early stage in the coursebook by starting off with personal metacognitive analysis and self-evaluation. This is a deliberate lead-in to stronger feelings of ownership later in the course. As this is only a "lead-in," there are no specific *linguistic decisions, procedural decisions,* nor *thematic decisions* for the students to make this early on.

Table 2: Evaluation overview of page 5, "Self-assessment"

	Strong	**Weak**	**Absent**	**Notes**
Cyclical		X		
Boundaries	X			
Ownership	X			
Procedural Decisions			X	
Thematic Decisions			X	
Linguistic Decisions			X	
COOPS Coordination				Low coordination with COOPS, but this is a pre-module exercise.

4.3.3 “About me!”

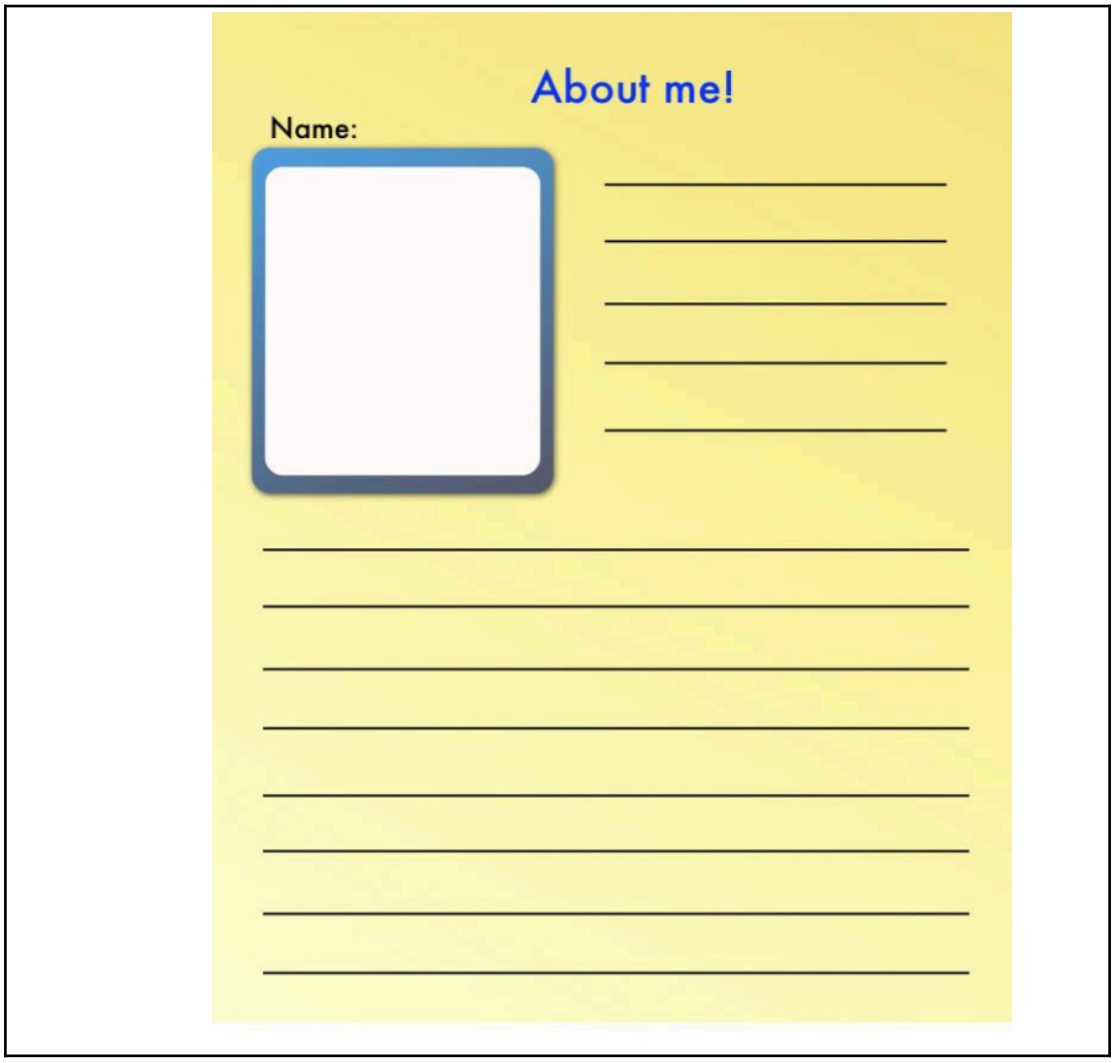


Figure 29: Copy of page 6, “About me!”, pre-Module work

This page is thematically linked to the previous page’s activity. Having already asked the students to self-assess themselves, here the students are asked to write in their new coursebook—to provide them a sense of autonomy, the content for this writing task is entirely up to the students. However, because of the way these two pages are set up, page 5's self-evaluation becomes a primer for this page’s task. Additionally, the students are asked to draw a picture of themselves, and/or paste a photo of themselves. This activity is meant to be completed in a cheerful style—it is definitely not meant to be endeavored with a serious or a confrontational tone.

This may look like a simple writing activity *“About Me!”* (Figure 29), with

an added drawing, but is important in the bigger picture for several reasons:

- Metacognition (carried over from the self-analysis page)
- Real personalization (this coursebook starts off with "me"—instead forcing students to care about generic "Jack and Betty" characters)
- Goal-setting (implicitly activated)
- Captivating/motivational
- Confidence building (explained below)

Design rationale: It is important to note that this singular activity alone would likely not be as epistemologically successful as the 2-part combination: the "Self-assessment sheet" + "About Me" writing task. Why is that? This pair works together thematically and pedagogically; the self-assessment warms them up with a non-threatening meta-cognitive activity filling in star-based evaluations of themselves. With their confidence unharmed, or perhaps even elevated, they move on to the more challenging writing activity. However, because the activity builds on what they were just thinking about (themselves), they should find it relatively easy—certainly more accessible than a cold start. In this way, this coursebook begins the students' journey with a keen focus on *them* and *their needs*, and not about unknown generic characters doing (boring) self-introductions.

This design style opens a pedagogic trend in this coursebook that always starts with singular (single point) ideas that build up into descriptive phrases, and then into more complete and better articulated sentences. Naturally, as we will see below, the ten cycles per Module in *OL2* have been designed to be a similar metalinguistic continuum (gradation of complexity) that initiates from singular ideas, then moves on to appropriate lexical items, then to sentences, and then finally to full presentations on the content that the students built up from those singular ideas – this all happens in an orderly, clearly visible, and systematic format per Module of study. The planned closure of each Module is an elaborate process (perhaps the most metacognitive of all the tasks). The design features of the Module closure will be

detailed later in this chapter.

Page Evaluation (see Table 3)

This is not a repeating cycle. It completes the lead-in from the page before. The boundaries are easy to understand. Write about yourself and draw a picture (or paste a photo). *Ownership* is strong as this is fully personal, and it completes the lead-in from the previous page. There are *linguistic decisions* and *thematic decisions*, but perhaps no significant *procedural decisions*.

Table 3: Evaluation overview of page 6, "About me!"

	Strong	**Weak**	**Absent**	**Notes**
Cyclical			X	
Boundaries	X			
Ownership	X			
Procedural Decisions			X	
Thematic Decisions		X		
Linguistic Decisions		X		
COOPS Coordination				Relatively weak. Not representative of COOPS as a standalone page.

4.3.4 The Module introduction

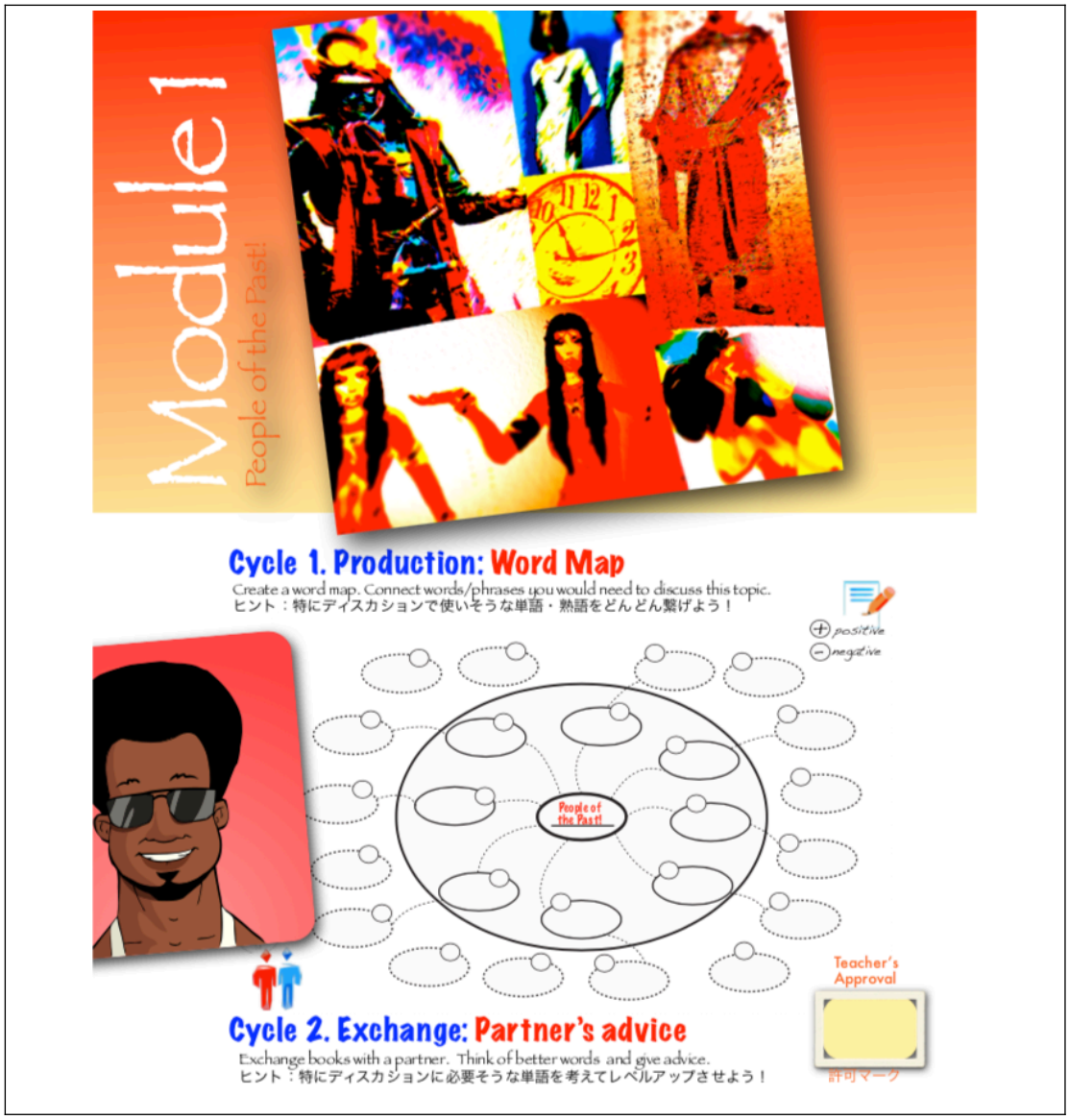


Figure 30: Cycle 1 and Cycle 2, Page 11

Every Module begins with an attention-grabbing graphic that sets the tone for the Module's topic. In this case (Figure 30), the theme is *People of the Past*, and so I included various historical images of people from around the world. The primary instructions are in English. Note, the Japanese instructions throughout the coursebook are deliberately not direct translations of the instructions, but they are often comments that "break the fourth wall," so to speak. They provide hints and commentary to the students in their native language, in the tone of a friendly

narrator. This is an atypical feature of this book that sets a pleasant mood and well-matches the friendly and colorful graphics.

"Cycle 1. Production: Word Map"

Rather than starting off the students' learning with a reading passage, or a quiz—which is the more orthodox starting points of a language coursebook chapter—in *OL2*, the students start their work independently with autonomous introspection and metacognitive action, trying to remember clues to the provided theme (in this case, memories *from their past*)—perhaps they will remember history lessons or documentaries that they had seen on TV, etc.

A critical feature of this cycle is the inclusion of the *emotional valence* bubbles (+ or -) that are attached to the top of each word bubble. This valence checking task is my twist on the standard word mapping design. This emotional valence task requires the students to decide if the lexical item that they had written in the bubble has a positive or a negative meaning *for them*, at the time of their writing. It is hypothesized that allowing the students to make these judgments over historical figures and concepts, so early on in their studies, further enhances a sense of autonomy and ownership toward the material.

The addition of this emotional valance choice task here plays a key role in establishing a positive foundation toward this material; it gives the students a sense of authority while being a small enough challenge as to not overwhelm them, nor likely have them lose confidence. It provides a sense of, "Yes, I can do this –and hmm… *I am* making the decisions here… I can enjoy this…"

Cycle 1 begins with an introductory picture that is designed to be colorful yet ambiguous; prominent yet not dominant. I chose this chronologically-based theme format because it will influence the students to continuously use the appropriate tense (in this case the past tense), allowing them to focus on building specific grammar skills in a non-tedious way–that is to say that they won't ever realize that they are explicitly "studying how to use the past tense". In this way, I was able to covertly mandate the grammar to be learned top-down, yet as will be made clear

below, the students only see the bottom-up aspects of the course because the ensuing pedagogical flow is so focused on their *personally perceived* needs.

This initiatory activity starts off at the conceptualization of basic ideas (singular words) that appear in their minds as they imagine how they might use language in their upcoming presentations. Students are explicitly told to be creative (break out of any cultural constraints they might have), as there is no "right or wrong" answer for these presentations. This often comes as a mild culture shock to some Japanese students because typically there is only ONE right answer in Japanese schools—rather famously, even an insignificant deviation from the *authorized-to-be-correct-answer* is often penalized in Japan. However, for this course, students soon enough learn to appreciate the new sense of freedom—as it gradually becomes the newly established norm for the culture of this particular teaching style.

"Cycle 2. Exchange: Partner's advice"

When the students fill out the majority of the bubbles (typically taking 5-10 minutes), the books are exchanged with a partner. This is a low-pressure task, as there are no right or wrong answers – it is a simple peer-to-peer advice-giving session. Ensuing conversations examples are: "What about connecting this person over here?", "I would have thought this person deserves a plus (+) mark. Why did you give him a minus (-)? Oh, I see…". This is likely a low-pressure task, but I propose four justifications for this task:

- The students will be using English to justify their decisions
- The students will be using English to provide advice to their partner (or ask questions)
- While negotiating meaning regarding the theme's lexical items, the students will likely be solidifying their knowledge, gaining confidence via the meaning negotiation process
- It provides a real opportunity for peer-bonding, which is essential for students' mental fitness—especially at the beginning of any course full of strangers, as it naturally subdues potential anxiety across the entire student body.

"Teacher's Approval"

In the lower right corner, there is a *Teacher's approval* box. When the students finish their pair work, they bring their filled in book pages to the teacher. If the written content is not "deep" enough, I ask the students to re-visit the topic and write more substantial content—then bring it back for approval. The students cannot move to the next page without this approval. So, while the students are provided with concrete autonomy, they are also provided with concrete boundaries—they learn that they must earn their evaluation with well thought out responses, and not just "fluff".

Page Evaluation (see Table 4)

This is a repeating cycle. Learners get better and faster at this mapping per cycle they attempt. The boundaries are clear; they must fill out the map, but how it is done is completely up to the learners. *Ownership* is very strong as this is fully personal—and the emotional valence bubbles make a significant difference on a psychological level, fusing the lexical items with the learners' deeper thoughts—while balancing their knee-jerk reactions to the creative ideas that pop into their heads. There are *linguistic decisions, procedural decisions, and thematic decisions.* However, being the first step in the module, the decisions and potential repercussions are deliberately kept light, to keep this a low-pressure task (compared to upcoming tasks).

Table 4: Evaluation overview of page 11," Cycles 1 and 2"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries	X			
Ownership	X			
Procedural Decisions	X			
Thematic Decisions	X			
Linguistic Decisions	X			
COOPS Coordination				Low coordination with COOPS, but this is a pre-module exercise

4.3.5 “Cycle 3a. Production: Word categories”



Figure 31: Copy of page 12, “Cycle 3a”

The students look at their word maps, then they are asked to do mental manipulation of their lexical items (Figure 31). They must divide the lexical items into one of four sub-genres of the theme:

- Cultural connections
- Feelings and emotions
- Connections to society
- Phrases and idioms

As the students mentally manipulate these lexical items, they gain new pathways of thinking regarding these lexical items. The justification processes in their heads must juggle ideas and make the sub-genre a logical connection for each of the lexical items. As they do this, they will likely also conjure up new lexical items. When that happens, I ask them to include those newer lexical items on this page and figure out where they should be included in their word map. (This is often the case with sub-genre task sections C. and D. as their original word map is unlikely to have lexical items from these perspectives from the onset.)

"Teacher's Approval"

When students feel ready (when most of the page is filled out sufficiently), students bring their book page to the teacher for approval. I review the page and engage each student with a random selection of questions regarding various points on their page. In this way, I attempt to gauge how well each student is currently grasping the theme—because my questions are random, the students cannot, and therefore do not attempt to, prepare for my questions with pre-set responses. I ask students to re-visit the content if I don't feel that the provided content is "deep" enough to move on to the next activity. When their content reaches an appropriate quality, I give them the approval mark and they move on to the next page.

Page Evaluation (see Table 5)

This is a repeating cycle. Learners get better and faster as they learn to think in terms of these four specific genres' perspectives, and how lexical items can fit logically into each of the four genres. The boundaries are clear; there are four specific genres that they must fit their lexical items into. *Ownership* may not be very strong the first time they do this cycle. However, by the second and third time they do this activity (in future Modules), students' feelings of ownership build up naturally as their brains become accustomed to this way of thinking.

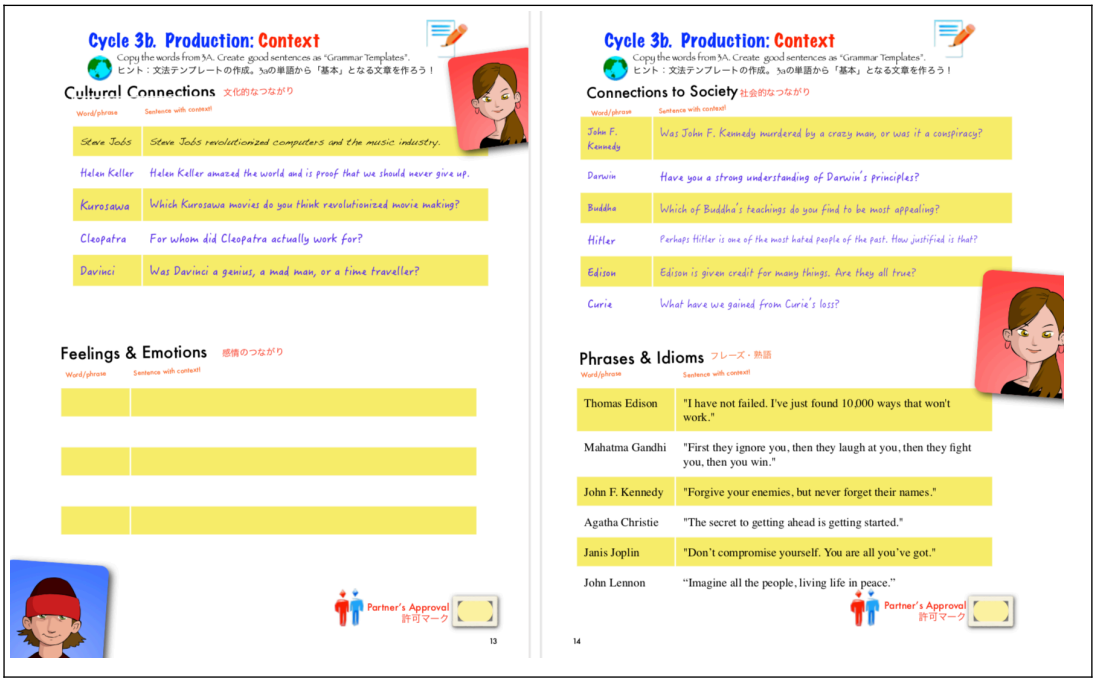
There are *linguistic decisions and thematic decisions* to be made by the learners, but probably not many significant *procedural decisions*, as the four genres

are already set.

Table 5: Evaluation overview of page 12, "Cycle 3a"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries	X			
Ownership	X			Gets stronger with progression
Procedural Decisions			X	
Thematic Decisions	X			
Linguistic Decisions	X			
COOPS Coordination				Works well enough but there is rooms for improvement

4.3.6 “Cycle 3b. Production: Context”



Cycle 3b. Production: Context

Copy the words from 3A. Create good sentences as "Grammar Templates".
ヒント：文法テンプレートの作成。3aの単語から「基本」となる文章を作ろう！

Cultural Connections 文化的なつながり

Word/phrase	Sentence with context!
Steve Jobs	Steve Jobs revolutionized computers and the music industry.
Helen Keller	Helen Keller amazed the world and is proof that we should never give up.
Kurosawa	Which Kurosawa movies do you think revolutionized movie making?
Cleopatra	For whom did Cleopatra actually work for?
Davinci	Was Davinci a genius, a mad man, or a time traveller?

Feelings & Emotions 感情のつながり

Word/phrase	Sentence with context!

Partner's Approval 許可マーク

13

Cycle 3b. Production: Context

Copy the words from 3A. Create good sentences as "Grammar Templates".
ヒント：文法テンプレートの作成。3aの単語から「基本」となる文章を作ろう！

Connections to Society 社会的なつながり

Word/phrase	Sentence with context!
John F. Kennedy	Was John F. Kennedy murdered by a crazy man, or was it a conspiracy?
Darwin	Have you a strong understanding of Darwin's principles?
Buddha	Which of Buddha's teachings do you find to be most appealing?
Hitler	Perhaps Hitler is one of the most hated people of the past. How justified is that?
Edison	Edison is given credit for many things. Are they all true?
Curie	What have we gained from Curie's loss?

Phrases & Idioms フレーズ・熟語

Word/phrase	Sentence with context!
Thomas Edison	"I have not failed. I've just found 10,000 ways that won't work."
Mahatma Gandhi	"First they ignore you, then they laugh at you, then they fight you, then you win."
John F. Kennedy	"Forgive your enemies, but never forget their names."
Agatha Christie	"The secret to getting ahead is getting started."
Janis Joplin	"Don't compromise yourself. You are all you've got."
John Lennon	"Imagine all the people, living life in peace."

Partner's Approval 許可マーク

14

Figure 32: Copy of pages 13-14, “Cycle 3b”

When the students have an adequate list of lexical items on the previous page, and they have been able to justify their reasonings to the teacher, they are now ready to write their ideas down in sentence form (Figure 32). Cycle 3b spans two pages and requires each student to write six original sentences per sub-genre from the previous page (the tasks: A, B, C, and D). I remind the students that while they structure these sentences, they might as well design them to be phrases and sentences that they will be able to use in their upcoming presentation on this topic. In this way, I remind them of their goal, keep them aligned with their purpose, and comfortably challenge them. This keeps them aligned and slightly on their toes. (Conversely, at times when I happen to forget to provide this advice to the students, a large portion of students “fall back” on old habits, presumably from prior language courses, and simply write up sentences that pragmatically fit the task, yet clearly have no big-picture thematic connection with their upcoming presentation. In short, they are wasting the opportunity to prime their brains for the larger over-arching connections that they will soon need to be making. From this I deduce that long-term teaching arcs of compounding complexity and content are not common pedagogical traits in

these students' past lessons. Indeed, my own memories of foreign language studies are mostly of disjointed sentence building, designed to re-produce the immediate teaching point for assessment purposes, but not as part of a grander over-arching thematically compounding climax. Keen attention to details regarding the cognitive development of students within teaching arcs such as these set the *OL* series apart from traditional coursebook designs that are more focused on immediate results.)

"Partner's Approval"

Until now these corner boxes were *"Teacher's Approval"* boxes, but here it is a peer that provides the evaluation and approval. This is likely more for the benefit of the peer than the author of the sentences. The chosen partner reads each sentence and determines if it is good enough for inclusion in the upcoming presentation. It provides the partner with a sense of autonomy and perhaps more importantly, a sense of responsibility toward the other student, and responsibility to the flow and content of the course. In this way, this is likely a natural rapport building activity.

Page Evaluation (see Table 6)

This is a repeating cycle. Learners get better and faster as they learn to think and write sentences along the lines of these four genres, and how lexical items can fit logically into them. The boundaries are clear; there are four genres of writing that the students must fit their lexical items into. *Ownership* starts getting stronger from this cycle because they are extending their thoughts (lexical items) into full sentence structures. They are witnessing their ideas come to life in a more meaningful way. There are *linguistic decisions, thematic decisions, and procedural decisions* being made here.

Table 6: Evaluation overview of pages 13-14, "Cycle 3b"

	Strong	**Weak**	**Absent**	**Notes**
Cyclical	X			
Boundaries	X			
Ownership	X			Strong ownership begins with this cycle
Procedural Decisions			X	
Thematic Decisions			X	
Linguistic Decisions			X	
COOPS Coordination				Very good example of COOPS. Integrates and compounds learning with other cycles

4.3.7 “Cycle 4b. Exchange: Grammar Ideas from the teacher”

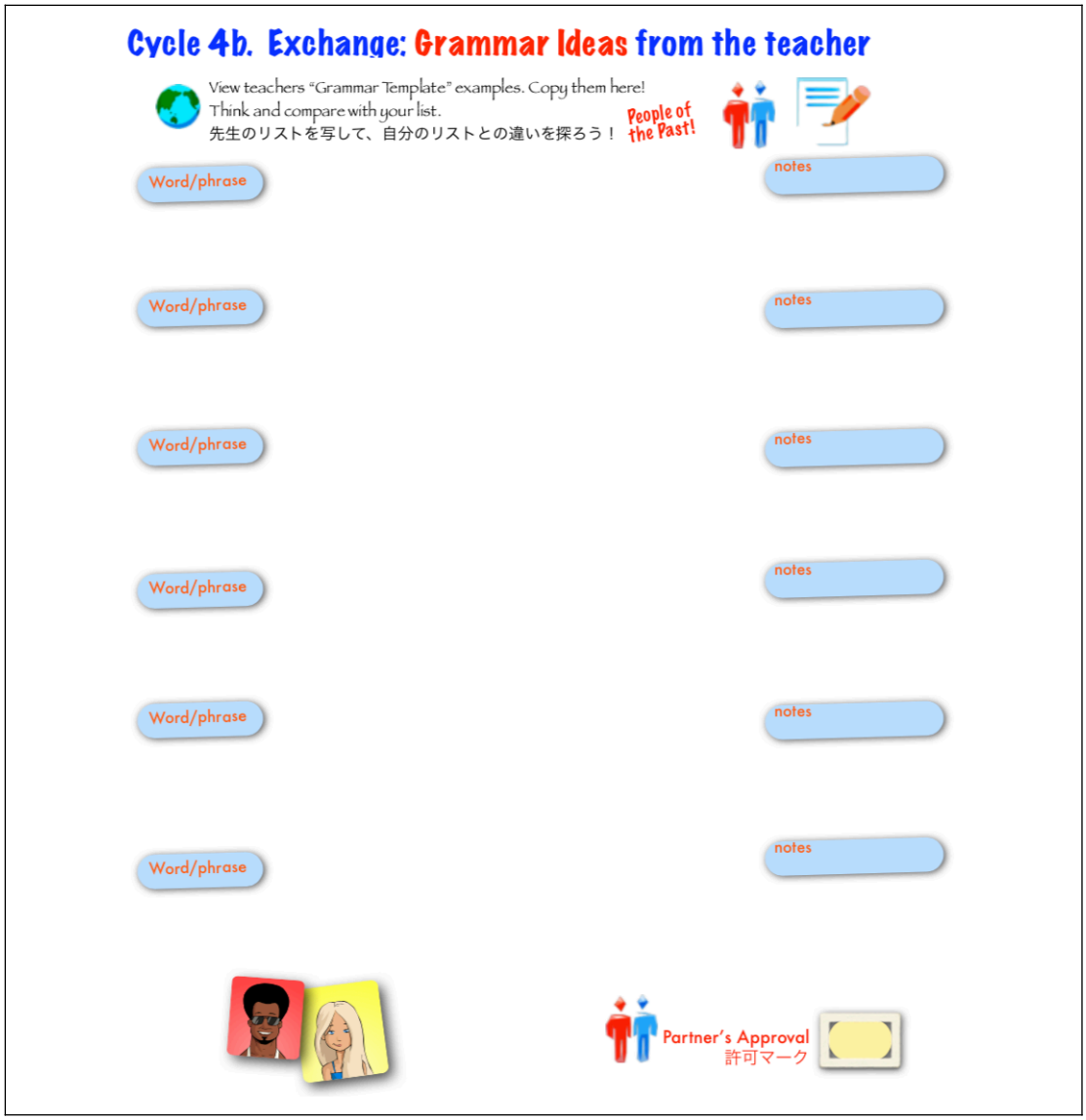


Figure 33: Copy of page 16, “Cycle 4b”

By this time of this cycle in the Module (Figure 33), the students have likely built some confidence regarding the sentences they might be able to use in their upcoming presentation. However, at the same time, they are probably not entirely sure of themselves either. Capitalizing on this inherent dichotomy of developing confidence and lingering insecurity, here the teacher can offer advice costumed as casual banter → *“Well, if I were to do a presentation on this topic I would say, ‘XXX and YYY’”* and run down a list of ideas seemingly off the top of his/her head. Over the years, this approach has been visibly successful in grabbing students’ attention

while veiling its true pedagogical intent—so it does not seem at all pedantic to the students.

For this "set up" to be most effective with the students, the teacher must have a pre-planned set of sentence structures and lexical items to be introduced to the students that well-match the current theme. The students, however, are unaware of this pre-planning and they are simply asked to take notes on what they think might be useful for them. They take in the advice as if it was a set of personalized "Christmas presents" (in stark contrast to a boring lecture on a grammar point). However, this is truly a Focus on Form lesson that is tactfully camouflaged.

How is this different from a grammar lecture? Theoretically, why is this approach more appealing for the students (and therefore likely to be less stressful for the teacher and the students)? Because of the order of this pedagogical "set up", the needs for mastery of those specific grammar points are viscerally felt by the students *before* the teacher starts talking about them. This order of events makes a critical difference in the minds of the students.

This process—*in this particular order*—raises: (a) mindsets of desire, (b) motivation to learn more, (c) autonomy to carry out what is necessary to gain mastery, and that leads to, (d) ownership and personal responsibility for the learned content. This virtous chain of events can occur because the decisions made by the students regarding what is to be written down for future use comes from the students while they are viscerally sensing a real and personal craving for this input—while also building a desire for the specific skills required to scratch those itches. All of these matters compound in an advantageous way that matches the student-perceived over-arching goals of the Module.

The intention of this page's design, therefore, is to make sure the students are aptly affected by their own naturally occurring personal feelings of insecurity toward the upcoming presentation—and then to use that extra energy in a positive way by enticing them to take responsibility on a more enthusiastic level than may have been typical of them until this point in their language learning careers.

Page Evaluation (see Table 7)

This is a repeating cycle. Learners begin to expect to hear good advice from the teacher by the second and third time they come to this point in their learning cycles. There are no *rigid* boundaries—that would likely be counter-productive for this task's intent. But, although the task may be unconventional, the objective is clear, and therefore so are the "soft boundaries"; the objective is to gather hints from the seemingly ad-libbed teacher's talk, and then take notes on this page regarding those camouflaged learning points. *Ownership*: When I discuss this activity in teacher training workshops, I tend to say this activity "makes me feel like Santa Claus on Christmas Day". I convey it this way to the teachers in the workshops because for this activity, the language learners treat this activity as if they are receiving little gifts from the teacher. Not all of the "gifts" are applicable to each of the students, so they get to choose which of those "gifts" they want to keep. In this way, the students are able to foster feelings of ownership due to the autonomous "choice" aspect that is built into this activity. Moreover, there is a natural desire to acquire helpful ideas at this point for the sake of their presentation group members as well, so a real sense of responsibility is also nurtured (but not forced upon the students). There is consistently a plethora of *linguistic decisions, thematic decisions, and procedural decisions* being made during the implementation of this task.

Table 7: Evaluation overview of page 16, "Cycle 4b"

	Strong	Weak	Absent	Notes
Cyclical	X			Learners begin to expect to hear great content next cycle
Boundaries		X		
Ownership	X			Very strong ownership.
Procedural Decisions	X			strong
Thematic Decisions	X			strong
Linguistic Decisions				strong
COOPS Coordination				Very strong. It is the first "top-down" activity, but the students are needy for input by this time, so they happily engage with this task.

4.3.8 “Cycle 5a. Production – Group Master List!”

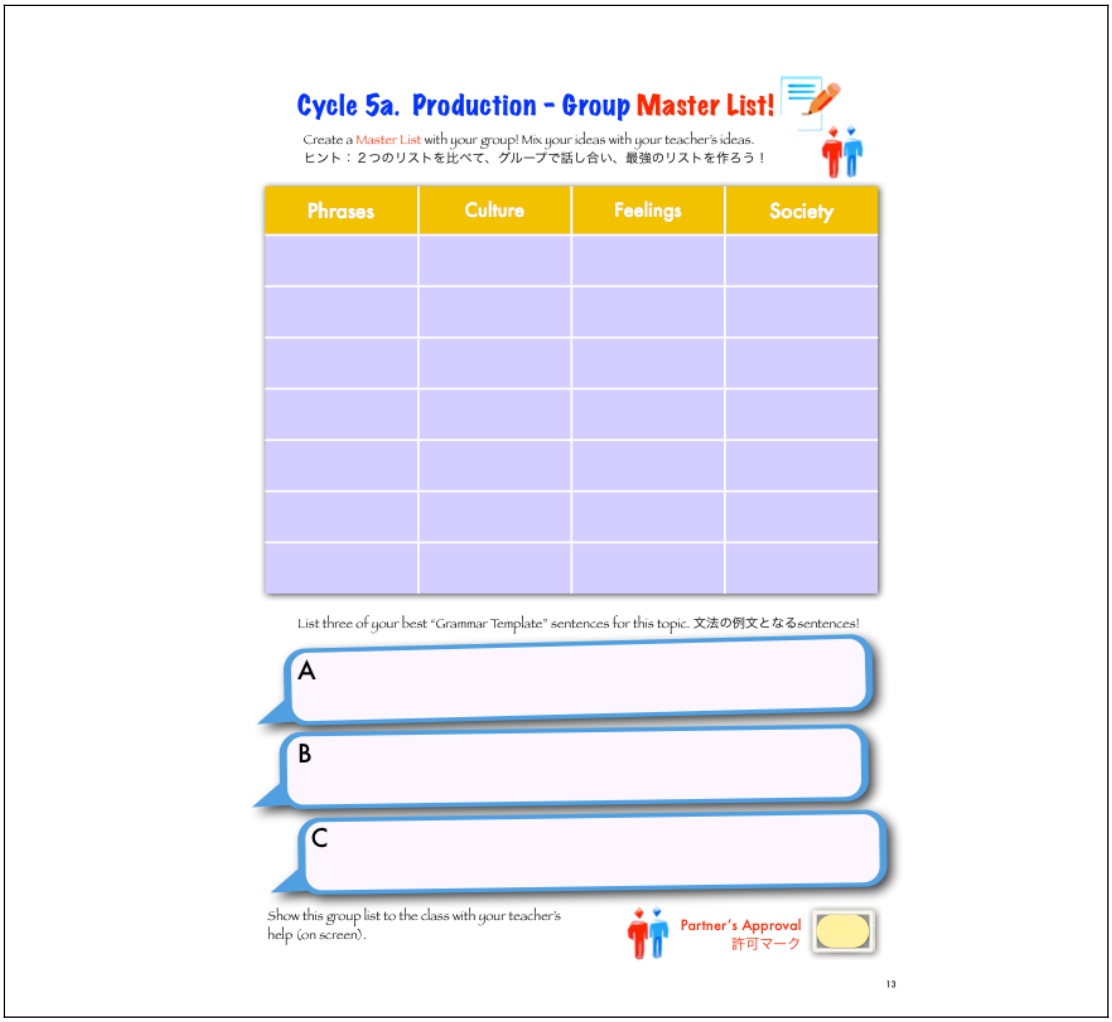


Figure 34: Copy of page 13, “Cycle 5a”

Students are either self-selected into, or are randomly put into, presentation groups of 4-6 students. In their presentation groups, they brainstorm ideas, mixing their ideas with peers' ideas and the teacher's ideas from the previous task. The goal is to fill in as much of this chart as possible (Figure 34). This would then become what we refer to as their “Master List”—a compiled list that they heavily rely on and often refer back to for locating and remembering all their lexical item choices from previous work -- for the sake of their upcoming group presentation. The master list therefore serves as a sort of bridge that links previous work to future group work. Also, three of what we refer to as “Grammar Template” sentence structures are

decided upon at this stage—this adds a concrete and explicit level of *focus on form* to the students' language studies. It is there to fulfill standard academic needs of a language classroom as well as to fulfill the psychological needs of Japanese students whom, out of habit, still strongly feel the need to memorize grammatical structures in their language studies. (As will be discussed later in this thesis, some students clearly requested explicit grammar instruction because such rote learning made this aspect of the classroom teaching feel more like "proper study" in their minds—and this provision gave them a sense of security. It was deemed important to cater to such fundamental psychological needs of these students, too—and not just concentrate efforts on progressive pedagogical designs.)

Page Evaluation (see Table 8)

This is a repeating cycle. Unsurprisingly, the language students typically only begin to appreciate the existence of this Master List from its usage in their second and third Module; during the first Module, it seems that some learners only find this to be a tedious activity because they have yet to experience the benefit of having this list when it comes time to design their group presentation.

The boundaries of this task are rigid. The Master List has only four categories—they directly represent the four genres that the student had just written about at the sentence level. The Master List is the centerpiece of the COOPS protocol in action—the students and teacher have co-created these boundaries, conforming to the design of this Master List, from which the upcoming group presentations will come to life.

Ownership: The student are in full control of what goes into their Master List. But, feelings of ownership really begin to show when they arrive at this activity in the second and third Module; enthusiasm toward this Master List creation is visibly higher during the second and third Module.

There is an abundance of linguistic decisions, thematic decisions, and procedural decisions being made during this task.

Table 8: Evaluation overview of page 13, "Cycle 5a"

	Strong	Weak	Notes
Cyclical	X		
Boundaries	X		
Ownership	X		Ownership grows with time.
Procedural Decisions	X		
Thematic Decisions	X		
Linguistic Decisions	X		
COOPS Coordination			Very high. It is the centerpiece of COOPS in this book. Naturally leads to in-depth thinking and further learning potential

4.3.9 “Cycle 5b. Class Grammar Templates List”

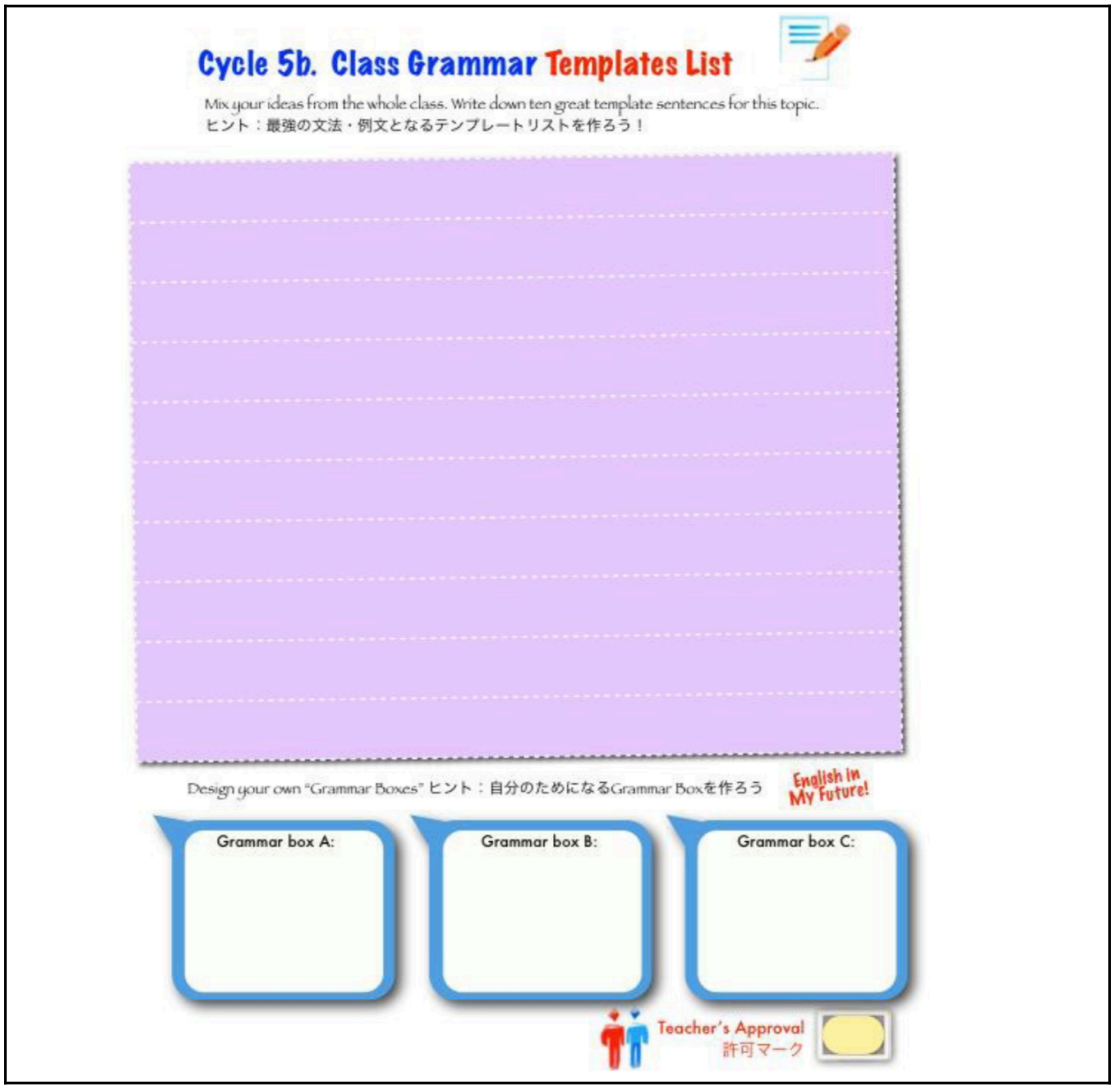


Figure 35: Copy of page 14, “Cycle 5b”

After the lexical items are settled on in the Master List on the previous page, the group of students move on to create a template list of sentence structures (grammar templates) to be used in their upcoming group presentation (Figure 35). There is a standard Teacher’s Approval box at the bottom of the page.

Page Evaluation (see Table 9)

This is a repeating cycle. But, not being a popular page, it gets used progressively *less* with each Module. The boundaries are set, but are flexible.

Ownership: In contrast with the Master List usage, by the time of the third Module, many students go so far as to skip over this Cycle entirely—these are diligent students, but by the third Module they have learned to see this as an unnecessary and mostly redundant page. By this time in the syllabus, many of the students (perhaps rightfully so) seem to just want to get on with the design work. There are many potential *linguistic decisions, thematic decisions, and procedural decisions* to be made. However, as discussed above, later in the year, this entire cycle is often skipped over by the students. (This issue, and the remedy, will be discussed later in the dissertation.)

Table 9: Evaluation overview of page 14, "Cycle 5b"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries		X		
Ownership			X	Ownership dissipates with time.
Procedural Decisions	X			
Thematic Decisions	X			
Linguistic Decisions	X			
COOPS Coordination				High, but seen as redundant by students. Not a popular page.

4.3.10 “Cycle 6. Group-owned: Presentation!”

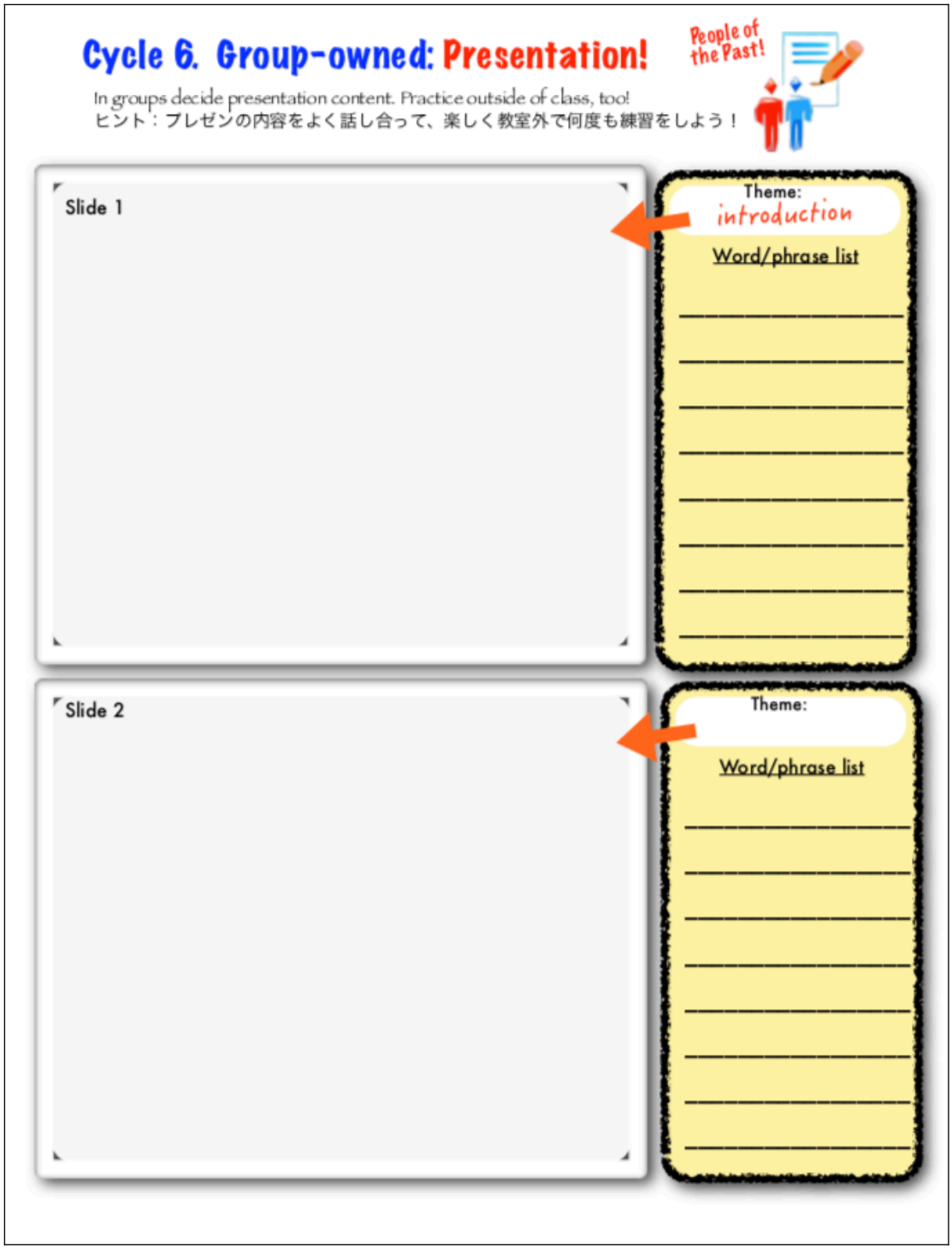


Figure 36: Copy of page 19, “Cycle 6”

The group presentations are limited to six slides per group. Each of those six slides is given a representative title (referred to as each individual slide’s “theme”).

There are two of these slide designing templates per page—along with the yellow sidebar used to house the lexical items to be referred to by the students as they present each slide (Figure 36). Students initiate their presentation designing process in their groups during class time, then they are expected to complete their designing and presentation practice outside of class—with the expectation that each group will find the time to complete one or two practice sessions on their own before the day of their presentation. In practice, the first Module is always "hit and miss" regarding the diligence of the students and their practice time—as can be clearly seen by the presentation outcomes of their first go at it. But by the time of the second and third Module, the substantial rise in quality of the group presentations not only indicates language learning and but it also seems to indicate more diligent practice sessions conducted by the students likely out of a sense of duty toward their groupmates, and the entire class—as they come to realize what a waste of time boring/tedious presentations can be. There seems to be a point early on when students come to the realization that if they are going to do these presentations, they might as well do a good job and make it entertaining for the sake of everyone involved.

I openly encourage this transformation in attitude by amenably praising the more content-heavy and/or entertaining and well-planned out presentations—the students quickly latch on to this even though they know that doing so does not directly affect their official course evaluation. Data regarding these points and noteworthy implications—especially in comparison with *OL2*—can be found in Section 8.9.

Page Evaluation (see Table 10)

This is a repeating cycle. The students learn to appreciate sidebar usage (their list of lexical items for each slide's theme) as they do their first presentation. Unsurprisingly, the students do not use the sidebar very well during their first Module's presentation. However, their sidebar usage improves in the second and third Module. The boundaries here are rigid and clear. These six slides are what they will be presenting with—they are not allowed to do more, or less. The sidebar must

be used to list the lexical items that they will be using per slide. *Ownership*: Because each student takes part in the artistic designs and the *logical flow* of the presentation, there are strong feelings of ownership, perhaps even on a deeper level than until now. There is a plethora of *linguistic decisions, thematic decisions, and procedural decisions* being made during this task.

Table 10: Evaluation overview of page 19, "Cycle 6"

	Strong	**Weak**	**Absent**	**Notes**
Cyclical	X			This is the most cyclical of all tasks.
Boundaries	X			
Ownership	X			
Procedural Decisions	X			plethora
Thematic Decisions	X			plethora
Linguistic Decisions	X			plethora
COOPS Coordination				Along with Cycle 5, this epitomizes COOPS

4.3.11 “Cycle 7. Presentation Feedback”

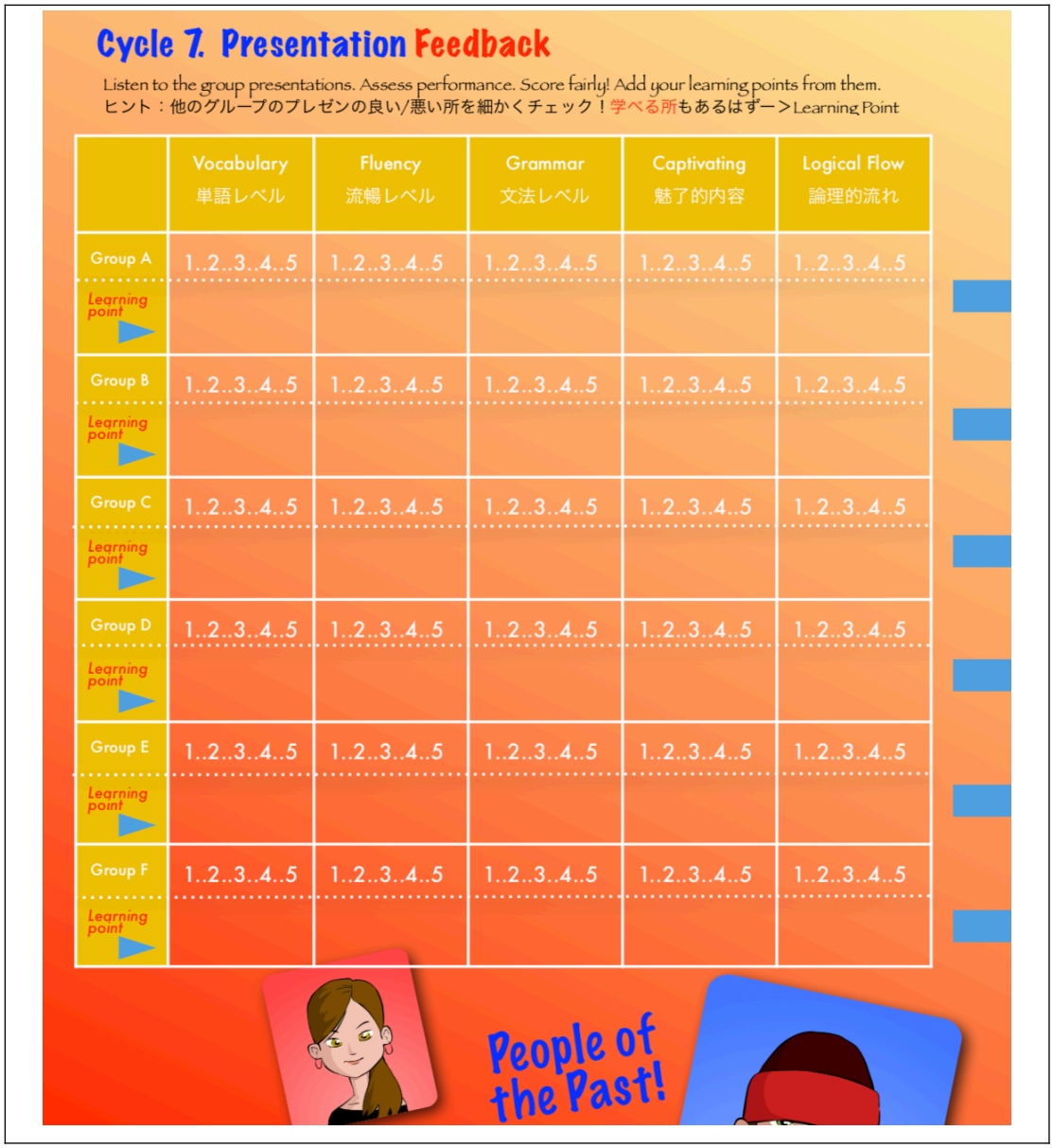


Figure 37: Copy of page 22, “Cycle 7”

Most presentation classes that I have witnessed seem to bore the students who are not presenting; the non-presenters seem to drift off into a daydream at some point during the hour. To win the war with this die-hard issue, this page goes head-on and *requires* the sitting students to actively assess their peers’ presentation content and presentation skills (Figure 37). The points assessed by the sitting students are *Vocabulary, Fluency, Grammar, Captivating,* and *Logical Flow*. (The presenting

students will also be doing *self-assessment* with the same criteria, when they return to their seats.) While the first three assessment points are the more typical points of assessment in Japanese EFL classes, the final two assessment points – *Captivating,* and *Logical Flow*—are probably not commonplace in Japanese school peer assessment, if they have peer assessment at all.

I hypothesize (after trailing this detailed method of peer assessment over several years) that the addition of these two final criteria points into the line-up make a significant difference in the outcomes, over time. Why? Because, as the learners assess their peers (and themselves) on aspects such as *logical flow*, they are forced to look beyond the grammatical points and lexical items; they need to examine how logically the content and the language usage interweaves to convey a coherent story—*convincingly*.

Regardless of how deeply each students' understanding evolves through the course, the fact of the matter is that on a philosophical level and on a psychosocial level, the *raison d'être* of human language is called to attention. Because of this inherent psychosocial phenomenon attached to the language learning, when ensuing Module 2 and Module 3's presentation designing cycles come around for the students, the students will often be enticed to be even more logical with their presentation designs—inclusive of, but not limited to, improved lexis, improved illustrations, and more logically justified language usage.

Because each Module's activities' designs are unwavering and cyclical—and therefore predictably compounding for the students—we can expect that with each progressive Module, the embedded cycles will be able to nurture further cognitive growth in regards to how the learners construe and evaluate their own demonstrations of their learning by means of improved *logical flow*, improved fluency, and perhaps even an added layer of confidence.

Page Evaluation (see Table 11)

This is a repeating cycle. Learners learn to appreciate the purpose of this assessment activity the more times they complete them. Learners begin to appreciate that they are gaining a "sharper eye" for these assessments, so there is also a positive metacognitive element emerging because of this cycle. The boundaries are clear. However, the specifics on how to assess on the 1-5 scale is left up to the students. The teacher only provides advice regarding what the middle score of "3" should be → "…for the rubric to be the most useful, a *3* should be what you deem to be an average/normal performance." Other than that, the students devise their own personal rubric system for the sake of added autonomy and feelings of ownership. Ownership: Peer assessment and self-assessment were completely new (unheard of) exercises for many of the students in the study. These assessment tasks were set up in the coursebook to resemble "diary-like" entries, making the task feel personal. To further this element of personalization, I asked the students to keep their marks hidden from the groups that they are assessing. In this way, the students maintain a sense of privacy that naturally brings on even more intimacy with this page. The right-side page in this 2-page spread gets filled with jotted-down private notation regarding the performances—especially with good presentation ideas and techniques that they might want to "steal" for their own upcoming presentations. In this way, there is a strong sense of ownership and a growing sense of responsibility regarding the task on this 2-page spread. Because the learners are given ample freedom for this task, linguistic decisions, thematic decisions, and procedural decisions may be made during the task, but there is no guarantee nor expectation that all three will be covered. However, the likelihood of ample coverage of all three points will rise as the Module numbers increase—as the learners' understanding and appreciation of the task deepens, and they take better notes.

Table 11: Evaluation overview of page 22, "Cycle 7"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries	X			
Ownership	X			
Procedural Decisions	X			
Thematic Decisions	X			
Linguistic Decisions	X			
COOPS Coordination				Very good. Particularly strong feelings of ownership

4.3.12 “Cycle 7b. Presentation Feedback notes”

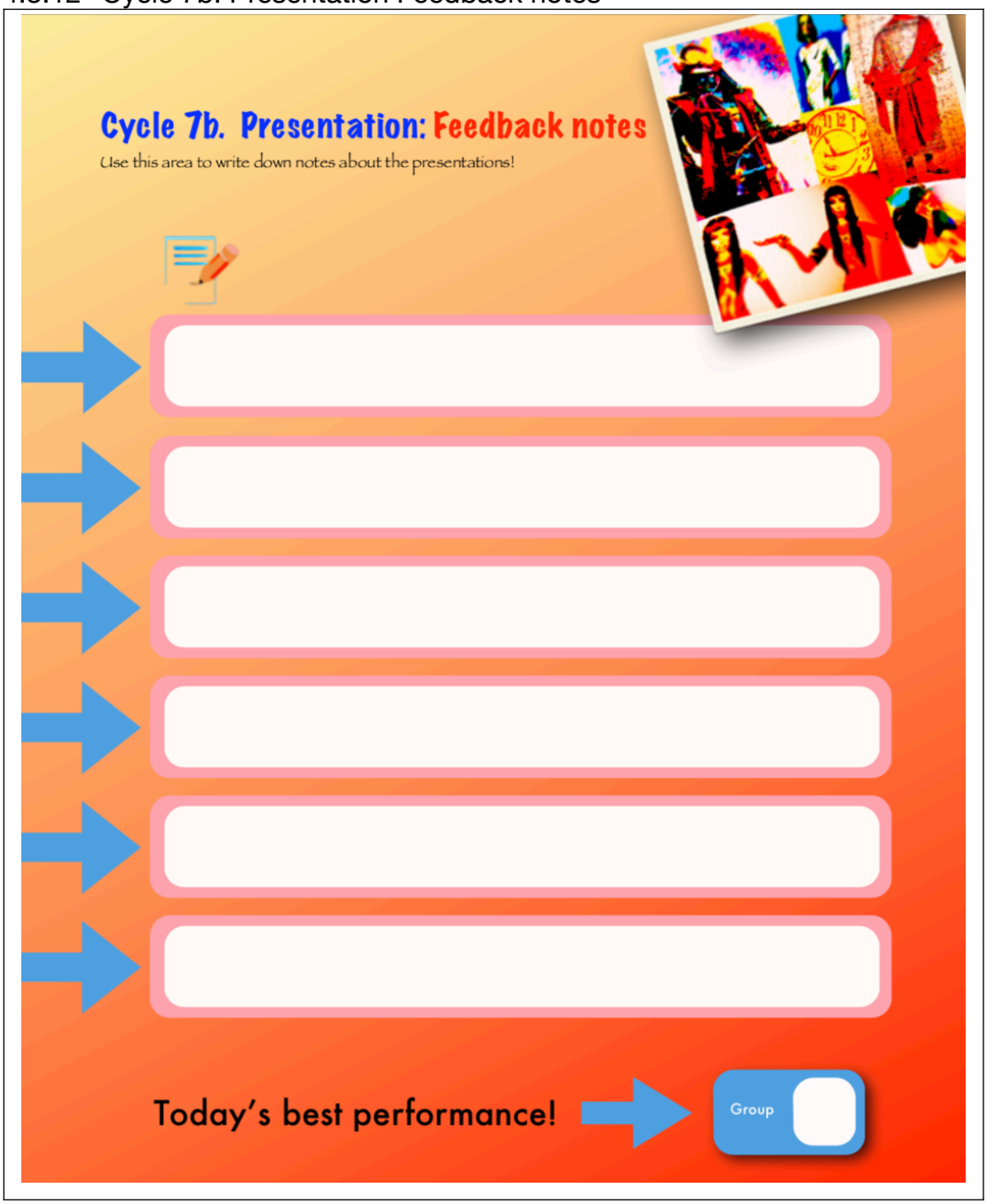


Figure 38: Copy of page 23, “Cycle 7b”

Cycle 7b is the right-side page of a 2-page spread, to be used concurrently with 7a. On this page (Figure 38), the students are asked to actively take notes on what they saw, with a particular focus on what they might want to “steal” for the sake of improving their upcoming presentations. At the bottom of the page is the

"Today's best performance!" box. Students tally up the scores from Cycle 7 and use their notes to decide which group they found to have the overall best performance that day. The group that was voted the best is given proper recognition for their efforts – they are asked to stand up to receive genuine applause, and to take a bow. Over time, it was evident that this assessment and simple award ceremony became a dominant force of motivation for the students as they openly cheered, and often wrote about how many times they were chosen for "best performance" in their journals—and even wrote about it in their year-end course evaluations.

Page Evaluation (see Table 12)

This is a repeating cycle. Learners quickly learn to take better notes. More importantly, this is where groups get voted for "best performance". Although this tally does not officially affect the semester marks, this is *clearly* an important task for the students—as a motivation and morale booster. The boundaries here are not rigid. The students are free to take notes in any way they find it to be beneficial for themselves. Some students prefer to take notes in Japanese because it is quicker for them to write in Japanese. However, the *game element* of this task is rigid, as only one group can be chosen as the best performance group.

Ownership: There is an interesting balance between free note-taking, 1-5 scoring (from 7a), and the "best performance" outcome. This is an exciting task for the students, and feelings of ownership are unambiguous; this task shares similarities with popular sporting events. Because the learners are given much freedom for this task*, linguistic decisions, thematic decisions, and procedural decisions* may be made during the task, but there is no guarantee that all three will be covered. However, the likelihood of ample coverage of all three points will rise as the Module numbers increase—as the learners' understanding of the task deepens, and they take better notes.

Table 12: Evaluation overview of page 23, “Cycle 7b”

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries		X		
Ownership	X			Very strong
Procedural Decisions		X		Likely (but not guaranteed)
Thematic Decisions		X		Likely (but not guaranteed)
Linguistic Decisions		X		Likely (but not guaranteed)
COOPS Coordination				As a part of the "climax" of the syllabus, it embodies COOPS well.

4.3.13 “Cycle 8. Group Work: Design a test!”

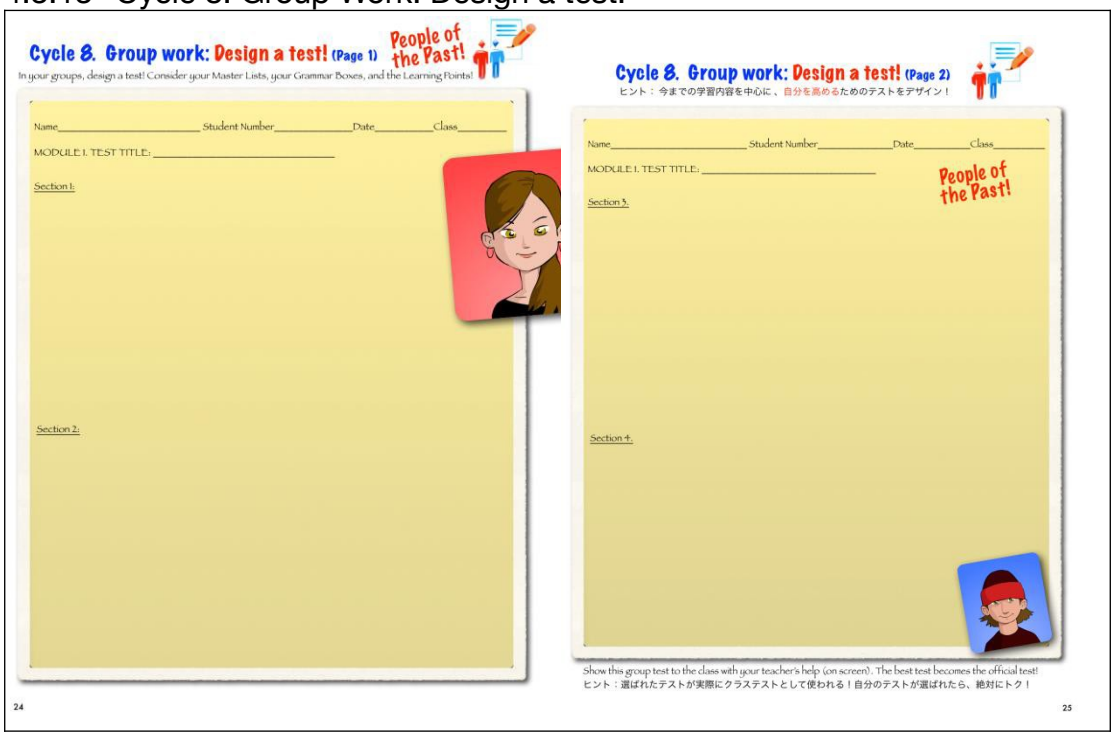


Figure 39: Copy of pages 24-25, “Cycle 8”

After the presentations are fully evaluated, and the students have had a chance to calm down from the excitement, the presentation groups get back together one more time to group-negotiate a four-part test, based on *their own* presentation content—inclusive of vocabulary and grammar points from their own group’s Master List, which likely has content received from the teacher. The groups are given full autonomy over their test designs. (Figure 39)

This opportunity to design a genuine test (albeit, a test worth only 10 points), is strongly representative of COOPS syllabus design protocol. The students are fully autonomous on a vital task, within the two binding rules. For this task, I typically nudge them with advice saying that “at least one section *should* be focused on lexical items,” and “…one section *should* be a long answer/paragraph-level question. But, it is really up to your group.” Students typically follow my two points of advice and are then free to focus creative thoughts with the other two sections – examples from past student test designs are: crossword puzzles, information gaps, listening-based

questions, and drawing questions.

Page Evaluation (see Table 13)

This is a repeating cycle. Much like teachers-in-training, the students learn to make better tests as the Modules and their compounding experiences progress (Module 1 → 2 → 3). The boundaries are easy to see. They are asked to design a 10-point test – so the groups must decide how many points each of the four sections should be assigned, and how many points should be assigned to each question within each of those sections. This group work naturally entails creativity and negotiation skills—and likely calls for some appreciation for teachers' work, which is something they may not have thought about until this point during their language studies.

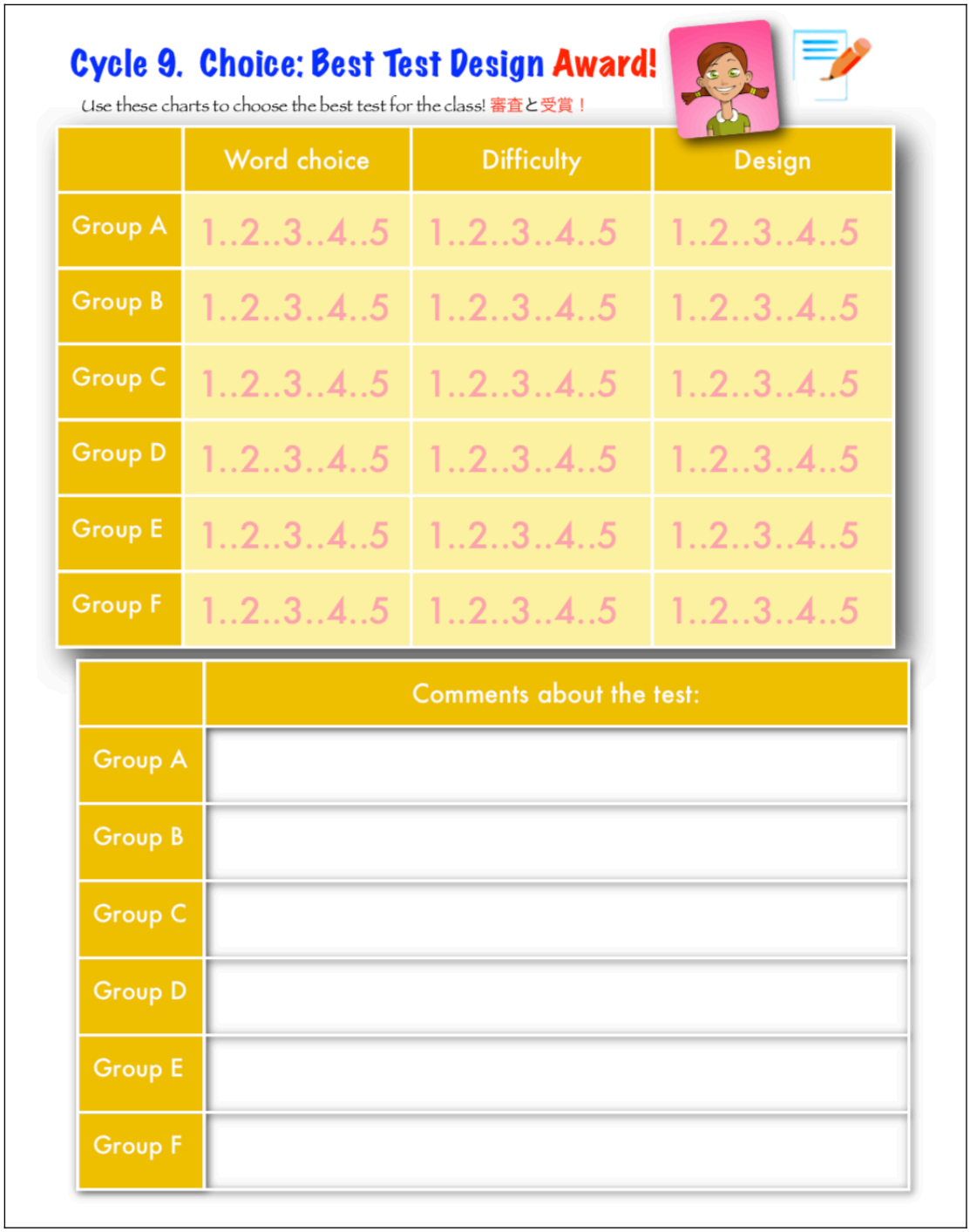
Ownership: Perhaps due to the afterglow effects of their presentations and best performance assessments, the students have a built-up affinity and bond to their presentation content—and now they are asked to design a test on that content. Therefore, autonomy and feelings of ownership are unsurprisingly high at this point—perhaps those feelings are even peaking at this point in the module.

This task is perhaps the best example of COOPS-level *linguistic decision, thematic decision, and procedural decision* making. All three elements are integral to this task.

Table 13: Evaluation overview of pages 23-25, "Cycle 8"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries	X			
Ownership	X			Strong affinity
Procedural Decisions	X			Very high
Thematic Decisions	X			Very high
Linguistic Decisions	X			Very high
COOPS Coordination				As a part of the "climax", this embodies COOPS very well.

4.3.14 “Cycle 9. Choice: Best Test Design Award!”



Cycle 9. Choice: Best Test Design Award!

Use these charts to choose the best test for the class! 審査と受賞！

	Word choice	Difficulty	Design
Group A	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group B	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group C	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group D	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group E	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group F	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5

	Comments about the test:
Group A	
Group B	
Group C	
Group D	
Group E	
Group F	

Figure 40: Copy of page 26, “Cycle 9”

As each group comes up to the front of the class and does a “live commercial” (a short promotional presentation), their peers are sitting at their desks with this rubric for assessment (Figure 40). They assess for *Word choice, Difficulty,*

and Design. And, they also notate supplementary information to help them make their final decisions. These notes are also pragmatic: (1) they help students realize new ideas and/or motivate them to improve their *next* test design, and (2) and, when the raw scores alone create a tie between groups' evaluations, the notes are used as the tie-breaker—the students are asked to review their notes and make their tie-breaking judgment based on their notes.

Page Evaluation (see Table 14)

This is a repeating cycle – what they learn from the test assessment, and the subsequent design promotion process (their group "commercial"), leads to future improvement in both test design, and in promotional skills—this is therefore likely to be a confidence booster for at least some of the students. The boundaries are set in the rubric, but there is freedom of interpretation. *Ownership* is built via the autonomy stemming from peer-to-peer evaluation and self-evaluation. *Linguistic decisions, procedural decisions,* and *thematic decisions*, may be made, but the likelihood of this coverage is context-dependent. Therefore, it should not be assumed that all three areas will be automatically covered.

Table 14: Evaluation overview of page 26, "Cycle 9"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries		X		
Ownership	X			
Procedural Decisions		X		
Thematic Decisions		X		
Linguistic Decisions		X		
COOPS Coordination				Much less restrictive, so good for autonomy. But, not a good example of COOPS. Waning productivity in motion.

4.3.15 “Cycle 10. Testing, Reflection, and New Plans!”

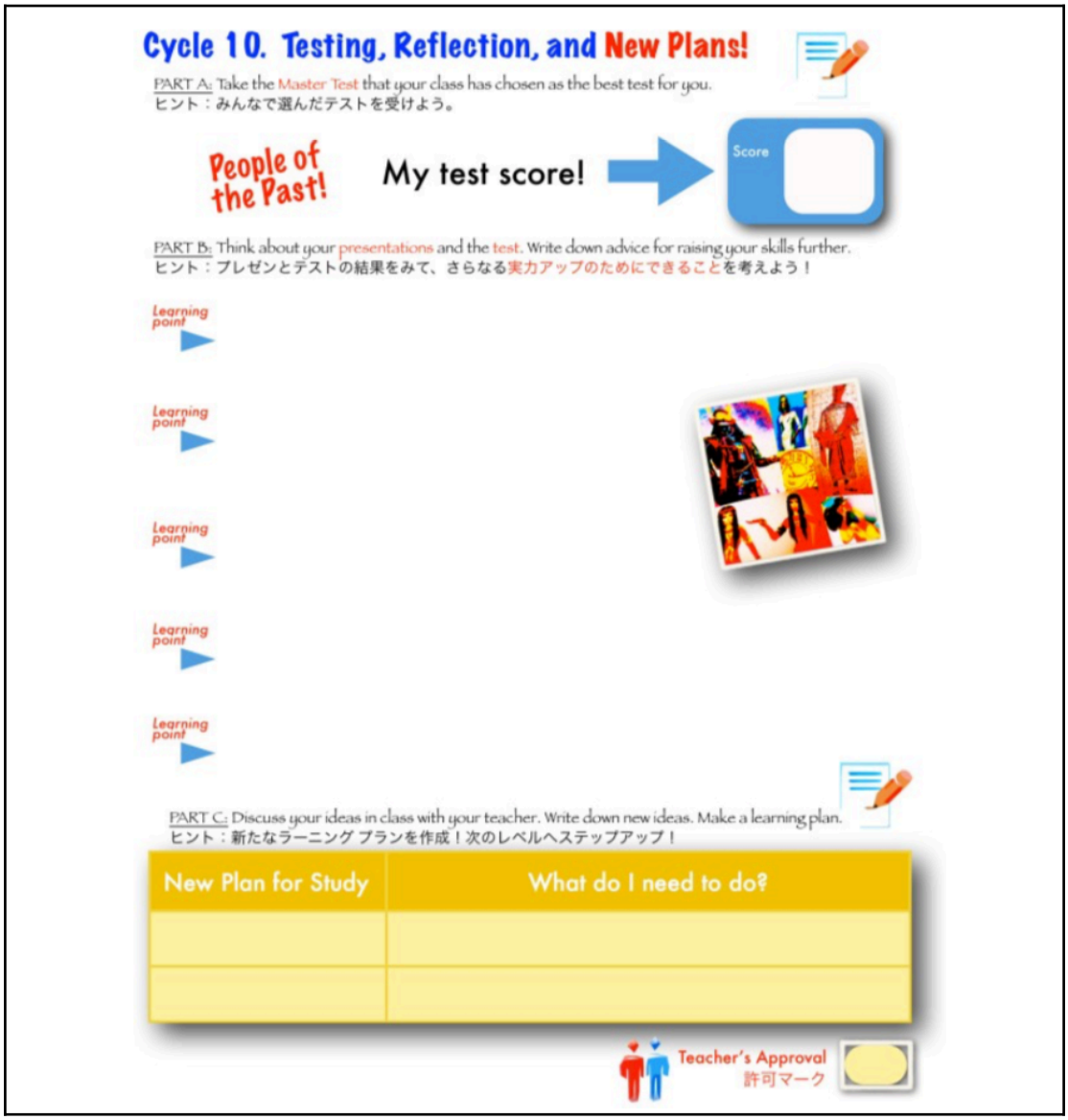


Figure 41: Copy of page 27, “Cycle 10”

This is the final cycle in the module (Figure 41). It starts off with notation of their 10-point test score. Then, it asks the learners to reflect on their presentations and their test results. Upon reflection, they then write down advice to themselves for raising their own skills even further. Five prompts are provided for this metacognitive task. From that list of five, the learners are then asked to deliberate and then focus on the two most important pieces of advice that they gave themselves. They then use those two points of advice to create two new plans for study, with

specific details on what they need to do to reach these two goals that they have set.

I chose to have the students initially come up with five ideas. The intent is to give them a chance to work on this as a broad metacognitive activity. Then, by asking them to deliberate and select only two, the students must learn to prioritize their thoughts. As this is one of the final parts of the cyclical routine across the module, it is hypothesized that this final and highly personal metacognitive task would be helpful in nurturing their critical thinking skills—on a very personal level. And, to help ensure that this nurturing of critical thinking is really happening, there is a *teacher's approval box* at the bottom of the page. In practice, if the two student-designed plans are not well thought out —they seem to be meaningless "fluff" so to speak, or conversely, the plans are overly ambitious for the student, the student is required to revise their plans until they reach level-appropriateness.

I believe the importance of this "student-based study-plan negotiation time" with each of students can easily be overlooked. But, I also believe that this is one of the most important jobs for the teacher within the pedagogical design of this coursebook; it ensures student growth via strategic student-led planning, and it provides quality bonding time between the teacher and student. It is during this time that the student is strong-armed into realizing that the teacher really will not just simply "sign-off" on any study plan. It leaves the students with the impression that the teacher really does care about the growth and the individual thoughts of each student.

By the second and third Module, it becomes apparent that several—if not the majority—of the students begin to devise well thought out level-appropriate plans. (I would like to add that this pedagogically integrated growth-expecting bonding session is especially important for the apathetic university students who must study in huge lecture halls where there is little to no direct teacher-student communication, nor is there any expectation of individualized care from the teacher.)

Page Evaluation (see Table 15)

This is a repeating cycle. It is recognized--and therefore mostly assumed that the level of critical thinking that occurs in this reflection task when Module 2 and Module 3 comes around, progresses to deeper levels than during Module 1 (their first attempt). The boundaries: There is a distinctive pattern of boundary-building designed into this page. At first, the students are asked to generate five learning points from this module (this entails a metacognitive process relying on a genuine review of their past work). But then, they are asked to cut this list down to just two points, and design self-advantageous plans around those two points. With this boundary-building and plan-building strategic task, the students are being asked to grow cognitively while the boundaries tighten. This design embodies a fundamental essence of the *Optimized Ownership Protocol.*

Ownership comes naturally because this page is literally all about—*and only about*—their own progress as language learners and how *they* plan new strategies for their own linguistic skill growth for their own futures. Because of the inherent requirement for honest metacognitive reflection (this task requires each student to reflect upon the path they took to get to this point, and find ways to improve their skills based on that reflection) it is likely that *linguistic decisions, procedural decisions,* and *thematic decisions* will be made by most of the students by the time they complete a teacher-approved solution to this task.

Table 15: Evaluation overview of page 27, "Cycle 10"

	Strong	Weak	Absent	Notes
Cyclical	X			
Boundaries	X			Strong essence of COOPS
Ownership	X			
Procedural Decisions	X			
Thematic Decisions	X			
Linguistic Decisions	X			
COOPS Coordination				Strong coordination with COOPS returns here, as students revisit their past work--while receiving approvals from peers and the teacher.

4.3.16 “(OPTIONAL) THINK AGAIN”

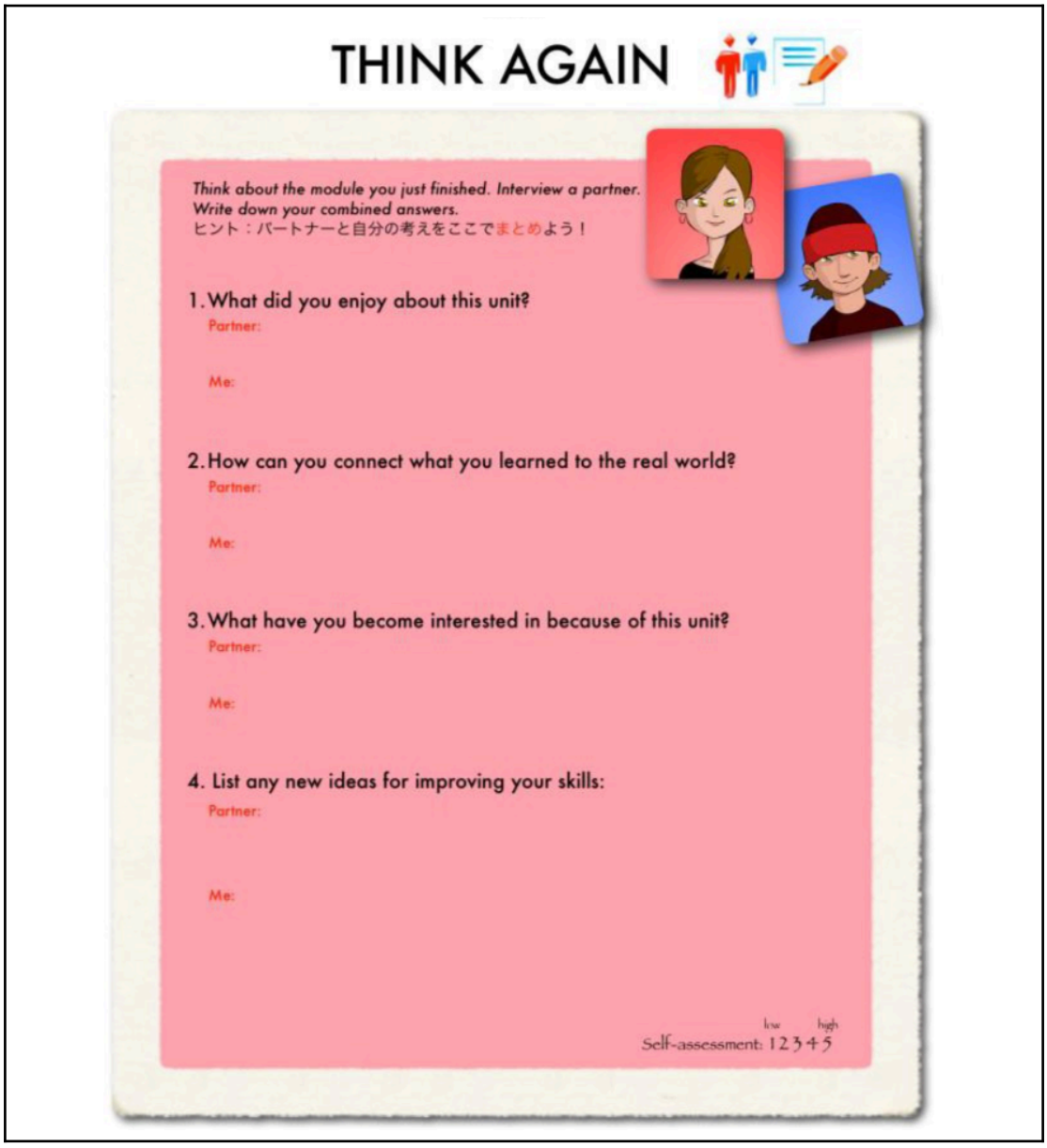


Figure 42: Copy of page 28, “(Optional) Think Again”

As an optional task, if semester time allows for it, this page (Figure 42) asks the students to once again reflect on their past learning and then *consider how to connect it to their future*. Emotions are called upon with the first question, “What did you enjoy about this unit?”. Then, real-world connections are called upon with, “How can you connect what you learned to the real world?” Metacognitive skills are called upon with the third question, “What have you become interested in because of

this unit?" This is an intriguing question because while asking the question, it is pushing the nuance that this Module *must* have affected them in some way—and coerces the student to consider such affect. And the final question is, "List any new ideas for improving your skills."

This page was explicitly designed to revisit thoughts from the previous page (Cycle 10), but it builds upon that revisited content with a significant twist—the questions on this page are developed as interview questions for their partner. So, what started out as a personalized metacognitive task concludes by being transformed into a peer-to-peer growth activity—based on content that was recently negotiated with and approved by the teacher. This allows the students to take this task on with a good level of confidence and authority. This interview activity is therefore likely an active confidence builder for many of the students. Moreover, this interview is designed to be conducted in their L2s (English). This ensures that personally meaningful content that is richly embedded in their personal development is being conveyed and negotiated in their L2—making their L2 that much more relevant in their lives.

The realization of this deeper level L2 relevance to their lives and L2 skill development should not be expected to come from the simplistic (yet common) language drills and grammar exercises of most major coursebooks. This may sound like an audacious statement, but to visitors of Japan it is plainly obvious that the average Japanese person's English competence and willingness to speak in English is extremely low even after an average of 6-8 years of traditional Japanese public school EFL lessons. The Japanese EFL context (in combination with Japan's xenophobic past) is one of the most significant, if not *the* most significant, impediments of English language competence and willingness to use English in Japan. To combat this diehard impediment, I hypothesize that a purpose-designed-for-EFL, cyclical, and metacognitive syllabus such as this one can attain a level of "between the lines" development, which can expedite the growth of confidence, competence, and the willingness to speak.

4.4 Concluding remarks

While *OL1* was based on what I thought at the time were the most logical steps to take for linguistic cognitive development, *OL1* did not systematically provide ample maturation opportunities for the students to nurture a singular thickening strand of efficiently compounded learning that could lead to enhanced cognitive development in their L2. The revisions that made its way into *OL2* still allow the learners to go through their learning in steps in a logical sequence (from a cognitive development perspective), but *OL2* users are also systematically provided with the opportunities to make critical pedagogical choices at major milestones in each Module. The learners are given the power to assess and choose what is important about the language *and* content (the procedures and the themes), making the learning have a focus on product *and* process.

All of this happens in a well-controlled environment. However, the students are led to focus on the autonomy and growing responsibilities that they are being provided with, so they likely do not notice that they are working within a well-controlled environment.

Learners acquire the skills to assess each other's progress (as part of a "*trinity of assessment*"→ solo, peer-to-peer, and teacher-student assessment). They gain feelings of empowerment from these assessment tasks, and, because they are allowed to evaluate both solo and group progression, these tasks seem to be naturally motivational as they create a cycle of positive growth within the students. Learners then move on to the designing, practice, and performance of their group presentations, giving them a strong sense of closure and fulfillment, stemming from their feelings of mastery and autonomy.

There is also an emerging "team sport-like" element that strengthens their task engagement. Next, the learners take their voted-upon self-designed tests (fulfilling the need for a solid "academic" activity), and finally, they self-assess and peer-assess their progress while keeping in mind better ideas for upgrading their presentation skills for the upcoming module. The teacher then approves those new plans.

Compared to *OL1*, student-reflection time in *OL2* has more than doubled. This amount of reflection and planning time spent after the "climax" of the module seems to be a key element in deeper cyclical L2 growth. I cannot overstate the importance of this area in the syllabus design. By the end of each module, much of each student's *personally meaningful* content, which is now becoming richly embedded in their personal development, is being "glued together" *with their L2.* In *OL2*, their L2 is the means to further their cognitive development, so their L2 naturally takes on a large role in the lives of the students.

As discussed above, if students are "stuck" in an EFL context, their EFL course must somehow compensate for the significant differences between a native speaking context and an EFL context. When in an EFL context, only a purposefully designed compensation strategy with a planned syllabus that not only closely follows human cognitive development steps, but also allows for rich and meaningful autonomous growth via feelings of ownership, can attain this level of L2 development and L2 usage with efficiency (substantial growth in a comparatively short time). This seems to be what the *OL2* package is accomplishing while fulfilling the theoretical needs and requirements that were outlined in the previous two chapters (as directed by the COOPs protocol). The later half of this thesis will display and discuss data from various sources that support these notions.

Before that, Chapter 5 will stand as the final literature review chapter in this thesis. In it I will go into deep discussion of cognitive growth theory, how cognitive growth can be tracked longitudinally, and why it is imperative for language teachers to be knowledgeable about the dynamic changes in growth rates when in in high vs. low support contexts.

5. Cognitive assessments: Lectica, Dynamic Skill Theory, Knowledge Quartet, and Microdevelopment

One essential component of this doctoral study is the review of how and why it may also be beneficial to assess the growth of students' cognitive complexity as expressed in their L2s. And, why this might be a cutting-edge path for syllabus designs and the evaluation of EFL courses. Focusing on "demonstrated cognitive complexity in the L2" is a non-traditional approach to assessing language learning that requires a robust and purpose-designed assessment protocol for it to be truly useful. This chapter reviews the tool (Lectical Assessment) that was adapted for this assessment protocol. It will also discuss the the theoretical underpinnings of this type of analysis. Although Lectical Assessment was specifically developed to study complexity growth changes in participants' native language (their L1), I will argue that the tool can be readily "re-tooled" for L2 contexts and for gauging the efficacy of EFL coursebooks. (Note: A large portion of the content from Chapter 5 has been previously published within conference proceedings and with professional development workshop materials.)

5.1 An introduction to Lectica – and, why use it for L2 contexts?

Although I had known about Lectica as a company, and although I had known that their work was built on neo-Piagetian theory (namely, Fischer's Dynamic Skill Theory) through my time at the Harvard Graduate School of Education, I never saw a compelling reason to use Lectical Analysis in my own research, nor in my own teaching. However, I had an epiphany one day and traveled to the Lectica headquarters in Massachusetts, USA—to see if the co-founder (Theo Dawson) would share the same enthusiasm that I had regarding the use of their analysis

system for L2 contexts. I discussed my idea with Dawson, and she agreed to two things: (1) I had brought her a novel approach to the use of Lectical Analysis, and (2) that she would be interested in helping me conduct this research. She was equally interested in the results of my proposed study.

5.1.1 How does it work?

Lectica provides cognitive complexity assessment marks on a scale from zero to 13, in decimal increments, denoting the cognitive complexity of each student's responses to interview questions. From the responses, Lectical Assessment provides growth scores in the following areas:

- ❑ Vocabulary, spelling, grammar, capitalization, punctuation
- ❑ Argument Relevance
- ❑ Repetition
- ❑ Fragments
- ❑ Persuasiveness
- ❑ Judgment justification
- ❑ Theme deliberation
- ❑ Theme disagreement
- ❑ Theme evidence
- ❑ Theme inquiry

Lectica's analytical system of assessment is based on Dynamic Skill Theory principles. The specific conditions for the ruler increments of the assessment system are directly lifted from Dynamic Skill Theory. I will provide a review of Dynamic Skill Theory in Section 5.3.4. Table 16 shows Lectica's ruler in the ranges from Grade 8 to Post-graduate. With over 45,000 collected responses to their tests over the past two decades, Lectica has been able to concretely define what answers look like for each 1/4 of a level (0.25); previously, definitions in Dynamic Skill Theory were only marked at integer levels. Regarding Lectical coders' accuracy, according to

Dawson, "Our minium standard is 85% agreement between raters within .2 of a level" (Dawson, 2018). As will be seen in the table below, the decimal point may be dropped altogether--for example, a score of 11.25 can instead be written as 1125.

Table 16: Lectical ruler: Grades 5 to Post-doctorate (Lectica website, 2018). Used with permission.

Lectical Range	When making decisions, individuals should be able to...	Coursework demands	Role types
1200+	• identify and leverage patterns across multiple systems to simplify decision-making under complexity • employ a wide range of strategies and tools that optimize solutions while mitigating human limitations		multinational CEO, Leader of large country
1180-1225	• identify patterns across multiple systems • embed iterative, distributed decision-making processes, • reduce impediments to good decision-making under complexity		corporate CEO, larger organization
1160-1205	• take into account the numerous nested interconnections among variables that characterize highly complex situations • expertly employ iterative, distributed decision-making processes • deftly determine the level of collaboration or perspective seeking that's optimal for a given context	Post-doc	executive, CEO medium-size organization
1140-1185	• identify multiple relations between *nested* variables • take into account change over time • consider the level of collaboration and perspective seeking that is optimal for a given context • maintain awareness of common impediments to good decision-making in organizations	Ph.D., M.D.	senior leader, CEO small organization
1120-1165	• work with both individual and group perspectives • identify and attempt to balance competing factors, such as human needs versus organizational needs • identify implicit and contextual causes • identify common impediments to good decision-making in organizations • shift frame of reference to explore alternate ways of seeing a problem	M.A.	upper level leader
1100-1145	• identify and seek the perspectives of *relevant* stakeholders • take a skeptical approach (because there is always bias or faulty information) • identify multiple relations between variables • identify own frame of reference • maintain self-awareness	undergraduate	mid-level leader
1080-1125	• identify and avoid basic forms of bias • attempt to remain objective or impartial • identify several psychological or interpersonal causes • weigh multiple perspectives or factors	grades 13–16	entry level leader

1060-1105	• consider the motives or biases behind perspectives, evidence, or actions	grades 10–13	pre- management
1040-1085	• consider pros and cons • identify a few psychological or interpersonal causes	grades 9+	
1020-1065	• compare evidence • try to keep an open mind	grades 8+	skilled labor
1000–1045	• consider consequences • think about what's reasonable or logical	grade 7+	
980–1025	• listen to or understand two sides of a disagreement • distinguish between facts and opinions	grade 7-10	semi-skilled labor
960–1005	• think about facts, evidence, or opinions • compare opinions	grade 6-7	unskilled labor
940–985	• try to understand other people's ideas • think about what makes sense	grade 5-6	

The questions Lectica asks are either subject-based (such as *science class* questions based on their current science classroom content), or the questions are built around philosophical conundrums such as can be seen in Kolberg's morality questions *("Should you steal a drug to save your spouse's life if you have no means to pay for the drug? Justify your response either way."*). The responses that are received for all of these questions are run through the Lectical Analysis system, and scored on the zero to 13 complexity scale, in decimal increments (Table 16). With this Lectical assessment system, the assumption is that you *can* choose to have a one-off student assessment, but the system really begins to shine when you have a longitudinal plan—a plan that methodically assesses participants' growth over months or years.

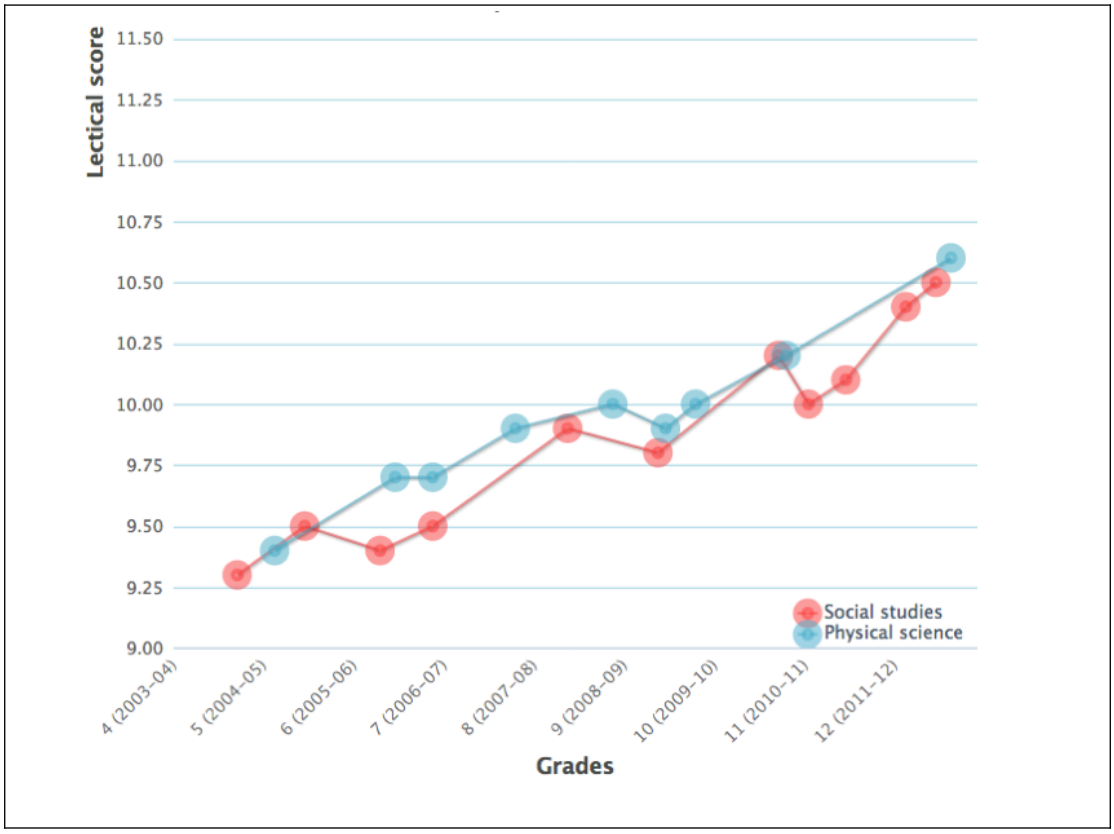


Figure 43: Growth patterns per student are domain specific(Dawson & Stein, 2011). Graphic from Lectica website, used with permission.

One typical format is to implement a pre-test and a post-test per teaching term, such as a semester, or year—then compare the charted complexity scores, per participant, per designated term (Figure 43). In this way, a participant's demonstrated growth can then be charted for however many terms you wish to follow them for.

Note that each student's growth pattern and speed is fully domain specific. That is to say that the realities derived from this form of cognitive growth assessment is far from the overly simplistic *G-scores* derived from common IQ tests. This is apparent in Figure 43.

Although the general trends between the two lines in the figure may seem to be close at first glance, there are significant divergences in growth between Social Studies and Physical Science over time for this particular student; the student's performance around the 5th grade was similar across the two subjects, but we see a gap between 6th and 7th grades, and if we follow those marks, we can see that the

student's Social Studies marks are almost always lower across the years that were tested. An IQ test would not be able to provide data at this resolution, nor would it be able to differentiate the differences between subject performances.

Moreover, it should also be noted that these Lectical Analysis scores can be divergent from students' official school marks. This is because Lectica assesses each student's demonstrated in-depth understanding of the tested for subjects (expressed in Table 16). Schools on the other hand, assess their students however they see fit, according to their own school policies. It is unlikely that a standardized school assessment format would include a system of cognitive growth analysis (akin to Lectica's).

In real life terms, a student who does well with rote-learning may be evaluated highly within, say, a Japanese public school's system, yet they may rate poorly on a Lectical Analysis test. Why the discrepancy? These discrepancies can occur because on the Lectical test, the provision of rote-memorized information alone does not demonstrate an in-depth understanding of the content, and therefore the student would attain a rather low (understanding) score. In this way, Lectical Analysis sees though superficial learning; this evaluation system refocuses the goal of education from simplistic rote-memorization to *applied learning*. What becomes necessary for the students are the skills to spontaneously demonstrate their own in-depth thinking based on their in-depth understanding of the studied content.

5.1.2 The Lectica "epiphany" and the rationale behind it

Because the nature of the Lectical test is to assess students' spontaneous responses to cognitively complex questions that are level-appropriate per student, these question-based assessment interviews needed to be conducted in the participants' native language. Interviewing them in a non-native language would skew the data in unforeseeable ways—making the data unreliable. Lectical assessments had, therefore, not been conducted in EFL contexts prior to my project. My epiphany was, however, to flip this canon on its head. I wanted to use Lectical Analysis for my EFL students *in their L2*s.

Why test in their L2?

I have never been fully satisfied with traditional L2 assessment systems because, based on my own observations of language teachers in Japan, it tends to become more pragmatic for teachers to *teach to the tests* than it is to help students develop practical fluency (and raise their willingness to speak in English). Developing competent fluency in an EFL context is of course an uphill battle; this is an unavoidable consequence of teaching in an EFL context.

Moreover, high-stakes language tests are an inextricable part of Japanese high school and university entrance exams. Who would go to "English *fluency* classes" to prepare for these rote-based Japanese entrance exams? In Japan, that would be absurdly impractical – and therein lies a multi-faceted problem. The irony of Japanese students' high scores on standardized proficiency tests—yet extremely low *communicative competence*—is a long-standing theme in EFL education in Japan.

In stark contrast, Lectical assessment in the L2 would not allow for *teaching to the test*. As stated above, Lectical assessment sees though superficial learning. So, the only way to prepare for a Lectical exam in the L2 would be to train the students to have in-depth reasoning skills demonstrable in their L2s. *"Training for in-depth reasoning skills demonstrable in their L2s"* would literally be raising their real-world L2 competence. Therefore, the only viable "cheat" would turn them into better L2 language users. That is not a "cheat"— that is a win-win scenario.

5.2 Lectical Assessment's validity

As discussed above, Lectica's ruler is based on Fischer's Dynamic Skill Theory. The ruler has 13 levels of criteria denoting demonstrable levels of cognitive complexity. Regarding validity, it must be understood that Lectical scores are based on responses supplied by the participants; there are no EEGs, fMRIs, or other such brain analysis apparatus being used concurrently with Lectical assessment test.

Lectical scores can therefore only tell us *indirectly* what seems to be happening in the brain. This may not seem like a significant point, but it is. Unlike a

brain scan that can produce data regardless of the cognitive state of the participant, Lectical Analysis can only be as good as the data that is being provided by the participant. Moreover, the participant must be willing to do their best on each test for the results to be a useful measurement of their growth.

Lectical assessment is not only accepted in academic settings, but it is also popular in corporate and government settings. Figure 44 was downloaded from the Lectica website; it is a description of levels that appear in their corporate tests that are designed to see if applicants for a job are "upper management material," or not.

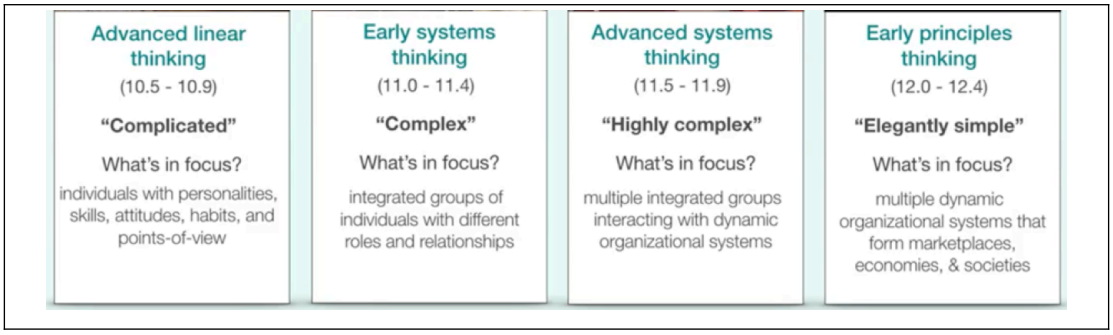


Figure 44: Four out of thirteen levels. Adults typically score somewhere in this zone.

The popularity of Lectical assessment in corporate and government settings reflect well on Lectica—and more to the point, show how this form of assessment cuts through superficial (practiced) knowledge and evaluates people on how much in-depth thinking they can spontaneously demonstrate. This is precisely what I wanted to see being evaluated in EFL classrooms. I wanted to see how much in-depth thinking EFL students could spontaneously demonstrate in their L2s, per pedagogical style, for a real show of pedagogical efficacy and to understand EFL students' growth potential (baseline growth).

5.2.1 Implications: A better way to look at communicative competence?

Is this the birth of a new paradigm of language proficiency testing? --one that better matches cognitive development? Traditionally, EFL teachers assess L2 proficiency with tests focusing on the "correctness" of vocabulary or the "correctness" of grammar. Moreover, even when there is a progressive form of

assessment, such as an oral exam, the assessment is typically conducted as if students' internal *thinking* in L1 came first, and then the *expression* of those thoughts came second, translated over to their L2 (Kelley, 2012).

The (unspoken) assumption seems to be that there is a linear progression to language learning—and that assumption mandates how that language usase is to be pragmatically assessed. However, from brain studies, it seems that we are regularly producing a myriad of *simulations* of experiences, based on past experiences (Murphy, 2017); the brain does not work in a linear format—it is much more organic and free-flowing.

There is an evolutionary purpose for these free-flowing and continued flow of organic mental simulations. These spontaneous simulations prepare us to be able to quickly respond appropriately to all potentially upcoming events, in prepared anticipation for what *might* happen in our immediate context—these simulations (much like chess players anticipating several moves in the future) cannot be linear to be of any practical use. From a linguistics perspective, the ability to communicate productively and coherently with others requires the ability to be mentally prepared with multiple simulations and be able to mentally manipulate linguistic concepts before they are mouthed. This is how proficient native language speakers engage in fruitful dialogue.

Therefore, the goal of the EFL language teacher should be to nurture their students into also being able to mentally manipulate concepts with spontaneity and to be able to produce logical statements (or questions) that reasonably correspond with the context they happen to be in.

5.2.2 But, why not linear?

The "beauty" of our capability to mentally simulate and manipulate high concepts is lost when the assessment of language skills and the teaching of language is assumed to be linear; the most pragmatic approaches with linear language teaching will strip away the myriad of divergent deeper thinking processes that should have occurred otherwise.

Linear teaching and linear assessment take the most direct approach to the linearly set goal. This is problematic, yet pragmatic. A pragmatic linear approach cannot be the best way to prepare the students for language usage in the real world, but it may likely be one of the best ways to prepare the students for upcoming standardized tests.

This argument illuminates the point that real-world communicative competence requires more than just a "reputable level" of vocabulary, grammar, and pronunciation. The concurrent provision of the same "reputable level" of vocabulary, grammar, and pronunciation to all the students in the classroom is what I would refer to as "equality-based teaching." On a theoretical level, "equality-based-teaching" can be described as being diametrically opposed to what I refer to as "*equity*-based teaching."

Suitable adaptive communicative competence requires a working knowledge of the contextual appropriateness of the language to be used. However, a working knowledge of contextually appropriate language is not easily taught *linearly* in rote. This is something that requires *dynamic* nurturing--in a living context as a life skill. For this to happen, multiple simulations and multiple mental manipulations must be allowed to occur in the students' minds dynamically, in context; students must therefore be put into contexts that will naturally nurture individualized in-depth thinking in their L2s.

The nurturing of individualized and dynamic in-depth thinking in the L2 is the essence of "*equity*-based teaching." With this argument, what now comes to light is this thought-provoking pedagogical dichotomy (equality-based language teaching vs. equity-based language teaching) that have dissimilar fundamental goals that evidently produce entirely different qualitative outcomes.

My current study highlights that not only can we gauge the dramatic differences in these teaching styles' outcomes, but that we can also use the ensuing data to help us develop a preliminary baseline of what "normal" cognitive growth progress looks like per student, and from there, work toward improving the quality and efficiency of that cognitive growth.

Traditionally, L2 teaching (and assessment of competence) per student has been conducted with little to no direct comparison with each student's L1 prowess. With no understanding of what each student is capable of in their L1, I fear a significant chunk of data is missing—data that would be helpful in better assessing each student's L2 competencies. This is why equality-based teaching can be so problematic. Correlation with students' L1 competence should not be disregarded--it should be centerstaged.

From a neuroELT perspective, equality-based teaching lacks attention to individualized growth patterns. The lessons are taught in low support contexts (meaning that the teaching does not conform to the individual needs of each student). In contrast, I propose that teachers should track what each of their students are capable of achieving in their L1—to gauge each students' genuine cognitive potential in linguistics—and then, using that data as a baseline, strive to close the gap between their demonstrable competencies in L1 with their demonstrable competencies in their L2. This approach would require a high support context that would assure *equity*-based teaching, which I believe is ideal for language students in EFL contexts.

5.2.3 In sum

Although it is ideal from a pedagogical perspective, asking an EFL teacher to assess each of their student's L1 competence every semester may not be a realistic endeavor. The second-best option, which is much easier to achieve, is for teachers to establish a rule of thumb baseline of what typical L1 growth looks like at the age group(s) they are teaching. Table 16 (p.192, from Lectica) is well suited to help establish this baseline. However, to really understand and assess pedagogical designs at the highest level, it is also imperative to learn what a satisfactory L2 growth *rate* for EFL students per age range is, and use those rates as benchmark rates for comparison with the newly emerging L2 growth data.

The assessment of *demonstrable growth rates of cognitive development in the L2* is a new paradigm for ELT, and it is only possible because of the complexity ruler from Dynamic Skill Theory and the trustworthiness of Lectica's data analysis. As for

my thesis, this has laid down the path for progressive coursebook design, and progressive coursebook evaluation.

5.3 Theoretical background

Traditionally, *Education* (and *Applied Linguistics*) has seldom looked to the brain for answers to its problems. This was reasonable for the past few decades because science still did not have the tools to provide answers to those questions. However, neuroscience has made impressive progress in the past few decades. Still, it can be said that neuroscience, as a form of academia, is in its infancy—especially when compared to other fields within science. Why look to neuroscience for help now?

5.3.1 The need for the neuroELT niche

Although it may still be in its infancy, attempting to learn from neuroscience is a worthy endeavor. With that in mind, Kurt Fischer at the Harvard Graduate School of Education started up his *Mind, Brain, and Education* program roughly two decades ago. The mission was not only to learn from neuroscience, but to begin a vibrant on-going dialogue between the fields of Psychology, Neuroscience, and Education. Being my mentor since my MA dissertation days, Fischer encouraged me to follow in his footsteps and establish a similar undertaking, but with the "Education" component of *Mind, Brain, and Education* changed to "English Language Teaching." Fischer's main theoretical contribution to the field is Dynamic Skill Theory – his seminal work will be reviewed in Section 5.3.4. With Fischer's mentorship, and with the help of like-minded colleagues (primarily based in Japan), we started up a niche in academia that we now call *neuroELT*.

What has neuroELT research accomplished? At the time of this writing, our growing group of several hundred enthusiasts has conducted 11 international conferences in Asia (Japan, the Philippines, Macau, and South Korea) attended by thousands of teachers from around the globe. One substantial long-term project of the group has been the establishment and continuous updating of 50 neuroELT

Maxims. These maxims have been primarily produced at our neuroELT Lab at the University of Kitakyushu with the help of interns under my supervision. The continued development of these maxims has had a strong influence on my syllabus designs and teaching.

5.3.2 The "Knowledge Quartet" repurposed

As discussed in Chapter 3, Rowland et al.'s (2005) work further proliferated use of the *Knowledge Quartet* in education assessment. The Knowledge Quartet is commonly used as a framework for the analysis and development of teaching. Being a quartet, there are four quadrants in the theoretical framework → foundation, connection, transformation, and contingency.

Each of the Knowledge Quartet categories are typically used to focus attention on that particular element of teaching. However, if the four elements were modified into a cline for university level students, it is compatible with constructivists' notions regarding cognitive development. Figure 45 shows that Foundation level teaching is equivalent to what I call equality-based teaching (and low support contexts). On the other side, we can see that Contingency level teaching is equivalent to what I call equity-based teaching (and high support contexts). This figure holds substantial implications for everything from curriculum design to classroom teaching approaches.

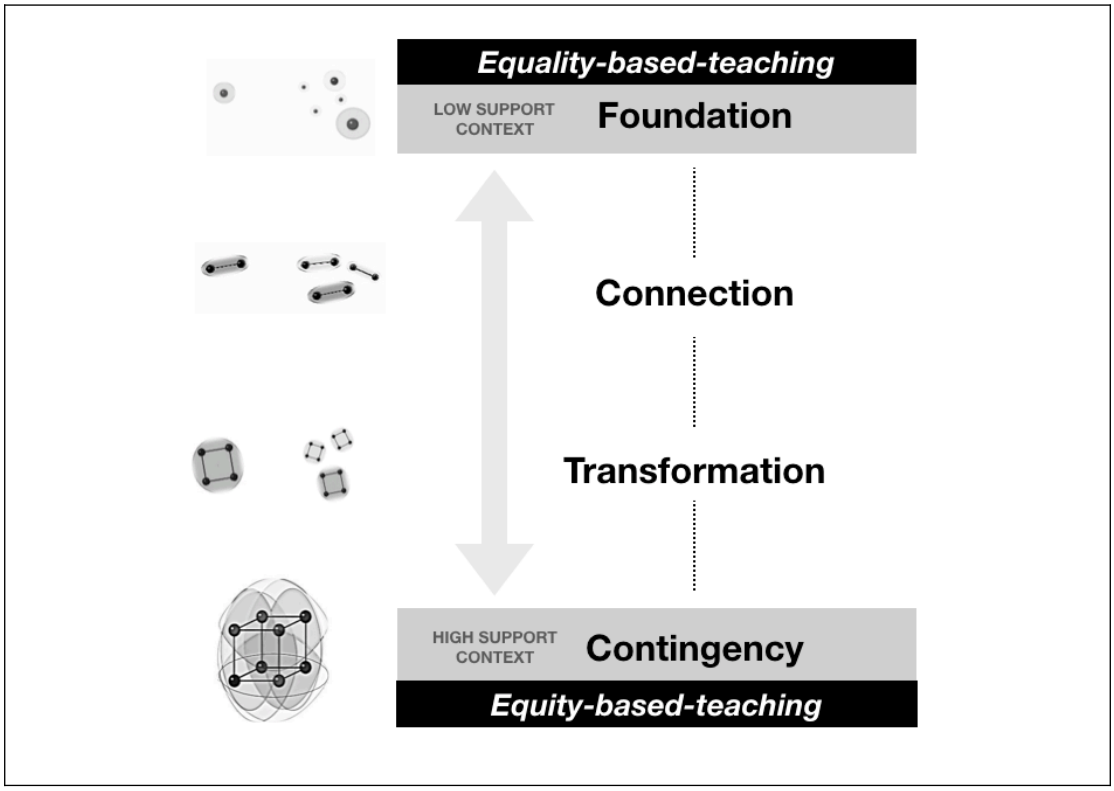


Figure 45: The Knowledge Quartet, Equality vs. Equity, and Low support vs. High support

In Figure 45, equality-based teaching (top of model) refers to the provision of *uniform* support regardless of the per-person attainability of that set goal. On the other extreme, equity-based-teaching refers to the provision of *personalized* support for each student to reach the set goal. This figure helps to visualize of how mainstream equality-based coursebooks differ from *OL2*, which is equity-based.

Upon review, I have concluded that many mainstream coursebooks aim at the Foundation level of teaching—providing information in a clear format that is pre-determined and therefore easily testable. Such an approach cannot venture much further down that cline in the figure.

In contrast to those coursebooks, *OL2* resides at the bottom of the cline, at the Contingency level; it asks both the student and the teacher to be spontaneous and self-matching with the learning content. There are no standardized tests for the learning because the specific learning points are not determined—the assessment must match the needs of the students dynamically, as those needs arise. In this way, Figure 45 depicts the gradation between equality-based and equity-based teaching

approaches, but what does this mean in practical terms? Below are four practical examples:

The Foundation level (at equality), provides the same coursebook content, time, and level of support per student. The burden of learning at this level is on the shoulders of the learners. It is the learners who are expected to make personal adjustments to fit the learning and make progress toward the learning goal. This level of support is known as a *low-support context* in Dynamic Skill Theory.

The Connection level (somewhat less equality), provides professional assistance in making meaningful connections between knowledge points, per student. This is a significant step above the *Foundation* level because now the burden of making those explicit connections is on the teacher, meaning at least some form of student personalization will be necessary to cultivate these connections within the students.

The Transformation level (approaching equity), provides the next step up from the *Connections* level. The teaching now systematically affects the students *transformatively*. This level of transformation is not likely to happen by chance, so there must be a significant amount of teacher-led equity building tasks that dynamically conform to the students' needs. In short, the teacher is likely providing individualized support for meaningful transformation to occur.

The Contingency level (full equity), there are significant dynamic adjustments *being made for the learners in real time*. This high level of catering to the personal needs of the learners is fully dynamic and self-adjusting. It is therefore diametrically opposed to the tactics at the *Foundations* level. This level of support is known as a *high-support context* in dynamic skill theory.

This section discussed some of the main differences between mainstream coursebooks and *OL2*. The next sections will discuss the mechanics of these differences.

5.3.3 How knowledge develops in the brain

From a biological perspective, how are learned connections being made in the brain? Is there an untapped solution to making these connections more robust? All too often there seems to be a "white elephant in the room" when it comes to professional development in Education. Treating the brain as a *black box* has been the long-standing norm (Dörnyei & Murphy, 2010). In Education journals and even teacher workshops, we tend to write, read, and discuss the content of papers that are focused on "*who said what about whom*," while not enough wordage is spent on how memories are physically formed in the brain. This is a peculiar phenomenon. It would seem brain function should be center stage in Education papers, conferences, and workshops. However, this is often not the case.

The reason for the exclusion of brain function from the literature is not much of a mystery—it is only very recently that neuroscience has been detailed and has become reliable enough to provide suitable data for use in Education research (Tokuhama-Espinosa, 2010). Therefore, historically, the bulk of mainstream Education literature has been composed without tedious notation of incremental developments in neuroscience.

Will the field of Education always be this way? No, I believe that is unlikely. *Educational Neuroscience* (inclusive of Harvard's *Mind, Brain, and Education*) is a relatively new field of roughly 20 years, but it is a growing field that is explicitly designed to rectify these issues. So, today that black box syndrome is perhaps more of a "grey box syndrome" instead (Dörnyei & Murphy, 2010:21).

In that spirit, this section shall bring to this literature review two components from Educational Neuroscience, with the intentions of: (a) helping balance the literature review and (be) to help to bring a few new perspectives to the table for deliberation.

Making connections: Human memory and microdevelopment

Human memory is notoriously unreliable (Schacter, 1997; Shaw, 2016), and our ability to *recall* information lies in the efficacy of our neuronal networks. Our

networks' efficacy is affected by, and dependent on, our immediate context—*and the results are often false anyway* (Gallo, 2006). In short, compared to computers, human memory is imprecise. So, how do humans learn anything with reliable accuracy? How does domain-specific knowledge development work?

To answer these broad questions, I should first address *micro*development. This is indeed the study of development, but it is focused on how stable connections are made in the brain on the scale of seconds, minutes, and hours—instead of weeks, months, and years. Demonstrable skill levels build up due to interaction with people and objects in context, but the skill levels tend to drop and rebuild, over and over again (Figure 46). This is the reality of microdevelopment—and, this is an inherent issue concerning the foundation of all learning.

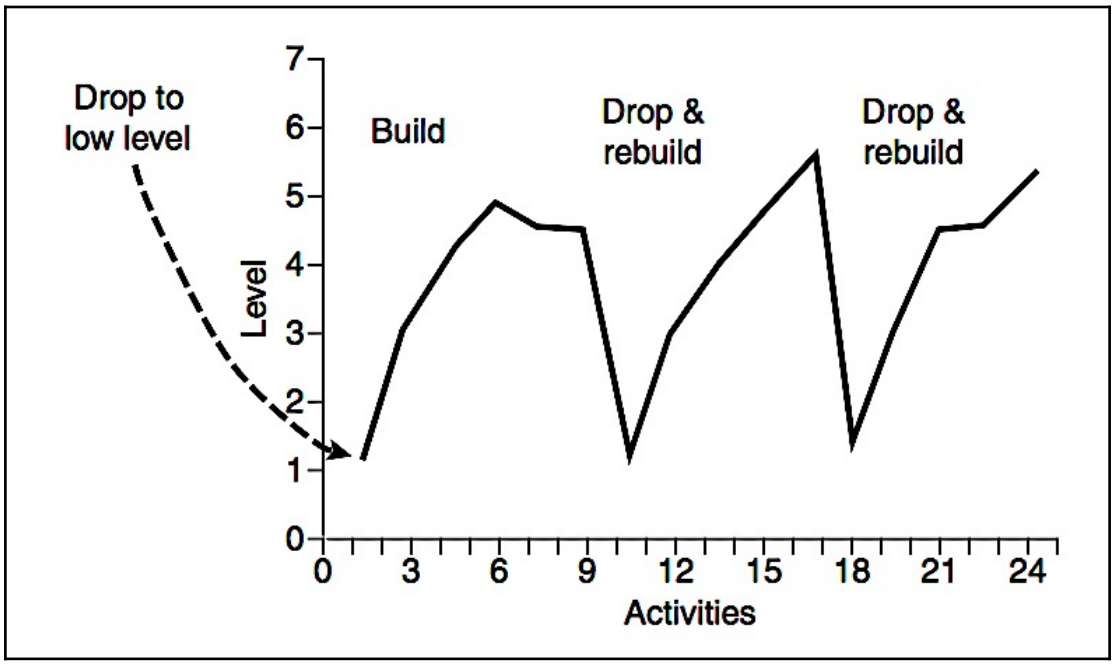


Figure 46: Microdevelopment: Natural peaks and dips (Fischer, 2008:143)

The next matter of relevance is that those peaks and dips in Figure 46 change their waveform as time passes and interchanges occur. Fischer has settled on identifying three representative waveforms, signifying the developmental models of novice, intermediate, and expert (Figure 47).

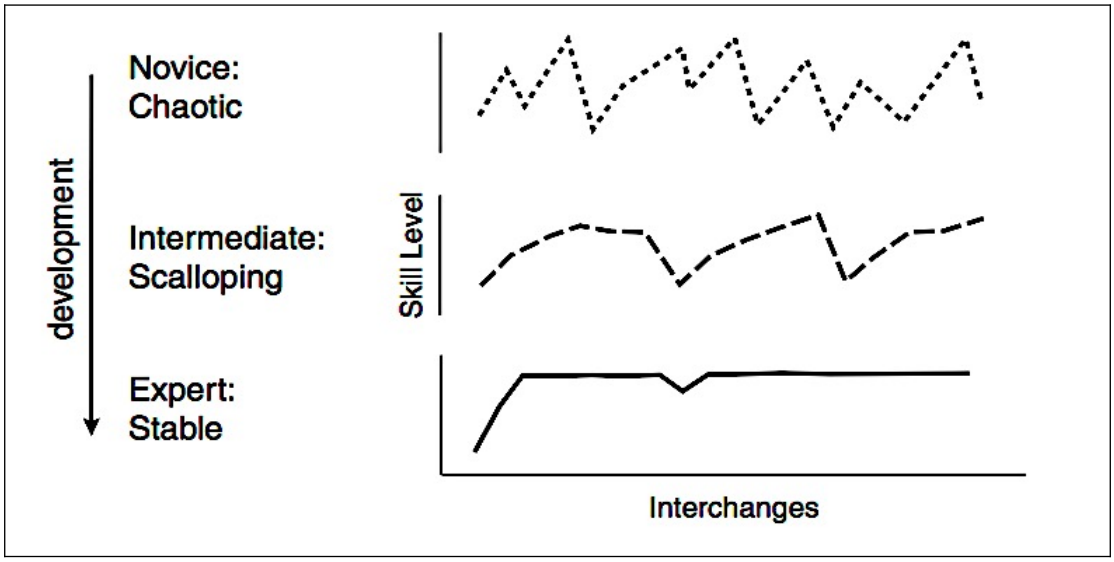


Figure 47: The three waveforms identified in microdevelopment (Fischer, 2008:144)

Without knowledge of the waveforms, teachers will likely (automatically) blame any number of other contextual variables. Those other contextual variables may very well be partly to blame for the peaks and dips in classroom performance, but teachers must be trained to realize that unstable outcomes are also due to naturally occurring biological functions—and that these unstable outcomes are not always due to controllable external variables. This is a somber yet highly significant issue that unfortunately seems to still be relatively unknown. So, to help remedy this matter for teachers in training, Murphy (2017) produced the following recommendations (Figure 48) for use in microdevelopment training:

(a) Teachers must realize that the consolidation of learning is never concrete. It is an ongoing *dynamic process.*

(b) Demonstrated levels of knowledge are the most erratic at the novice stage — it is not advisable to base firm conclusions on achievement during the novice stage.

(c) When a student displays a stable and consistent level of success, this may indicate that they are near, or already at, their mastery stage.

(d) For the most accurate assessments, there should be day-to-day longitudinal records. The outcome of a single test should not be used to judge levels of mastery.

(e) Levels of mastery should be judged by how stable a performance the student can provide over a *longitudinally-designed cycle of evaluations*.

Figure 48: Microdevelopment-based recommendations for teachers (adapted from Murphy, 2017)

In sum, from the literature, it is clear that teachers and coursebook authors ought to observe the nature of their own short-term development as well as students' short-term development. Erratic performance is the norm in the early stages of any skill development because the neural connections do not come readily.

Plateauing (a string of connections working efficiently) is a sign of mastery—but even then, dips should be expected; dips in performance should not be automatically penalized, nor automatically taken as an indication of ill-conceived pedagogical design. Moreover, longitudinally-designed cycles of evaluations, explicitly looking for the plateauing of skills, is the recommended method of assessment at the microdevelopment level. With that in mind, I shall move the review to how developmental connections are made on the “macro” level.

5.3.4 Dynamic Skill Theory – the basics

How do our developmental connections build on the "macro" level (over a semester, a year, or longer)? Fischer and Bidell (2006) discuss how primary developmental "growth sites" take 3-4 years to systematically cycle around the cortex (Figure 49), as can be measured by EEG data in the alpha band. Interestingly, the timing of the completion of each 3 to 4-year cycle perfectly matches Piaget's levels of development. This is not a coincidence; it is substantial scientific evidence in support of Piaget's seminal work (Fischer & Bidell, 2006: 388).

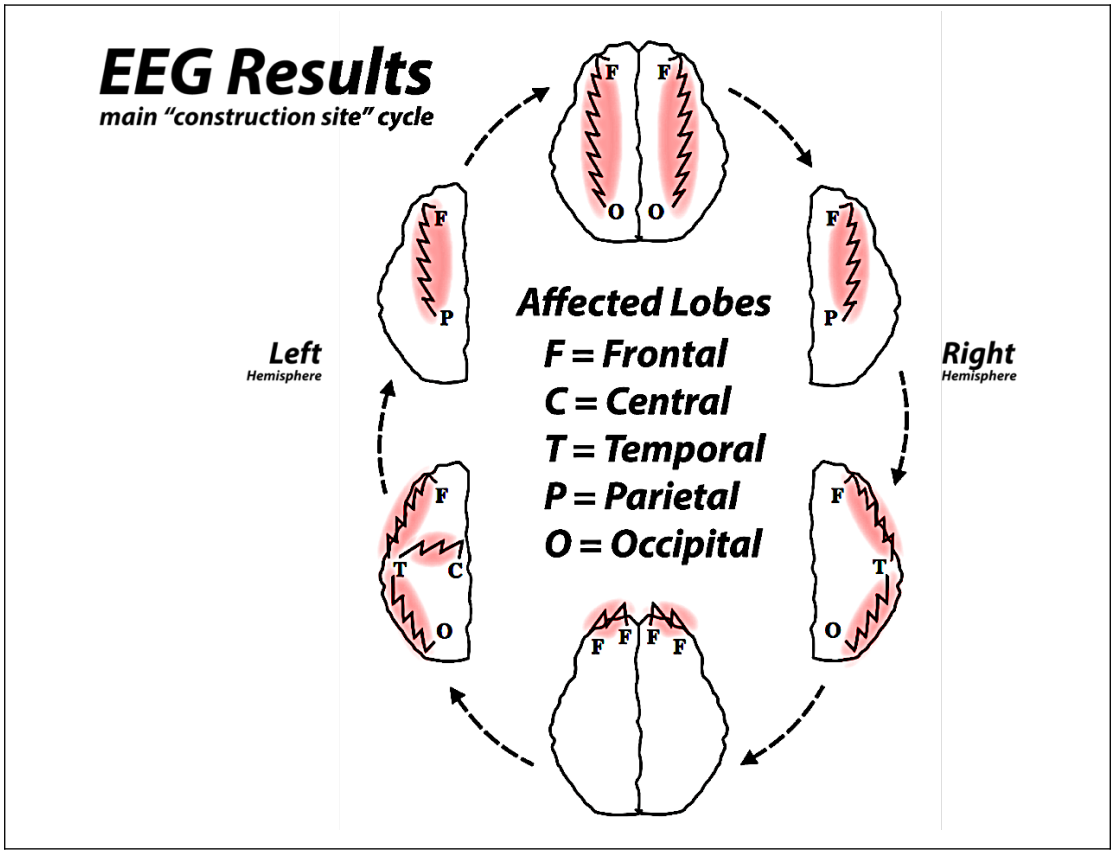


Figure 49: A graphic representation of the primary "construction sites" uncovered by EEG scans in the alpha band (adapted from Fischer & Bidell, 2006: 389)

With dynamic skill theory, Fischer (1980) was able to further refine Piaget's work regarding developmental stages and was able to define and differentiate optimal level trajectories (attainable only in learning contexts of "high support"), from functional level trajectories (attained in "low support" learning contexts). This

tangible differentiation is what was missing from Piaget's work.

For Piaget, the apparent lack of consistency in demonstrable skills trajectories was a thorn in his side. Fischer was able to solve this Piagetian mystery by realizing that there are separate demonstrable skill trajectories that are affected by the quality of the contextual support provided to the participant. This phenomenon is illustrated in Figure 50. I hypothesize that this "Piagetian mystery" still commonly baffles classroom teachers—and likely provides ammunition for Piagetian theory opposers. Fischer's solution to this mystery needs to be more widely circulated in Education circles.

Note: At the specific ag timing that each full rotation cycle in Figure 49 completes, the solid red line in Figure 50 peaks. That red line in Figure 50 is the optimal level trajectory, or, the trajectory of maximized potential at any given age (and as discussed above, conclusively lines up with Piaget's stages of development). Unlike the results of the microdevelopment data graphing, at the *macro* level of learner development, there is no plateauing within high support contexts across the developmental years.

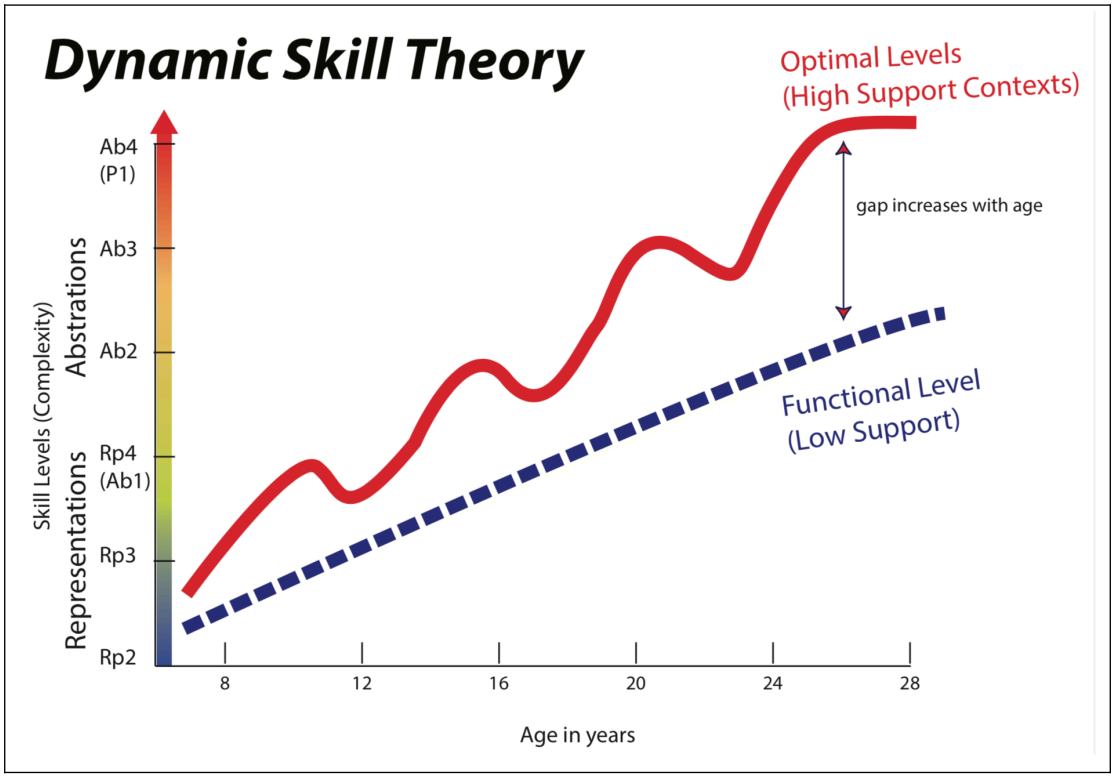


Figure 50: Dynamic skill theory: Optimal levels and functional levels (adapted from Murphy, 2009)

Research in longitudinal (macro)development, suggests that teachers should be taught that even when the best possible pedagogy is employed by the best possible tutor, it is still unlikely that students will show lengthy plateaued demonstrations of understanding (display a stable understanding) during their developmental years. In fact, over time, the students will demonstrate major dips in skill level—this is inevitable and unavoidable. It is imperative for this phenomenon to be taken into account while assessing students—and when evaluating the efficacy of the employed teaching strategies.

Moreover, the dotted blue line in the figure signifies the general trajectory demonstrated by students in (the much more common) low support contexts. In relation to the implications made regarding the modified Knowledge Quartet discussed earlier, the dotted blue line here represents student outcomes when at the *Foundational* level, while the solid red line represents student outcomes when a perfectly-adjusted-for-the-student *Contingency* plan is implemented. As alluded to

earlier, the implications of these differences are profound; these substantial differences in pedagogical approaches and their consequential outcomes should not be glossed over but should be embraced by teacher trainers and researchers for what they exemplify regarding pedagogical design.

Furthering "Connections" in the Knowledge Quartet

This final portion of Dynamic Skill Theory discussion further strengthens its association with the Knowledge Quartet. According to Fischer (1980, 2008) and other neo-Piagetian constructivists, all in-depth learning depends upon on the development of neuronal connections. After humans learn singular concepts, we then can progress on to a mapping of those two concepts (be it parity or disparity between the two concepts). Then, when a few mappings emerge in the brain, if the context calls for it, the mappings can move on to form a system of understanding. Finally, when the context calls for it, those various systems of understandings can bond together, forming what is called a system of systems of understanding (Figure 51). This is a well-documented flow of events that humans follow for learning any skill in any domain (Young, 2011).

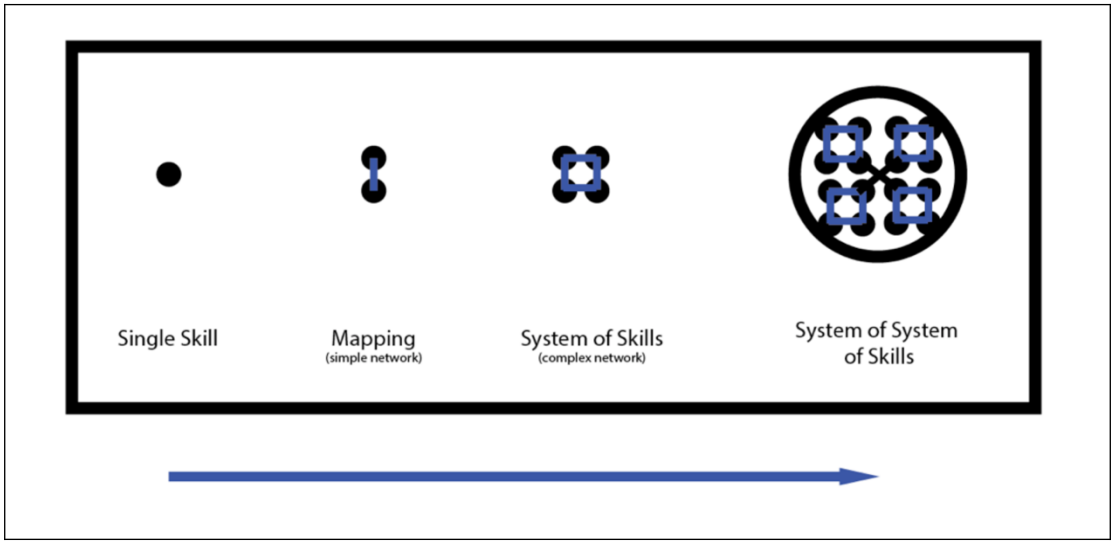


Figure 51: Four steps in development in dynamic skill theory (Adapted from Murphy, 2015)

As mentioned above, *Connections* in the Knowledge Quartet resonates well

with constructivist (and neo-Piagetian) ideas. Technically, Connections in the Knowledge Quartet corresponds well with what Fischer calls *Mappings*. While attaining new skills in any domain, it is not biologically possible to skip multiple steps of skill building—there are no shortcuts for skill building (Fischer & Bidell, 2008). Therefore, the most efficient and pragmatic teaching solution would be one that is configured with the skill theory pattern in mind--moving from single skills to mappings, to systems, and so forth, as in Figure 51.

To further explore how this works from a practical vantage point, Figure 52 displays the learning skills necessary to play the sport of baseball. The figure graphically exemplifies how a multitude of smaller individual skills need to connect and compound on the very long journey toward the ultimate goal of being able to play a game of baseball.

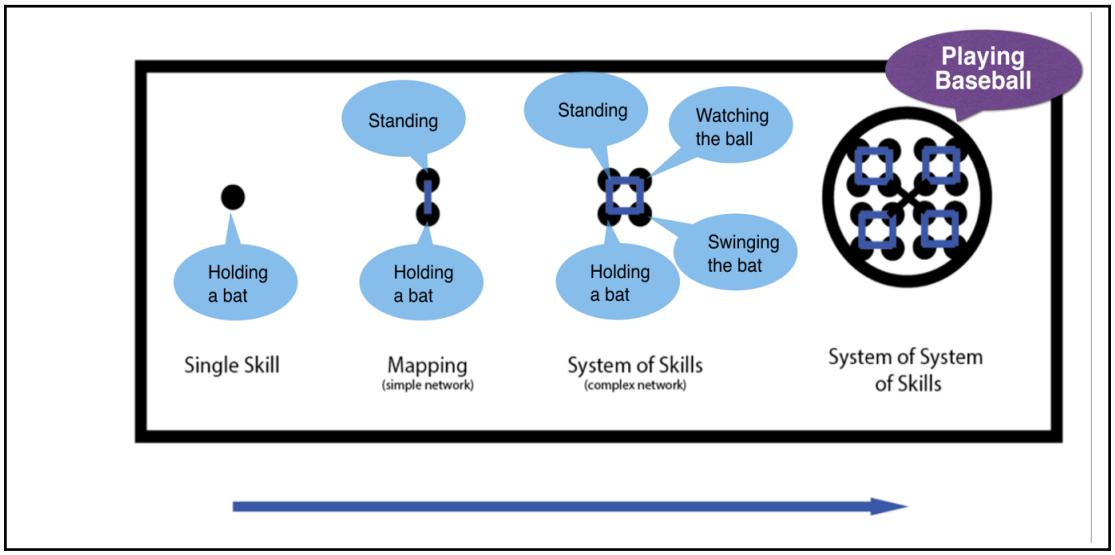


Figure 52: Using the sport of Baseball to exemplify how individual skills connect and compound to arrive at the goal. Several systems of skills must be created and then meshed by the end to finally achieve a "system of systems" – enabling the child to finally play baseball. (Adapted from Murphy, 2015)

From this practical example, it is evident that a total novice in baseball cannot begin to effectively play baseball with cognitive intent alone. Similarly, it would not be possible for an external person (such as a teacher or coach) to effectively *will* them to suddenly be able to play baseball either. All of those smaller individual points of understanding (each necessary skill) must be founded though

experience, then deliberately and carefully connected to build into mappings, and then into systems of skills.

In the end, if all goes well internally and contextually, a system of systems of skills *may* emerge—creating a customized set of skills specifically tailored for that person to be able to play their first game of baseball. All teaching (and assessment of pedagogy) should be based on this evidence-based "customized pattern" of skill building that leads to deeper learning.

Understanding, and following the patterns of customization that lead to learning, should provide the best return on investments, from the pedagogical perspective. This is neuroscience-based gravitas that is robustly interrelated to the meaning of Connections in the Knowledge Quartet as we move on to the next section.

5.4 In sum: How all of this informed the design of OL2

OL2 was designed with Dynamic Skill Theory's patterns of cognitive development in mind, albeit within an overarching cyclical format that binds the lesson outcomes together as an organic continuum, with the intent of creating the most learner-efficient design possible—hence the design and inception of the COOPs protocol. This intent may have produced a uniquely designed coursebook (at least on the theoretical level)—but as they say, the proof is in the pudding.

How shall the pudding be tasted?

A typical coursebook rubric is not designed to detect the types of innovations that are embedded into the *OL2* design, so readily available evaluation strategies would not be suitable for this task. More importantly, regarding student outcomes, there are no common standardized English test formats that are specifically designed to gauge the quality and specific type of cognitive growth that is hypothesized to emerge from *OL2*'s usage; the tracking of growth patterns in spontaneously demonstrable cognitive complexity per student in their L2s is not a part of the test criteria in any of the standard language tests that I have come across in my teaching

career. In fact, "The tracking of growth patterns in the cognitive complexity per student as demonstrated spontaneously in the L2s" is not even a genre of discussion within Applied Linguistics yet. Even within *neuroELT*, this dissertation marks the first large scale attempt at gauging EFL students' cognitive growth as demonstrable in their L2s. How can this complexity of demonstrable growth in the L2 be systematically and reliably assessed?

The answer to this conundrum is *Lectical Assessment*. Lectical Assessment is the only currently available system with the data and usable knowledge-base to accomplish this task competently. As such, is not a coincidence that Lectica was founded on constructivist paradigms, namely Fischer's (neo-Piagetian) Dynamic Skill Theory. Although the founders of Lectica had not initially planned on using their evaluation techniques to gauge L2 growth in EFL students, as will be revealed in the later half of this thesis, Lectica helped produce robust results for this study, and I am happy to report that there is now a growing body of data regarding the application of Lectical Analysis in second/foreign language contexts.

6. Research Design and Methodology

6.1 Summary of Research Design

As has been discussed thus far, over the course of the past decade, I have been tinkering with an idiosyncratic teaching methodology, evolving evaluation protocols, and a corresponding series of textbook design implementations (*OL2*). This study concerns the evaluation of the new teaching methodology by administering *OL2* to Japanese learners of English, all of whom were studying under my supervision, across two semesters (2014-2015). In total, I had *three* different groups of students, each group using a different incarnation of my coursebooks. The third group (which was only in the second semester) used a slightly revised version of *OL2*; this *OL2 Revision* (*OL2-R*) was based on my first-semester research conclusions. This is the epitome of—and the strength of—*action research* in action.

In this way, several groups spanning two semesters were studied. During this process, I collected a wide range of data using five primary instruments. Data was collected from: cognitive-based measures, language learning achievement measures, student evaluation measures of my own design, centrally administered standardized questionnaires designed and administered by the university, my own qualitative focus group interviews, and my students' year-end final essays (See Section 6.3 for a flowchart of events). The five instruments used for this data collection were as follows:

- Main student questionnaire (*online*)
- Lectical assessment (online)
- Focus group interviews (face-to-face)
- Standardized university questionnaires (paper-based)
- Student essays (paper-based)

The next sections shall clarify the exact flow of events—providing the details of the interventions and students' activities throughout the on-going and elaborate action research process. Also, see the *Limitations* section for a list of isses that were uncovered regarding these tools while doing this research.

6.2 Participants

6.2.1 Semester 1: English Course Participants

Seven courses at a university in southern Japan were selected for this study in the students' first semester of the year. Enrollment, gender ratios, and coursebook choice are listed in Table 17. Each class was comprised of roughly 30 students. The participants were mixed groups (*International Relations, Business, English-Chinese Studies, Humanities, English, and Law*), yet they were chosen for this study because they had relatively similar English competencies. All students grew up in Japan and had at least six years of Japanese school English education. The total participant count was 164 students. Enrollment figures are in Table 17. The *OL1* group (comprising of one out of the seven courses) had 27 students across the two semesters that were followed. The remainder, the *OL2* group (inclusive of *OL2-R*), had 137 students. Gender ratios for *OL1*, *OL2*, and *Total* were 15:12; 59:78; and 74:90, accordingly.

Table 17: First semester enrollment, gender ratios, and coursebook choice.

Coursebook	**Course Count**	**Semester 1 Enrollment**		
		(Total)	***(Male)***	***(Female)***
OL1	1	27	15	12
OL2	6	137	59	78
Total	**7**	**164**	**74**	**90**

6.2.2 Semester 1: Lectical Assessment participants

The Lectical Assessment participants were a subset of the 164 participants described above. Due to cost constraints, I chose a subset comprising of two representative courses (See Table 18), each using a different coursebook; Group 1

used *OL1* while Group 2 used *OL2*. Those two groups were comprised of similar first-year students at the university. Due to participant attrition across the semester, the final participant count for the Group 1 was 24 with a male-female ratio of 12:12. The final participant count for the Group 2 was unchanged at 30, with a male-female ratio of (12:18).

Table 18: The two classes chosen for Lectical Analysis

Group	Coursebook (OL1)	(OL2)	Semester 1: Lectica Pre-test Participant count	Semester 2: Lectica Post-test Participant count
1	✓		24	24
2		✓	30	30
Total	**1**	**1**	**60**	**60**

6.2.3 Semester 1: Focus group participants

For my focus group interview sessions (Table 19), 4 students per course (7 courses) were recruited on a volunteer basis. Each focus group had 4 participants per course—however, all but 2 focus groups had a 50:50 male-female ratio. The other two were 3:1 (female to male), and 4:0 (female to male). The focus group discussion minimum time allocation was 30 minutes, but the discussion often went on for a full hour or longer in a few cases. The interviews were bilingual; the participants were asked to freely choose the best language to enunciate their thoughts in.

Table 19: The seven focus group: numbers, gender ratios, and coursebook choice.

Group number	Coursebook *(OL1)*	*(OL2)*	Focus Groups, 30-60 min. discussions *(Participants)*	*(Male)*	*(Female)*
1		✓	4	1	3
2		✓	4	2	2
3	✓		4	2	2
4		✓	4	0	4
5		✓	4	2	2
6		✓	4	2	2
7		✓	4	1	3
Total	**1**	**6**	**28**	**10**	**18**

6.2.4 Semester 1: Main student questionnaire participants

As all of the language classrooms used had one computer per student, all 164 participants in all seven classes in the first semester participated in the online questionnaire (Table 20). However, as will be detailed below, due to attrition, not all of the 164 participants' data from the first semester was analyzed for this study.

Table 20: The main student questionnaire: numbers, gender ratios, and coursebook choice.

Coursebook	**Courses**	**Semester 1: Main Questionnaire**		
		(Participants)	***(Male)***	***(Female)***
OL1	1	27	15	12
OL2	6	137	59	78
Total	**7**	**164**	**74**	**90**

6.2.5 Semester 1: Standardized bi-annual university student questionnaire participants

The university mandates their standard paper-based questionnaire every semester (biannual). All seven classes participated in this questionnaire (Table 21).

Table 21: University questionnaire: numbers, gender ratios, and coursebook choice.

Coursebook	**Course Count**	**Semester 1 Enrollment**		
		(Participants)	***(Male)***	***(Female)***
OL1	1	27	15	12
OL2	6	137	59	77
Total	**7**	**164**	**74**	**89**

6.2.6 Semester 2: English Course Participants

My teaching schedule was significantly altered in the second semester of the year; only three of the original seven courses continued through to the second semester. Luckily, most of the students in those continuing courses carried over to the second semester, so longitudinal data collection was possible. Details are provided in Table 22. Additionally, and significantly, the revised version of *OL2 [OL2-R]* completely replaced *OL2* in the second semester.

Table 22: Second semester: Coursebook, enrollment, and gender

Course	***OL2-R***	**Semester 2 Enrollment** *(Participants)*	*(Male)*	*(Female)*
1	✓	30	13	17
2	✓	21	10	11
3	✓	32	20	12
Total	**3**	**83**	**43**	**40**

As in semester one, each course was roughly comprised of 20-30 students. As most students from these particular courses carried over from the first semester, none of the demographics changed significantly.

6.2.7 Semester 2: Final Focus Group participants

Nearing the middle of the second semester, the participants of the *first-semester* focus groups came forth and requested to have continued focus group sessions in their-then current second semester. So, at the request of the students, we arranged times to conduct one final focus group session for all three remaining courses; I incorporated that additional data into my study (Table 23).

Table 23: Semester two: Final focus group interview

Course	***OL2(R)***	**Semester 2 Enrollment** *(Participants)*	*(Male)*	*(Female)*
1	✓	4	2	2
2	✓	4	2	2
3	✓	4	1	3
Total	3	4	5	7

6.2.8 Semester 2: Main student questionnaire participants (Mid and Post)

As in the first semester, the classrooms had one computer per student. All present students were required to take part in the online questionnaire. There was one questionnaire session at the beginning of the second semester, and another questionnaire session at the end of the second semester. The sessions are titled MID and POST, respectively (Table 24).

Table 24: Semester 2: Main student questionnaire participants (Mid and Post)

Course	SM2: Main Questionnaire "MID" Participants	SM2: Main Questionnaire "POST" Participants
1	29	42
2	27	25
3	36	40
Total	**86**	**107**

6.2.9 Semester 2: Year-end "My Change" essays

Semester two participants were asked to reflect upon their changes from their first day in the first semester to their current state; the assignment was to write a one-page essay on how they changed over the entire two semesters (Table 25). As this was the second semester, all participants were now using *OL2-R*.

Table 25: Second semester: Coursebook, enrollment, and gender

Course	*OL2-R*	Semester 2 Enrollment *(Participants)*	*(Male)*	*(Female)*
1	✓	30	13	17
2	✓	21	10	11
3	✓	32	20	12
Total	**3**	**83**	**43**	**40**

6.3 Instruments and Procedures

A general overview of the procedures: For semester one, although there were seven different courses, all six courses that were using the *OL2* followed the same schedule, without deviation. The sole *OL1*-based course followed a similar, but slightly different schedule. Below is a side-by-side comparison of the *OL1* vs. *OL2* schedules (Table 26), including all the components of the data collection:

Table 26: Comparing OL1 and OL2 teaching schedules

***OL1* Schedule**	***OL2* Schedule**
Class 1: Introduction, Online Questions, Essay	Class 1: Introduction, Online Questions, Essay, **HOMEWORK pg. 6**
Class 2: Section 1- 3. **HOMEWORK** Project Book -2 projects	Class 2: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book 2 projects
Class 3: Online English Assessment day, **HOMEWORK,** Project Book 2 projects	Class 3: Online English Assessment day **HOMEWORK,** Project Book 2 projects
Class 4: Section 4-5, discussions; **HOMEWORK** Project Book do 2 projects	Class 4: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice
Class 5: Section 6-7, *Presentation designing time;* **HOMEWORK** presentation practice!	Class 5: ***Presentations!*** *Cycle 7* - 9; **HOMEWORK,** Project Book do 2 projects
Class 6: ***Presentations!***	Class 6: Cycle 10, ***Test!*** Discussion; no homework
Class 7: Section 1- 3. **HOMEWORK** Project Book do 2 projects	Class 7: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book do 2-3 projects
Class 8: Section 4-5, discussions; **HOMEWORK** Project Book 2	Class 8: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice
Class 9: Section 6-7, *Presentation designing time;* **HOMEWORK** presentation practice!	Class 9: ***Presentations!*** *Cycle 7 -9* **HOMEWORK,** Project Book do 2-3 projects
Class 10: ***Presentations!***	Class 10: Cycle 10, ***Test!*** Discussion; no homework
Class 11: Section 1- 3. **HOMEWORK** Project Book do 2 projects	Class 11: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book do 2-3 projects
Class 12: Section 4-5, discussions; **HOMEWORK** Project Book do 2 projects	Class 12: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice
Class 13: Section 6-7, *Presentation designing time;* **HOMEWORK** presentation practice!	Class 13: ***Presentations!*** *Cycle 7* - 9; **HOMEWORK,** Project Book do 2-3 projects
Class 14: ***Presentations!***	Class 14: Cycle 10, ***Test!*** Discussion; no homework
Class 15: Online Questions, Essay, Discussion, Final exam preparation	Class 15: Online Questions, Essay, Discussion, Final exam preparation
Week 16: **Final exam (online)**	Week 16: **Final exam (online)**

As can be seen in Table 26, each 90-minute class had a variety of interactive activities plus frequent homework with their Modules (chapters), that consistently climaxed with a group presentation, three times per semester. For semester two, there were only three participating courses. They used *OL2-R*, with the following schedule (Table 27):

Table 27: The schedule from the second semester

***OL2-R* Schedule**
Class 1: Introduction, Online Questions, Essay, **HOMEWORK pg. 6**
Class 2: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book do 2 projects
Class 3: Online English Assessment day **HOMEWORK,** Project Book do 2 projects
Class 4: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice
Class 5: ***Presentations!*** *Cycle 7* - 9; **HOMEWORK,** Project Book do 2 projects
Class 6: Cycle 10, ***Test!*** Discussion; no homework
Class 7: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book do 2-3 projects
Class 8: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice
Class 9: ***Presentations!*** *Cycle 7 -9* **HOMEWORK,** Project Book do 2-3 projects
Class 10: Cycle 10, ***Test!*** Discussion; no homework
Class 11: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book do 2-3 projects
Class 12: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice
Class 13: ***Presentations!*** *Cycle 7* - 9; **HOMEWORK,** Project Book do 2-3 projects
Class 14: Cycle 10, ***Test!*** Discussion; no homework
Class 15: Online Questions, Essay, Discussion, Final exam preparation
Week 16: **Final exam (online)**

6.3.1 Main student questionnaire (SurveyMonkey)

The instrument:

SurveyMonkey is an online research tool, often used for custom designed surveys. For this study, I devised a 38-question survey for the students to take in pre, mid, and post sessions, primarily to collect their background data, to test for changes in self-assessment marks, and to collect open-ended question responses regarding their personal development, and opinions about the pedagogy of the course (Figure 53 and Table 28). The survey was designed to be administered and to be taken bilingually; instructions were bilingual and responses were accepted in either language (Japanese or English).

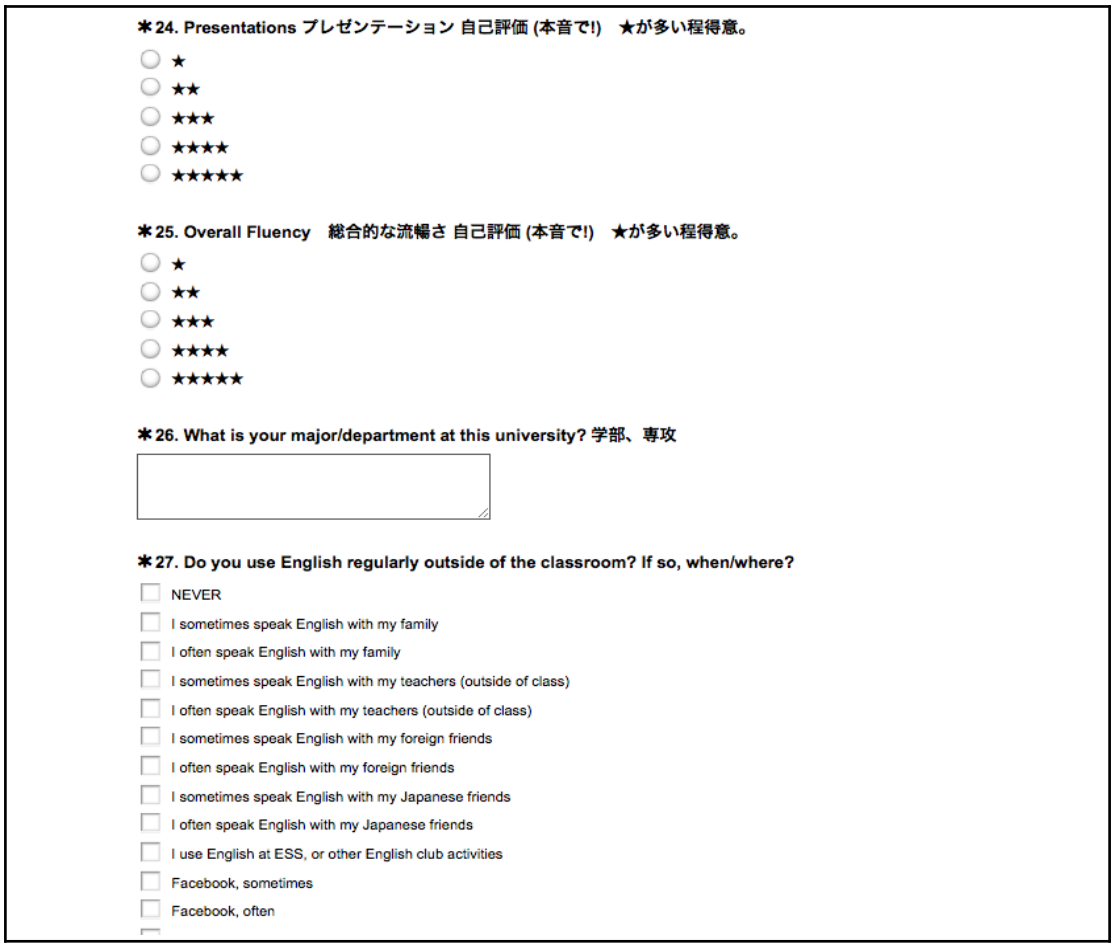


Figure 53: Screenshot from one portion of the main student questionnaire

Table 28: Main topic range coved in the questionnaire

Demographics	**Self-assessment (rated 1-5)**	**Projected English usage (likeliness rated 1-5)**	**Reflection (rated 1-5 & open-ended)**
Name	Reading	...with future family	Overall improvement
Age	Writing	...with future friends	Explanation
Gender	Listening	...with future business	
Department	Speaking	...with future internet usage	
+ misc variables	Grammar	...with future hobbies	
	Vocabulary		
	Spelling		
	Presentation Skills		
	Overall Fluency		
4 variables (+)	**9 variables**	**5 variables**	**2 variables**

Procedure:

1. Participants were assigned to take the survey in CALL (computer) classrooms up to three times across the span of two semesters—once at the beginning of semester one, once at the beginning of semester two, and once at the end of semester two.
2. With one student per computer, the students were required to log into the *SurveyMonkey* server and access my online survey.
3. Instructions on how to respond to the questions were provided bilingually. Personal assistance was also provided for participants who had trouble understanding the system and/or the questions being asked.
4. The remainder of the 90-minute class time was used to complete the survey as thoroughly as possible.
5. A 2-minute stand up and stretch time was scheduled at the 45-minute mark. Participants were allowed to chat briefly, to relieve stress and help them relax. This was to help combat potential fatigue issues.
6. At the end of the 90-minute class, all participants' work was confirmed verbally and online.
7. Responses recorded on the *SurveyMonkey* server were converted into SPSS

files, cleaned up, and statistically analyzed—resulting in quantitative and qualitative data.

6.3.2 Lectical Assessment

The instrument:

Lectica's online "RFJ assessment tool" is used in American school grades 4 through 12. For its own development and reliability, the RFJ test has undergone longitudinal testing involving over 8,000 students worldwide. RFJ psychometrics are robust; it can reliably distinguish 5-7 levels of performance in a typical classroom. I have provided a full example of this tool in the Appendix.

The online RFJ test is a battery of questions (two sets of different content but with identical design and complexity, issued randomly across the students) that takes 60-80 minutes to complete. It is a set of questions that focuses on reflective judgment (focusing on evidence, inquiry, deliberation, disagreement, perspectives, conflict resolution, persuasion, and truth). The RFJ test proposes a dilemma (scientific or social, depending on the test number) then documents how the participants pursue *resolution* of the dilemma. Each student's written dilemma resolution tactics and justifications reveal the level of cognitive complexity that they reached while in pursuit of their dilemma resolutions. The received responses were scored with the Lectical Assessment System by Lectical Analysts who were blind to my experimental conditions. The RFJ test is typically administered in English. However, for this study, the instructions were provided bilingually to maximize the reliability of the study.

In this way, the *cognitive complexity* of the written English responses of each participant was professionally assessed, in pre- and post-test contexts—providing trustworthy growth data, albeit in an unheard of manner—via demonstrable *L2 cognitive complexity*, and not by standard TESOL measures.

What is coded?

- ❑ explicit logical organization of the arguments
- ❑ implicit hierarchical order of abstraction of the concepts

Spelling, grammar, vocabulary, and response word counts were among other data sets that were collected by Lectica and analyzed by the coders, but those measures were not used for this study because those measures were designed primarily for native English-speaking test-takers, and did not suit this study.

By administering a pre-test (beginning of semester) and a post-test (end of semester), it was a simple matter to set up cognitive growth comparisons across the board for all of the participating students, and to see exactly which students demonstrated improvement, how much improvement they displayed, and in which specific areas they displayed their improvement. The coded data resulting from Lectical Analysis also made it possible to directly compare the efficacy of *OL1*'s usage against *OL2*'s usage. This "contest" between *OL1* and *OL2* provided some of the most intriguing and fascinating aspects of my study.

Comprehending the results of the RFJ test

Lectical Analysis RFJ scores around the 10~11 range are normal for the tested age group. The typical growth rate across the first year of university study for native English speakers in their native country is a 0.05 rise. These growth rates fluctuate according to age and context, but understanding the above numbers and growth rates is sufficient to follow the data that emerged from this study of Japanese university students.

Procedure:

This current study used the Lectica RFJ001 and the RFJ003 test (see Appendix). Each test provides a central dilemma, (i.e., "Some scientists think that it is okay for children to watch TV shows that…") and a battery of six main questions with sub-questions designed to uncover the current cognitive capabilities/ development of each student by forcing them to provide answers to cognitively challenging questions. The pre-test was taken on week three of their first semester. The post-test was taken during the second to last class of the semester. Lectical data was only collected during the first semester. The participants were assigned to take the questionnaire in CALL (Computer Assisted Language Learning) classrooms:

(1) With one student per computer, the students were required to log into the Lectica server and access the online assessment program. Students were each provided with personal login data, and after login, they were randomly selected for either the 001 or the 003 test.

(2) Instructions on how to respond to the questions were provided bilingually. Personal assistance was also provided for participants who had trouble understanding the system and the questions being asked.

(3) The remainder of the 90-minute class time was used to complete the questionnaire as thoroughly as possible.

(4) A 2-minute stand up and stretch time scheduled at the 45-minute mark; participants were allowed to chat briefly to combat potential fatigue issues.

(5) At the end of the 90-minute class, all participants' work was confirmed.

(6) Responses were collected by Lectica, coded, and then analyzed twice by Lectica's professionally certified coders to ensure accuracy.

(7) The coded data were then statistically analyzed via SPSS software.

(8) For the post-test, the procedure was mostly identical, but the RFJ test number was flipped accordingly, so that no participant took the same test twice.

6.3.3 Focus Group: Interviews and schedules

The instrument:

Below (Figure 54) is the discussion prompting template that was used to elicit responses from the participants of the four-person focus group meetings.

Focus Group
Discussion Template

Class progress:
"How is the class doing?"
"Which are developing the best?"; "Which are the worst?"
Cycle 1:
Cycle 2:
Cycle 3:
Cycle 4:
Cycle 5:
Cycle 6:
Cycle 7:
Cycle 8:
Cycle 9:
Cycle 10:

THINK AGAIN page:
Positive things about the course book:
Possible issue/difficulties with the course book:
Recommendations/suggestions:
Learning (Pedagogy):
"What tasks worked best in the class?"
"Which tasks could be further fine-tuned?"

"Any ideas/comments/final words?"

Figure 54: Focus group discussion protocol, used for every FG session

Procedure:

For the first semester, four students per course volunteered for the Focus Group (FG) sessions. Each FG met four times during the semester, and the discussions were digitally recorded with an Apple iPhone. Each group had one introductory ("Getting to know you") FG session, then there were 3 more serious reflexive FG sessions where the FG participants reported on each cycle in the

coursebook and detailed the progression of their classmates from their personal perspectives, providing to the best of their knowledge, the pros and the cons of the course methodology and its implementation in their classroom. The template used to elicit responses can be seen in Figure 54. All interviews were conducted bilingually. They were then transcribed into English; major points were coded and listed into tables. Table 29 is the interview schedule provided to the students. FG meetings have been underlined to illustrate the pacing in relationship to their learning schedule.

Table 29: Focus group interview schedule

Semester Schedule
Class 1: Introduction, Online Questions, Essay, *recruit* Focus Group Volunteers; **HOMEWORK pg. 6**
Class 2: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book 2 projects; *conduct* Focus Group meeting
Class 3: Online English Assessment day 英語実力テスト・オンライン**HOMEWORK,** Project Book 2 projects
Class 4: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice ***発表練習を小グループで!!***
Class 5: ***Presentations! (グループ・プレゼン),*** *Cycle 7-9*; **HOMEWORK,** Project Book 2 projects
Class 6: Cycle 10, ***Test!*** Discussion; no homework; *conduct* Focus Group meeting
Class 7: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book 2-3 projects
Class 8: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice ***発表練習を小グループで!!***
Class 9: ***Presentations! (グループ・プレゼン),*** *Cycle 7-9*; **HOMEWORK,** Project Book 2-3 projects
Class 10: Cycle 10, ***Test!*** Discussion; no homework; *conduct* Focus Group meeting
Class 11: Cycles 1 - 3b; **HOMEWORK** ~3b, Project Book 2-3 projects
Class 12: Cycles 4b -6; **HOMEWORK**, Finish Cycle 6 and Practice ***発表練習を小グループで!!***
Class 13: ***Presentations! (グループ・プレゼン),*** *Cycle 7-9*; **HOMEWORK,** Project Book 2-3 projects
Class 14: Cycle 10, ***Test!*** Discussion; no homework; *conduct* Focus Group meeting
Class 15: Online Questions, Essay, Discussion, Final exam preparation
Week 16: **Final exam (online)**

During the second semester, the same four students per course volunteered themselves for a final FG session. Each FG met three weeks before the end of their second semester, and the discussions were again digitally recorded with an Apple iPhone. Each group was instructed to genuinely reflect upon their class's development over the course of the *entire* year, not just the current semester, and then report on each cycle in their new coursebook (*OL2-R*). The FG members were instrumental to the revision process from *OL2* to *OL2-R*, so these participants were particularly eager to discuss their thoughts on the revisions that emerged, and what was left unchanged.

The participants detailed how they viewed the progression of their classmates, providing to the best of their knowledge, the pros and the cons of the course methodology and implementation, comparing and contrasting the differences between the two semesters (*OL2* vs *OL2-R*). The ensuing results were coded, tabulated, and analyzed.

6.3.4 Standardized bi-annual university questionnaire participants

The instrument

My university mandates a university-designed standardized paper-based questionnaire every semester. The main questions (originally written in Japanese) are in Figure 55, translated into English. This questionnaire seems to have been designed to gauge two main areas: teachers' effectiveness, and student engagement. Participants are asked to rate each of these entries on a 1-5 scale. No special training is provided for this task.

Main questions on standardized bi-annual university questionnaire (1-5 scale):

- ❑ attendance
- ❑ gained deeper understanding
- ❑ overall satisfaction
- ❑ appropriateness of course level
- ❑ appropriate speed/flow
- ❑ appropriate content
- ❑ instructor's passion
- ❑ instructor's usage of media
- ❑ appropriateness of coursebook
- ❑ instructor's speaking style
- ❑ pre-checked the syllabus
- ❑ checked the syllabus mid-term
- ❑ instructor followed syllabus
- ❑ course was empowering
- ❑ your diligence

Figure 55: Main questions, with responses on a 1-5 scale

Procedure:

(1) Within their normal classroom settings, the participants are given a bubble mark sheet, this sheet of standardized questions, and an optional blank sheet of paper (A5) for additional notation.

(2) The instructor reads the questionnaire instructions in Japanese.

(3) Participants are given as much time as they need to fill in the bubble mark sheets and are given the option to write notes on the blank sheets.

(4) The entire process typically takes no longer than 20 minutes. When the participants have completed their forms, the forms are collected into a single envelope, sealed, and delivered to the university office.

(5) The university office tallies the evaluations and produces mean scores in Excel files (per question, per class). Per-student responses are not obtainable from the university, making it impossible to longitudinally study per-student response changes.

(6) However, Excel files with mean scores per question per class are obtainable from the university.

6.3.5 Year-end "My Change" essays

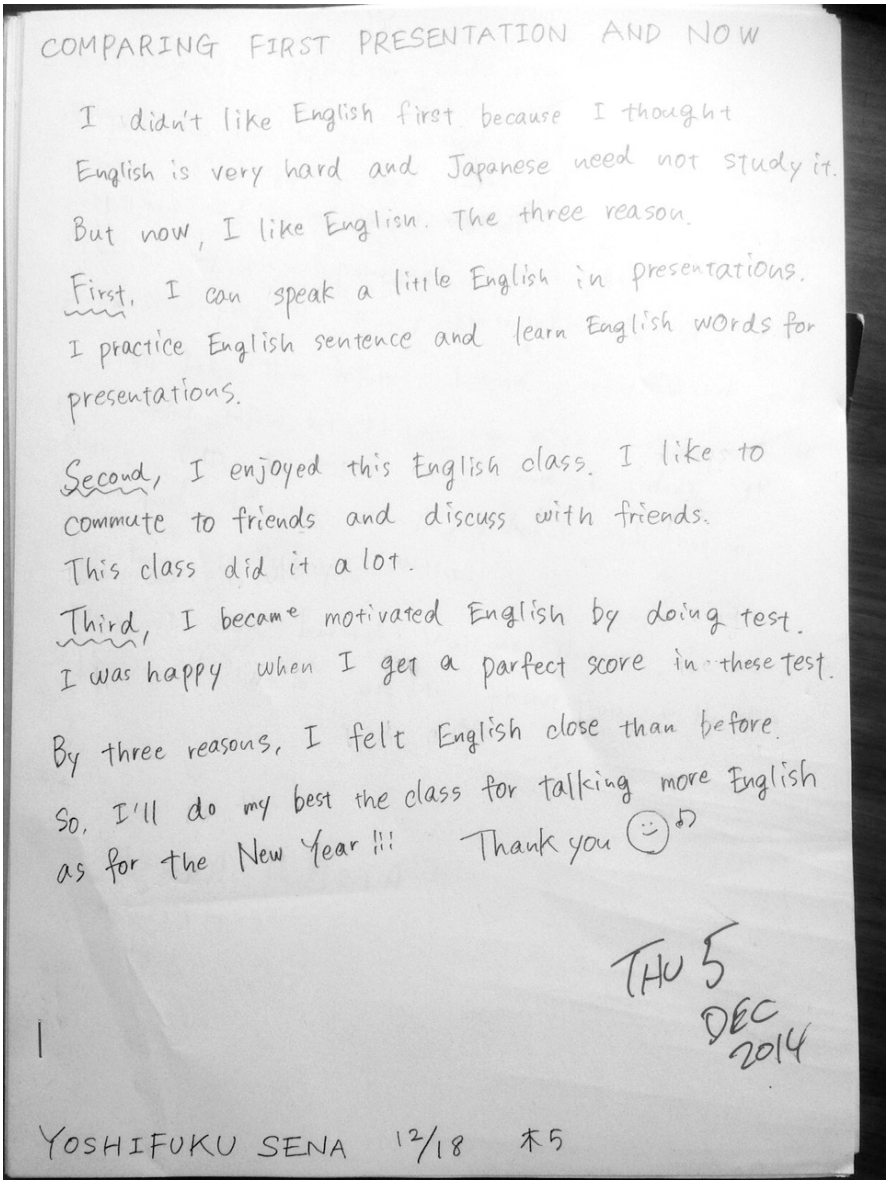
The instrument

This instrument was a reflexive one-page essay on a blank A4 sheet of paper with the writing prompt: *"Comparing First Presentation and Now."*

Procedure:

Eighty-four of the second-semester students participated in this open-ended reflexive writing exercise. As an in-class activity, the participants were asked to reflect upon their gradation of change across the two semesters. Specifically, they were given the writing prompt *"Comparing First Presentation and Now*." The assignment was nuanced for them to write up a one-page essay on what changes they noticed occurring across the two semesters and what justifications they could raise in support of those realizations.

By this time in the course, the students were accustomed to self-refection inquiries, so it was not necessary to spend time motivating them to write worth-while essays. Participants were asked to fill out the blank A4 page in an essay-like format, but there was no strict time limit for this procedure (Figure 56). All participants finished the assignment within 60 minutes. The completion times ranged from 20 minutes to just under 60 minutes.



COMPARING FIRST PRESENTATION AND NOW

I didn't like English first because I thought English is very hard and Japanese need not study it. But now, I like English. The three reason.

First, I can speak a little English in presentations. I practice English sentence and learn English words for presentations.

Second, I enjoyed this English class. I like to commute to friends and discuss with friends. This class did it a lot.

Third, I became motivated English by doing test. I was happy when I get a parfect score in these test.

By three reasons, I felt English close than before. So, I'll do my best the class for talking more English as for the New Year !!! Thank you ☺♪

THU 5 DEC 2014

YOSHIFUKU SENA 12/18 木5

Figure 56: Sample writing of the final reflexive essay. Pseudonyms are used.

6.4 Summary of main data types collected

The five instruments used:

(1) Main student questionnaire (*SurveyMonkey*)

(2) Lectical assessment

(3) Focus group interviews

(4) Standardized university questionnaires

(5) Student essays (reflexive)

Because of the complexity of the data set and the data collection schedules, I have provided a chronologically accurate summary of the data collected throughout this project, separated by semester. Data type categorizations are: qualitative, quantitative, and mixed (Table 30).

Table 30: Comparing OL1 and OL2 teaching schedules

Instrument	**Data type**	***[Semester one] Details: Variables***
Focus Group (Semester one only)	qualitative	20 main questions: transcribed, coded
Main questionnaire (Pre)	mixed	37 questions: bio data (1-14), Self-assessment ratings (15-32), course data (33-35), open-ended (36-37)
Lectica (Pre)	quantitative	6 qualitative responses coded into quantitative data
University standard questionnaire (Semester one)	mixed	16 main questions + free writing (open)
Lectica (Post)	quantitative	6 qualitative responses coded into quantitative data
Instrument	**Data type**	***[Semester two] Details: Variables***
Main Questionnaire (Mid)	mixed	
University standard Questionnaire (Semester two)	mixed	16 main questions + free writing (open)
Main Questionnaire (Post)	mixed	38 questions: bio data (1-14), Self-assessment ratings (15-32), course data (33-35), open-ended (36-38)
Final Focus Group	qualitative	20 main questions: transcribed, coded
Final Essay	qualitative	Single topic essay

7. Evaluation of the coursebooks based on student questionnaires and Lectical Analysis

This chapter and the following chapter (Chapters 7~8) offer a variety of evaluations of the two coursebooks (*OL1* and *OL2*). Whereas my Chapter 8 evaluations are based primarily on *qualitative* data, this chapter's evaluation is based on *quantitative* data. In this chapter, I will first review the quantitative data received from students via a standard university questionnaire (replies to multiple choice questions on bubble sheets), regarding the courses. Second, I will review the quantitative data received from my online questionnaire (*SurveyMonkey*), regarding the courses. Third, I will reiterate the rationale for using Lectical Analysis—then I will discuss the significant results from that portion of the study. I have chosen to use multiple data sources, over a range of modalities, to help better piece together the puzzle of complex reasons behind the performance gap that occurred between the two coursebooks, and to learn from them.

My primary studies were conducted on first-year students in the 2014-2015 period. To add credibility to the data and to my deductions, I have also: (1) conducted a full replication study of my primary study (the replication was conducted the year immediately after my primary study [the 2015-2016 period]), and (2) conducted secondary studies collected over a four-year span (2014 to 2018), which also focused on first-year students. The intent of these secondary studies was to help demonstrate that the data collected in the primary study (2014-2015) and the replication study (2015-2016) was not atypical for these university level EFL students; the secondary studies are meant to add credibility to my deductions based on the data from the primary study and from the replication study.

Interestingly, not only did these secondary studies help fortify my hypotheses from the primary study, but they also helped guide the formulation of additional hypotheses. These additional hypotheses will be discussed at the end of this chapter and will be carried over to Chapter 8 for further exploration and validation from a

qualitative perspective. In this way, Chapter 7 and Chapter 8 have a symbiotic relationship in explaining the differences between the two coursebooks. They also firmly ground and justify the rationale for the pedagogical tenets that I will present in Chapter 9.

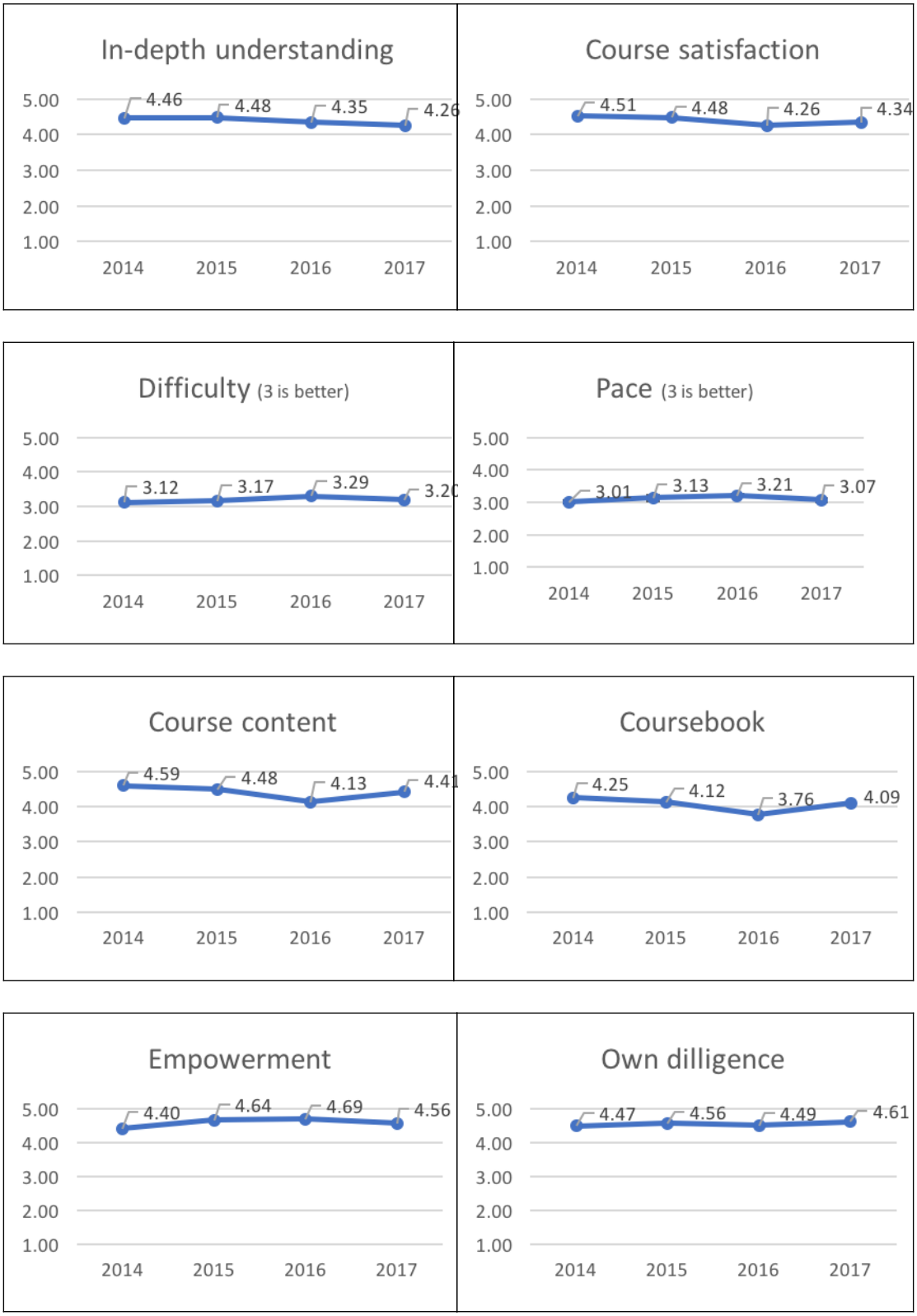
7.1 University student questionnaire

The primary studies were conducted in a single year at a university that administers a biannual questionnaire to their students. However, as explained in the introduction, I have a wide and varied range of data that was collected over a seven-year period. To maintain a coherent logical flow in this chapter, I will first illustrate the "bigger picture" then move on to the smaller noteworthy details. So, here I shall first present comparative data from the longitudinal four-year portion of my project. I collected and reviewed data from the year-end results of each year from 2014 to 2017. I focused on data attained from questions regarding the following themes:

- In-depth understanding
- Course satisfaction
- Difficulty
- Pace
- Course content
- Coursebook
- Empowerment
- Own diligence
- Study time outside of class

Trends across the 4-year study show strong support for two hypotheses put forth in the Lectical Analysis section (end of this chapter). Figure 7.1 is a *nine-part* data set comprising the entire 4-year longitudinal study of first-year students taking the *OL2*-based course. The nine-part data set is presented in this way to illustrate how flat (unwavering) the mean results were per year, per evaluation question. A full

listing of university questionnaire content is provided in the Appendix.



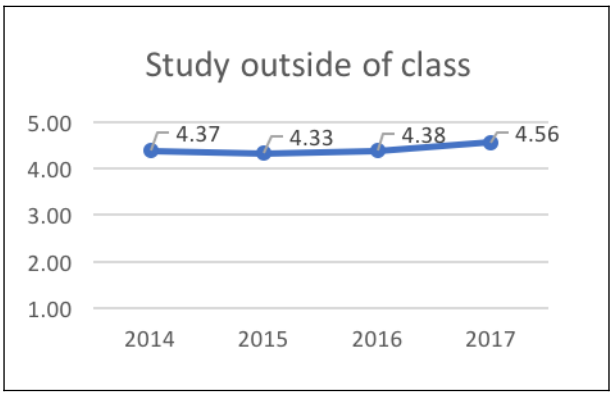


Figure 57: The nine-part dataset from the four-year longitudinal study of first-year students' attitudes toward the OL2-based course.

The nine-part data set comprising Figure 57 clarifies two points: (1) the means across all four years tend to be rather unwavering, and (2) more importantly, although the means are relatively flat across the years per question, those means differ widely *per question*—signifying that the students were not simplistically writing numbers down arbitrarily without reading the questions. This is an important first step in establishing the validity of the data. However, while this data provides an interesting backdrop for this study, the next sections build on this foundation and provide more compelling data.

7.2 Comparing *OL2* to world-class coursebooks

It is hard to imagine that there will ever be a perfect method of coursebook evaluation; there are too many variables across coursebook designs, goals, and contexts for perfect comparison protocols to be established. Likewise, it is not possible to perfectly compare *OL2* directly with other coursebooks because of the complex set of variables involved. Some of the most prominent variables would likely be regarding the relationship between the teacher and the coursebook within the given context (the classroom). However, rather than only comparing *OL2* to *OL1*, my thesis supervisor and I decided that it might be useful to compare the student-perceived efficacy of *OL2* with other coursebooks being used at the same university. So, to accomplish this (and to help level the playing field), I created a composite of the means of the questionnaire responses received by all of the other native English

speaking English teachers in the department that my primary studies took place (N=5). To help with validity issues, I used data from the same time-period (2014-2015). The coursebooks used by the native English speaking teachers were:

- ❑ Smart Choice 1 & 2 (Oxford University Press)
- ❑ InFocus 2 (Cambridge University Press)
- ❑ Conversations in Class, 2nd edition (Alma Publishing)
- ❑ First Class Service 1 & 2, Asia Edition **(**Cengage Learning Asia)

These coursebooks are relatively well-known and well-established coursebooks within Japanese ELT circles from that time-period—moreover, the first two coursebooks have global recognition. Table 31 is a side-by-side comparison of the mean results of *OL2*-based classes data and the "Department" mean. (The Department mean is a composite mean of the data from the classes taught by native English speaking teachers in the department where the primary studies were conducted, using the above-listed coursebooks.)

Table 31: Comparing OL2 class data with a composite of all NESTs in the department

Explanation	***OL2* Mean**	**"Department" Mean**
In-depth understanding gained	***4.46**	4.34
Course satisfaction	***4.51**	4.38
Difficulty of course (3 =good; more than 3 = difficult; less than 3 = easy)	***3.12**	3.49
Pace of course (3 =good; more than 3 = difficult; less than 3 = easy)	***3.01**	3.43
Course content	***4.59**	4.23
Coursebook/media	4.25	***4.42**
Empowerment	4.40	***4.50**
Own diligence	***4.47**	4.25
Study outside of class	***4.37**	4.04

As can be seen in Table 31, OL2-based classes scored better in 7 of the 9 themes; the asterisks symbol (*) denotes the best score in each row. One of the major design goals of *OL2* is optimized "L2 Maturation" or the production of more in-depth thinking in the L2 within the provided timeframe (raising the quality and efficiency of demonstrable cognitive growth in the L2). The results in the top row → "In-depth understanding gained" correlate positively with this goal. Interestingly, "Course satisfaction" and "Course content" also have substantially higher marks than the department mean. It is noteworthy that the *OL2*-based course, with its nonconforming pedagogical cycles of study, is being rated this highly by the students – even in comparison with world-class coursebooks such as those from Oxford University Press and Cambridge University Press.

However, despite the higher "course satisfaction" rating, and the better "course difficulty" rating, the question theme that specifically focuses on "coursebook" was slightly worse than the department mean. Given that the comparison was with some of the world's leading coursebook titles, and because 7 out of 9 themes scored higher than the department means, perhaps I can feel reassured that *OL2* was able to stand its ground even on the "big wig" playing field. Another reassurance is that this data is entirely congruent with focus group data that shall be discussed in Chapter 8, which explicitly provides points of revision for the upcoming rendition of the coursebook (discussed in Chapter 9). This upcoming version shall be referred to as *OL3* in this dissertation.

To add another perspective to this study, I have compiled each English coursebook rating in the department between the years 2014 and 2017 into means. Those means were composited into a "Department Composite" mean per year. That composite data is represented by the orange line in Figure 58. The blue line represents the *OL2* means for the four-year range. The standard deviation for *OL2* for this 4-year rage was 0.20, 0.21, 0.29, and 0.38, respectively. The standard deviation for the "Department Composite" was 0.23, 0.29, 0.29, and 0.21, respectively.

It is difficult to judge how significant this yearly difference is in real-life

terms, but it does imply that there certainly is room for improvement in the design of the *OL2*. It is hypothesized that these Japanese students are at least partially rating their coursebooks by the quality and/or the quantity of the *provided content* in the book—if this is the case, the *OL* series will have difficulty in gaining points in this category because the inherent pedagogical philosophy is for the *OL* coursebook series to be more like a dynamic study-diary that cyclically prompts deeper thinking—it is not meant to be a compendium of reading information to be memorized by students. Perhaps the addition of reading passages regarding neuroELT-based study tips would help raise these ratings while providing valuable study hints to the students. This would be a welcome addition to *OL3*.

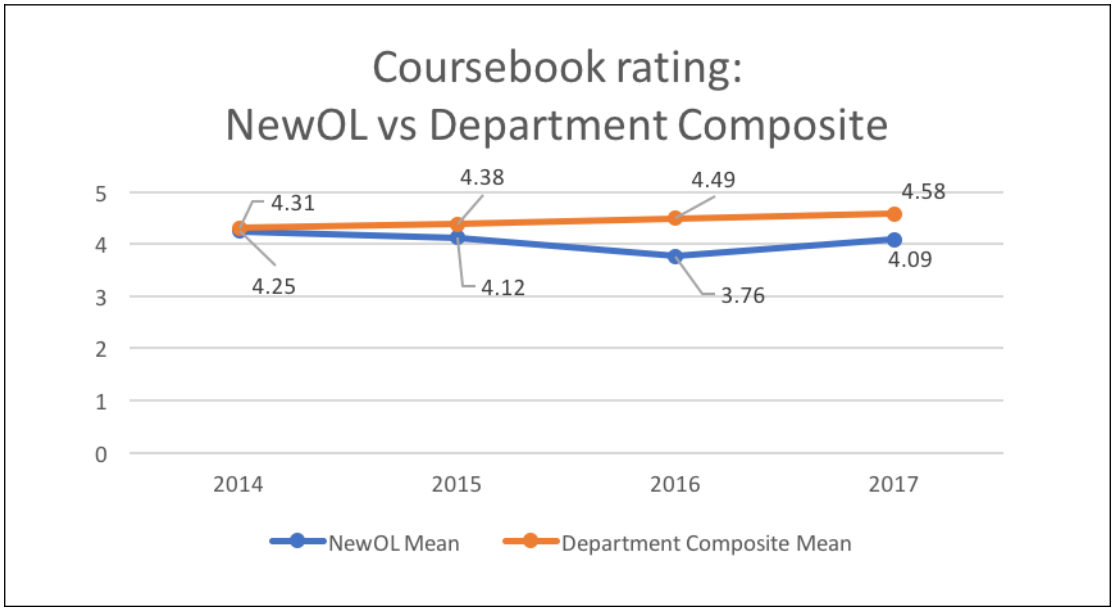


Figure 58: OL2 vs. Department composite, course ratings per year

Two other noteworthy points from Table 31 (p.240) are "Own diligence" (M = 4.47) and "Study outside of class" (M = 4.37). They are both remarkably high, especially in comparison to the department mean—which perhaps reflects the Japanese student experience of being spoon fed. Considering these higher means with the similarly higher means of "Course satisfaction" and "Course content," it seems *OL2* is holding its own not only in the classroom but also outside of the

classroom—signifying that *OL2* seems to be a transformative coursebook for at least some of the students. The implications of this will be discussed in more details in the Lectical Analysis section of this chapter.

I was particularly pleased by these points because another significant objective of my coursebook writing project was to encourage a naturalistic increase of student engagement. The higher evaluations here seem to indicate that the coursebook may have at least partially achieved this goal.

I had hypothesized early on that a high level of naturalistic student engagement could be accomplished by providing a more bottom-up (student passion-following) thematically organized design that is free-flowing (organic), yet at the same time is tightly bound to a solid systematic learning framework. *OL1's* design did have this goal in mind, but, being my maiden attempt in coursebook writing, the intended level of learning efficiency—and natural motivation to do more—was not achieved. The Lectical Analysis data makes these shortcomings of *OL1* very clear.

Further exploration of my theoretical conceptualizations (pedagogical tenets) and their implications will be discussed in Chapter 9, with Chapter 8 providing qualitative data (student interviews) giving critical insights into *how* and *why* everything generally played out better in *OL2*.

7.3 My SurveyMonkey student questionnaire

Although the university's questionnaire data (discussed in 7.2) was useful as supporting data for my primary study, I also created a purpose-designed online survey (using the online *SurveyMonkey* service) and conducted my own survey concurrently with the university questionnaire. This was necessary because the university questionnaire was not explicitly designed to evaluate coursebook efficacy. And, for the purpose of this study, I needed more concrete data regarding coursebook efficacy, with a focus on students' perspectives. Here I will review the primary study results (conducted 2014-2015). I will then discuss the replication study results (conducted 2015-2016).

For the primary study, I focused on 14 of the online questions; I have listed them in Figure 59. The full list of questions is much longer and can be found in the Appendix. The full list of questions is provided to show how I used the questionnaire to create a high-support context; the questions were deliberately arranged to build-up higher levels of engagement—to induce a deeper-thinking state in time to respond to the 14 key questions in greater depth."

The 14 key questions comprise two batteries of questions. The two batteries of questions were: (1) questions related to perceived current English proficiency, and (2) questions related to imagined future use of English (Figure 59).

The two batteries of questions:

1. English proficiency related questions:
   - Current reading skills
   - Current writing skills
   - Current listening skills
   - Current speaking skills
   - Current grammar skills
   - Current vocabulary skills
   - Current spelling skills
   - Current presentation skills
   - Current overall English proficiency
2. *"English in your future*!" related questions:
   - ➢ How likely do you think you will use English with your future family?
   - ➢ How likely do you think you will use English with your future friends?
   - ➢ How likely do you think you will use English with your future work/business?
   - ➢ How likely do you think you will use English on the internet?
   - ➢ How likely do you think you will use English as a "hobby," or out of fun (music, movies, etc.)?

The additional "closure" question (only asked at the end of the year):

- ➢ How much have you improved, compared to traditional English courses' improvement? [Traditional class improvement = 3]

Figure 59: SurveyMonkey questions that were selected for this study (full questionnaire in Appendix)

7.3.1 Results

The *SurveyMonkey*-based online questionnaire was conducted three times during the 2014-2015 study year (Pre-test, End of 1st-semester test, and End of 2nd-semester test), as a part of the primary study. I will provide a full review the data of the "Current proficiency" battery first. I will then provide a full review of the "Imagined future use of English" battery after that. I will also provide a review of the year-end-only "closure" question on *Improvement compared to traditional coursebook improvement.*

7.3.2 Current proficiency

Regarding the "Current proficiency" battery, first, let us review the raw data. For the composite of "Current proficiency", the mean results were: Pre-test (M = 2.32, SD = .67), End of 1st semester test (M = 2.42, SD = .57), and End of 2nd semester test (M = 2.76, SD = .59). The trend here was constant growth. Table 32 is the raw data in tabulated form.

Table 32: Results of the "Current proficiency" battery of self-assessment

Question Matter	2014 Pre-test Means	Standard Deviation	2014 Semester 1 Means	Standard Deviation	2014 Semester 2 Means	Standard Deviation
Reading	2.90	0.84	2.88	0.73	3.14	0.71
Writing	2.40	0.90	2.45	0.86	2.66	0.83
Listening	2.55	0.95	2.44	0.94	2.77	0.96
Speaking	1.90	0.89	1.94	0.85	2.38	0.87
Grammar	2.49	0.93	2.52	0.99	2.69	0.85
Vocabulary	2.51	0.84	2.47	0.81	2.68	0.90
Spelling	2.56	0.92	2.53	0.87	2.84	0.88
Presentation	1.74	0.86	2.20	0.85	2.94	0.92
Overall	1.96	0.84	2.31	0.78	2.64	0.84

Next, let us review SPSS-based inferential statistical data. Regarding the three tests' data (Pre-, End of 1st Sem., and End of 2nd Sem.) on "Current proficiency": for within-subject effects, $F = 30.346$, $p < .001$; for within-subject

contrasts, linear analysis showed F = 39.135, p < .001; quadratic analysis showed F = 9.591, p = .003. The between-subject effects showed F = 1511.969, p < .001. These analyses all returned highly significant results, meaning the overall results were trustworthy and that inferences can be made from these results. I have provided a graphical illustration of how the participants' self-assessment of "Current proficiency" changed over the course of one year of *OL2* usage (Figure 60).

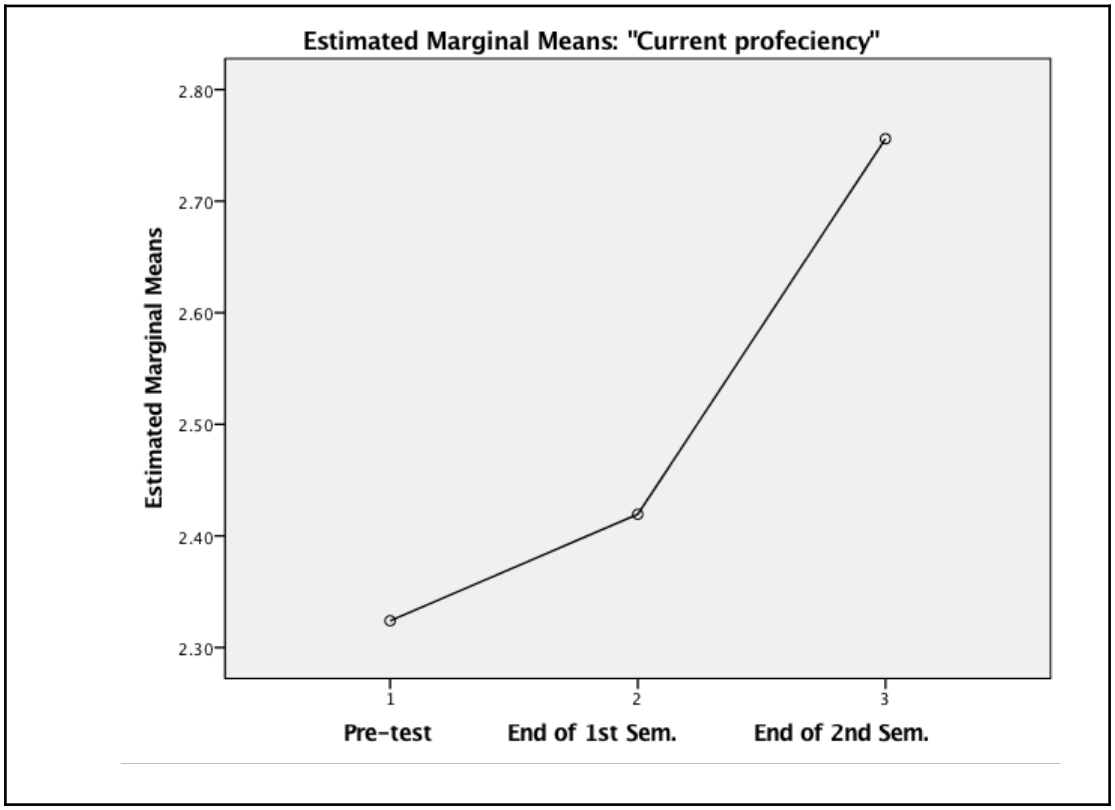


Figure 60: How self-assessment of "Current proficiency" changed across the year.

7.3.3 Discussion of "Current proficiency"

There are two points of interest regarding this portion of the data. One point is that during the course, the participants continuously self-assessed their then-current proficiency level higher than their previous self-assessment. This is particularly interesting because Japanese students are known to be culturally ingrained to be more "reserved" regarding self-assessment, so this upward trend is a powerful indicator of the perceived efficacy of the *OL2* procedures; the students of

OL2 collectively showed this upward trajectory spanning the entire year (as seen in Figure 60).

The second point is reassurance; the statistical analyses via SPSS registered high significance regarding this dataset. We can, therefore, assume that these trends will likely hold true for other users of *OL2*, and in fact, the myriad of other data collected in the primary study and the secondary studies (including longitudinal data, and the qualitative data discussed in Chapter 8), support this supposition.

7.3.4 English usage in my future

For the composite of the "English usage in my future" battery, the mean results were: Pre-test (M = 2.86, SD = .90), End of 1st semester test (M = 2.77, SD = .86), and End of 2nd semester test (M = 3.03, SD = .87). Constant growth was not the trend here; there was a drop at the end of 1st semester, but then a rise at the end of 2nd semester. Next, let us look at this data with inferential statistics. Regarding "English usage in my future" (Pre-, End of 1st Sem., and End of 2nd Sem.), for within-subject effects $F = 4.786$, $p = .010$; for within-subject contrasts, linear analysis showed $F = 3.839$, $p < .054$, and quadratic analysis showed $F = 5.768$, $p = .019$. The between-subject effects showed $F = 983.880$, $p < .001$. The linear analysis showed a positive trend (on the borderline of significance) at .054, while the quadratic analysis showed significant results at .019. Taken collectively, these results are significant enough to report. I have provided a graphical illustration of how the participants' self-assessment of "English usage in my future" dramatically changed over the course of one year of *OL2* usage (Figure 61).

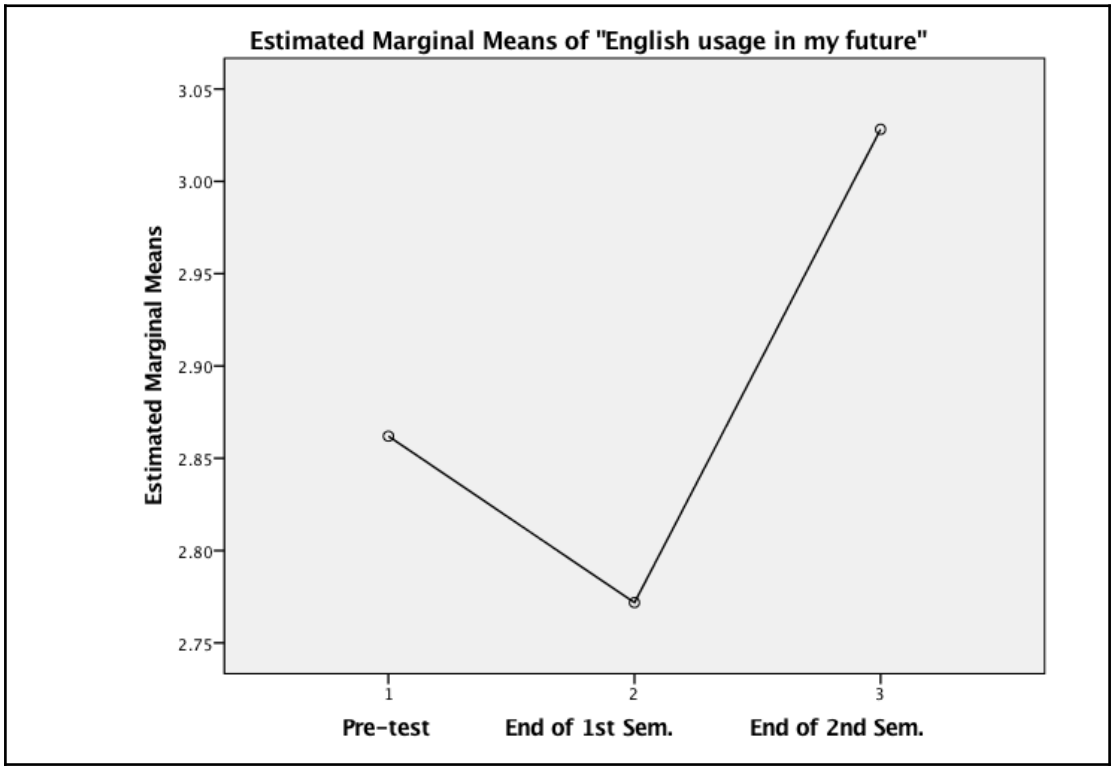


Figure 61: How self-assessment of "English usage in my future" changed across the year.

7.3.5 Discussion of "English usage in my future"

There are four points of interest regarding the participants' imagined "English in my future" data.

Point one: At the beginning of the course, the participants generally self-assessed at a mark under "3". It was clearly explained to the students in Japanese that "3 = a typical Japanese person's likelihood of using English in their future." Therefore, a mean below "3" signifies that collectively, they were not very confident about English improvement and/or could not imagine themselves using English at what they perceived to be a normal level of competence for Japanese people in their own futures. Additionally (as discussed above), because they were Japanese, they may have been reporting their self-assessments in their customarily "reserved" manner. Whatever the reasons may be for this mean result, this pre-test mean score serves as the benchmark for this portion of the study.

Point two: Using that first mean score as the benchmark, we can see that at the end of the first semester, there was a clear and substantial drop in their reported confidence regarding future English usage. Perhaps the veracity of using English for real communicative purposes instead of relying on rote-memorization (as they had done so often in the past), may have kicked in (Figure 61).

Point three: However, by the end of the year, there was a dramatic rise in reported confidence levels. Their final mean for this question shot past the benchmark and even rose above "3"; this is a meaningful gain not only because of how sharp a rise it is but also because it puts their means above what they perceived to be the societal norm in Japan.

Due to the influence of Confucianism (primarily *filial piety*), it is almost taboo for students to suggest that they are on par *or above* adults, yet the notion of superiority over the perceived norm was established here; they collectively put their self-assessment above what they perceived to be the norm in Japan. This is a significant (and persuasive point) that is being made by the students—I believe it is safe to say that this dramatic change in student confidence is at least in part due to the design and implementation of *OL2* (over a prolonged period), while overcoming the inherent issues of being a first year university student.

What clues can we uncover here regarding this affect? It seems the efficacy of the embedded cycles of *OL2* may have provided the students with some L2 automaticity that they could now "feel"—giving them confidence in spontaneous English usage. I believe this "new" confidence regarding spontaneous usage may have been inspiring and is what may have helped push the mean over the (taboo) "3" threshold. Observing this equation from the opposite side, this may signify that the participants realized that the skill of "spontaneous L2 usage" is lacking in the average Japanese adult—and, they realized that they were now "above" this. If that is the case, I would like to agree with their collective assessment.

Furthermore, it seems that the participants not only gained confidence but also began to display signs of deeper metacognition regarding their learning and their performance levels. The qualitative data in Chapter 8 explains these phenomena in

greater detail; in Chapter 8 the participants plainly discuss how, by the end of the year, they gained confidence in their English skills in multiple ways. They also went on to explain in their own words their own understanding of the (biological) mechanics of how this new pedagogy must have worked for them.

Point four: The data suggests that prolonged exposure to the cycles within *OL2* raises student confidence and/or raises students' ability to envision themselves using English in several different contexts in the future (see Section 7.1). This dramatic rise—from the lowest point to highest point, across the course of the second term—is particularly noteworthy (again) in the Japanese cultural context because Japanese students are traditionally trained to be "reserved" during self-analysis. Moreover, it required a full year of exposure to *OL2's* cycles for the benefits of the design to manifest in this way, suggesting that longer-term usage of *OL2* pedagogy is likely to be more beneficial as an agent of change--compared to single-semester usage.

There are likely two reasons for the time required for these students to gain confidence when studying with *OL2* (and produce such a dramatic rise in confidence when it does appear):

(1) In the first term, the students were not used to the unique pedagogical flow of events, so it took time for them to become accustomed to the cyclical and deliberately convoluted intricacies of *OL2's* tasks; this can be seen as a break-in period. But by the second term, became evident to many of the students that there were no more pedagogical surprises waiting for them, and so they became relaxed—while at the same time this "developed knowledge" of what-is-to-come-next surely helped them concurrently develop their own *pre-emptive strategies* regarding the now-fully-foreseeable flow of events. Surely the participants began developing these pre-emptive strategies specifically because of the regularity of the pedagogical flow in *OL2*—but this does not mean that the majority of the students were necessarily conscious of their developing pre-emptive strategies. Much of their pre-emptive strategy building may have been

going on at an intuitive level, based on the automaticity-building effect of regularity, which went hand-in-hand with the nurturing of their confidence levels—regardless of how explicitly conscious they may have been regarding any of these points. So, although their changes in confidence levels and the catalysts (such as via solo assessment and peer-to-peer assessment) may or may not have been apparent to them at a conscious level.

(2) Being designed with the concept of *L2 Maturation* and pedagogical efficiency in mind – leading to "deeper thinking in the L2" as main design goals (the upcoming Lectical Analysis data will show that these design goals have been met), the students needed ample time to create effective strategies for these tasks that demanded substantial cognitive growth within them. However, once reliable strategies for these cognitive demands were in place, the students were able to use them repeatedly, on demand, and with high efficiency→ specifically with, and because of, the regularity of the cyclical *OL2* pedagogy.

It would seem that this collection of observations help to at least partially explain why *OL2* outperformed *OL1* when the efficiency (rate) of cognitive growth as demonstrable in the L2 was compared between the two groups; the *OL2* students gained confidence via the development of pre-emptive tactics while they were developing efficient and reliable strategies to keep up with *OL2's* cognitive demands in their L2. Because of the regular cyclical design that was unwavering across the course of the year, these strategies became standardized and over time naturally created substantial automaticity in the brains of the *OL2* students in regards to their own L2 skill building and their own L2 usage. The regularity of the challenging cycles gradually allowed them to became ready for spontaneous and meaningful discussions in the L2 when called for.

On the other hand, while the initial design goals were similar, in retrospect, *OL1's* pedagogical flow was not cyclical enough and was generally more disjointed. Therefore, within the provided timeframe, it was not able to nurture enough

linguistic automaticity and linguistic confidence to allow for spontaneous and meaningful L2 usage on a level that paralleled *OL2* users.

7.3.6 The replication study (2015-2016)

To add an extra level of rigor to the project, I replicated the entire portion of the primary study described above in the following school year (2015-2016). The replication study provided interesting results that added depth to this thesis and helped nurture hypotheses regarding the efficacy of the pedagogical designs of *OL2*.

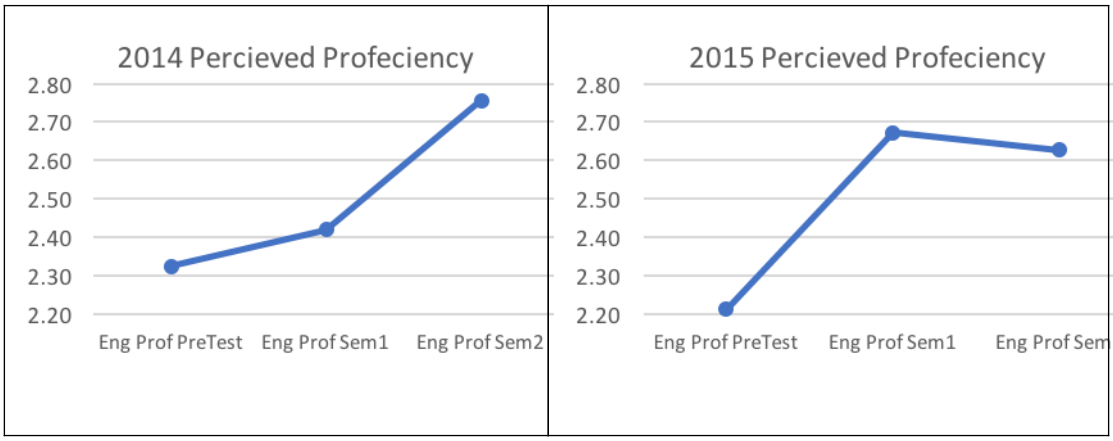


Figure 62: Students' perceived proficiency mean growth, primary study vs. replication study.

Regarding *Perceived Proficiency*, the primary study group and the replication study group tell different stories (Figure 62). The noteworthy points are, (1) the incoming freshmen in both years start off their year with "reserved" self-assessments—as expected. (2) By the end of the 1st semester, both groups show a confidence boost. The replication study group showed a much more pronounced boost, jumping up 3~4 times higher than the primary study group's boost in the 1st semester. (3) By the end of the year, the primary group also had experienced a significant boost; the trajectory is similar to the replication group's boost in their 2nd semester. It can be said that by the end of the year, both groups showed similar concluding levels in confidence, but it took *longer* for the primary group to achieve this confidence. (4) When viewed in the graph, the replication group clearly shows a slightly decreased mean by the end of their school year. However, the graph's visual story may be a bit deceiving. Considering that the 1st semester's *SD* = .90 and the 2nd

semester's *SD* = .86—which are wide deviations on a mean that moved from 2.67 to 2.63--it may be rational to interpret this as a *plateauing* mean response rather than a noteworthy drop.

I have a reasonable hypothesis for the radical difference in trajectories between the primary study group and the replication study group. Overall, for the primary group, *SD* = 0.67, 0.67, and 0.59, respectively (*M* = 0.61). For the replication group, *SD* = 0.90, 0.86, and 0.87, respectively (*M* = 0.73). The replication group's great rise in the 1st semester (compared to the primary group's trajectory) should probably be attributed to my personal growth "as a teacher accustomed to using *OL2"*—in short, I had better rapport with the students, and more confidence with the teaching method. By the end of the 2014-2015 school year (the end of my primary studies), I had a full year of focus group input, data reviews, and in-depth personal realizations about the strengths and weaknesses of *OL2*. Because "Personal growth as a teacher" was also a desired outcome of this dissertation project, I cannot deny that the multi-modal data (resulting from researching *OL2* during the 2014-2015 school year) did not provide significant insight into pedagogical strategies that could have positively affected my teaching techniques and classroom demeanor.

In retrospect, I found it fascinating that, much like the students in my study, I seemed to have developed automaticity regarding pre-emptive strategies for *OL2* usage (but as their teacher). The confidence that I gained from understanding the "ins-and-outs" of *OL2* likely allowed me to enter the classroom in the 2015-2016 year with higher expectations for myself and higher expectations for my students. Moreover, it is highly likely that my explicit knowledge of the trajectory in the primary study affected my teaching in a way that pushed student confidence levels higher than the previous year—this may have significantly affected students' perceptions and their reporting.

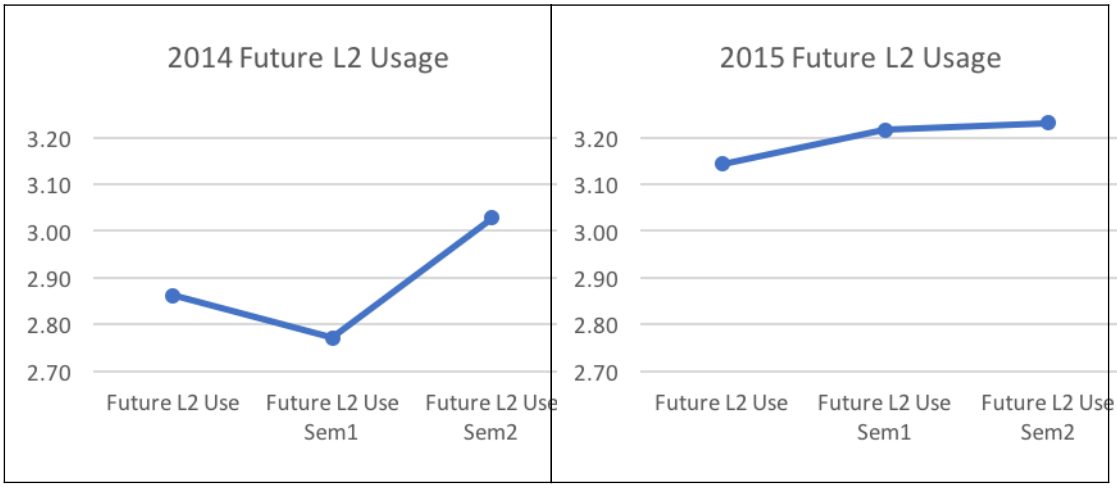


Figure 63: Students' confidence in future English usage, primary study vs. replication study.

Compared to the *Perceived Proficiency* data trajectories, the *Future L2 Usage* trajectories were vastly different for these two years (Figure 63). The trajectory for the 2014 *OL2* group was as follows: (1) the students joined the class at the beginning of the year and set a starting mean mark (M = 2.86), (2) the students lost a bit of confidence by the end of the 1st semester, but then (3) showed a sharp rise by the end of the year. The trajectory for the 2015 *OL2* group was as follows: (1) the students starting mark was at a much higher point than the 2014 group's (M = 3.14), (2) by the end of the 1st semester, unlike the 2014's data, it continued to rise, and (3) in the 2nd semester, it rose further; although considering that SD = 0.95, it can be viewed as a plateauing rather than a significant rise at the end. Overall, the 2014 group's SD = 0.90, 0.86, and 0.87, respectively. The 2015 group's SD = 0.96, 0.87, and 0.95, respectively. Logically, the explanation for this difference is identical to the explanation provided above. My personal confidence levels and expectations for the students in the second year very likely had a profound effect on the students' confidence levels.

In sum, the general trends for 2014 were: a slow start, but a sharp upward turn by the end of the year. The general trends for 2015 were: quick to rise, then a plateau. In addition to these intra-year studies, the university questionnaire data (the 4-year longitudinal study that I conducted with first-year students using *OL2* that focused only on year-end reports), shows that there is a clear year-to-year plateau for

every question that was studied, even though these were a different group of students each year. The above 2-year intra-year growth trajectory study regarding *Perceived Proficiency* and *Future L2 Use* adds meaningful depth to the overall project by showing how—within the timeframe of a single year—students' perceptions do radically (and dynamically) change. And, it show calls to mind how student perceptions are potentially affected by the teacher's own developing understanding of the possible outcomes of the pedagogical design.

Thus, it seemed essential to add a study on a potential of a "Halo effect". In the final year of my action research I added a study of how students' *course evaluations* and their *motivation to do better in class* may have been affected by their rapport with the teacher. Those results will be explored in detail in Section 8.11.

7.3.7 The year-end final question

Next, I would like to focus on the "closure" question that I added to the 2014 questionnaire and the 2015 questionnaire—only at the end of each respective year. That year-end question was, "How much have you improved, compared to traditional English courses' improvement?" By the end of the year, the students had been undergoing a variety of 1-5 assessment tasks each month (inclusive of peer-to-peer assessment and self-assessment). It should be noted that throughout the year, it had been thoroughly established that "3 = norm". And for this question, I took the time to explicitly explain that "3 = the norm for traditional English class improvement".

At the end of the primary study (2014-2015), the results to that closure question were as follows: M = 3.70, SD 0.75. For the replication study (2015-2016), the results were are follows: M = 3.59, SD 0.65. Both years' participants reported that the *OL2*-based class produced superior results compared to their perception of traditional English classes. This is a recurring theme within this chapter. The next chapter will use qualitative data to shed light on *how* the students perceived these differences. The next section shall look at student L2 growth from a cognitive developmental perspective (using Lectical Analysis to uncover L2 maturation rates within the participants).

7.4 Lectical Analysis results

For this study, I used Lectical Analysis in a way that was as unconventional to Lectica as it was for the field of Applied Linguistics. The results were used to study student growth as well as to compare coursebook efficacy—coursebook evaluation has never been a very popular research area within Applied Linguistics, so it seemed like another worthwhile avenue to explore for this doctoral thesis.

Standardized tests—even tests that are purpose-designed for specific coursebooks by the publisher—are prone to the issue colloquially known as "teaching to the test." This is not only due to the book publishers' designs. Teachers are known to *teach to the test* for pragmatic reasons; as a consequence, it could also be the students who are cramming their coursebook content just before the test. It would be hypocritical to blame students for being duly pragmatic.

Tests such as these (ones that can be taught to, crammed for, and then forgotten) surely have little validity when assessing students' actual proficiency and growth. When coursebooks are assessed with quantitative data, a common method of assessment is by comparing student outcomes with established test baselines—and sometimes rightly so. But by the same logic as above, students' scores on "crammed tests" and/or students' scores on tests that were spoon-fed by the teachers are unlikely to be valid for coursebook evaluation. To help alleviate this weighty issue, I have taken an approach to evaluation that cannot easily be "taught to," nor can it be directly studied for by the student beforehand. With this form of assessment, I am looking at the cognitive growth as demonstrable in students' L2s. This growth data from the "demonstrable cognitive growth in the L2" can then be compared to the means per coursebook (in this case, *OL1* vs. *OL2*) to show the differences in their efficacy—this efficacy test is therefore based on the demonstrated growth in complexity of L2 language usage usage per student, per coursebook.

Lectica's tests for cognitive growth measurement are typically administered in the native language of the participant. Their long-term goal has been to map-out the most accurate growth patterns in humans, and their massive datasets along with their assessments strategies have proven high validity. Naturally, the most reliable

data can only be derived from the collection and assessment of responses made in a participant's native language. Yet, I have turned this around and made the domain of L2 the testing field, and used ANCOVA statistics to pursue further validity. By comparing Lectica's typical growth rates of university students (their L1-based testing) with the L2-based growth rates that I uncovered with my own study per coursebook, I was able to compare L2 maturation rates per coursebook. The details of this method are below.

7.4.1 Understanding the data

On the Lectical Analysis scale, scores in the 10~11 range are normal for tests in the undergraduate university age range in the L1. Growth speeds vary on the gradation between children and adults. According to Lectica, the typical growth of a single year of study for typical native English speakers in their undergraduate years (in the L1) is a speed of 0.05 ~ 0.08/year on their Lectical scale. To simplify things, let us purely use the upper end of the range as a benchmark (0.08 growth per year). With an established benchmark of 0.08/year for undergraduate student growth in the L1, the results of *OL1* vs. *OL2* growth rates across pre- and post-test scores were compared —but to reiterate, my testing was done in the participants' L2, not in their L1. Therefore, rather than using t-tests and ANOVA, I used ANCOVA in SPSS— comparing the post-test results of the two groups with the pre-test results filtered out as the covariate for the ANCOVA calculations. By adjusting the post-test scores for any pre-test differences, I was able to statistically screen out the unwanted effects of the potentially different starting points for the two classes. In this way, the playing field has been leveled, and comparisons between the two groups (*OL1* users and *OL2* users) became valid.

First, let us view the raw numbers, then move on to the ANCOVA results. In one semester, the *OL1* class (N=24) produced a 0.01 growth mean (the difference between the pre-test vs. post-test scores means) in the Lectical scale after 15 weeks of study with *OL1*. This is a growth speed of 0.04/year (compare to benchmark L1 growth speed of 0.08/year [1/2 of benchmark growth]). The *OL2* class (N=30)

produced a 0.08 growth in the Lectical scale after 15 weeks of study with *OL2*. This comes to a growth speed of 0.32/year (compare to benchmark L1 growth speed of 0.08/year [4x benchmark growth]). The ratio between the growth speed of the means of these two data sets is therefore 1:8 (*OL1*: *OL2*). Figure 64 graphically illustrates the difference in growth speeds between *OL1* and *OL2* with the pre- and post-testing outcomes (compare orange line vs. blue line growth, per book).

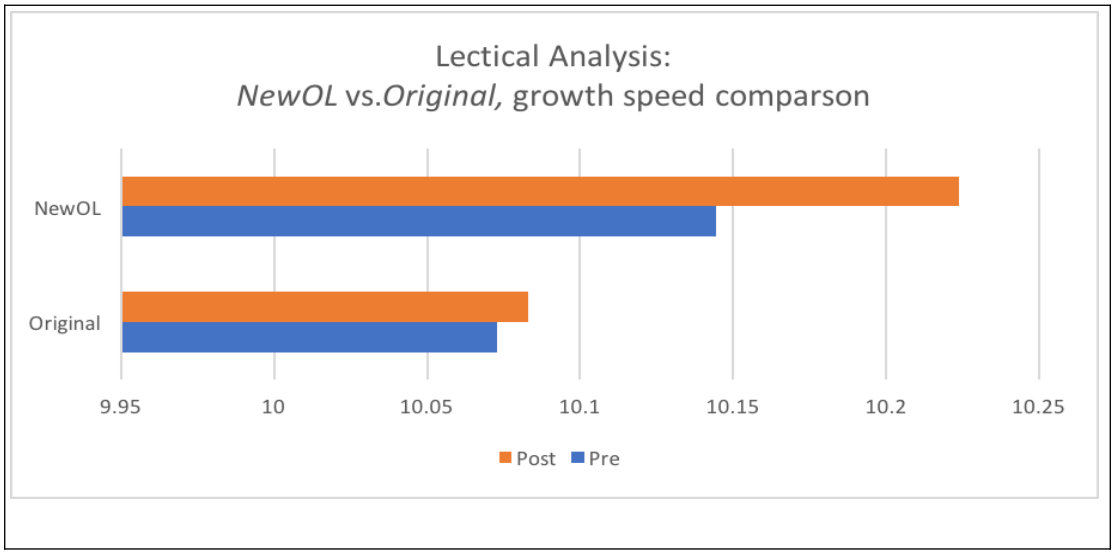


Figure 64: The OL2 (NewOL) group showed a roughly 8x faster rate of demonstrated growth over the OL1 (Original) group, on the Lectical scale.

In Figure 64, it is evident that the mean starting point of the two groups (pre-test mean) differs, requiring the ANCOVA analysis strategy to make it a fair comparison. So, let us now view the ANCOVA-based results. With this analysis strategy, there was a significant difference between the outcomes of *OL1* (M= 10.095, SD = .041), and *OL2* (M = 10.214, SD = .36); F = 4.702, p = .035. The effect size was medium (Partial Eta Squared = .084). In ANCOVA, the mean for the *OL1* class was 10.095 with a standard error of .041. For the *OL2* class, the mean was 10.214, with a standard error of .036. Thus, both the descriptive statistics and the inferential statistics produced what we can deem as significant results; these results show that students using *OL2* benefitted significantly more from the teaching than the students being taught with *OL1*. Not only did the raw data show this, but even

with the employment of inferential statistics using ANCOVA, the *OL2* group demonstrated a significantly faster growth rate in their L2 than the *OL1* group did.

7.4.2 Presentation marks vs. cognitive growth: The need to establish *L2 Maturity*

In this section, I will add to the mix my own (teacher assessed) evaluations that I gave per group presentation during the primary study (2014), per coursebook. I will display the trends that were uncovered from that data. Then, I will discuss how classroom marks do not always correlate with actual growth, why this may occur, and the practical implications.

First, let us view a direct comparison between *OL1* and *OL2*'s mean scores per group presentation across the first semester of the 2014 school year (the primary study year). Figure 65 is a line graph that shows the trends across the three presentations that occurred during that 15 week period. The blue line represents the *OL1* student group work presentation mark's trend (using the mean of all groups that presented). The orange line does the same, but with the mean of all the *OL2* presentations held during the primary study.

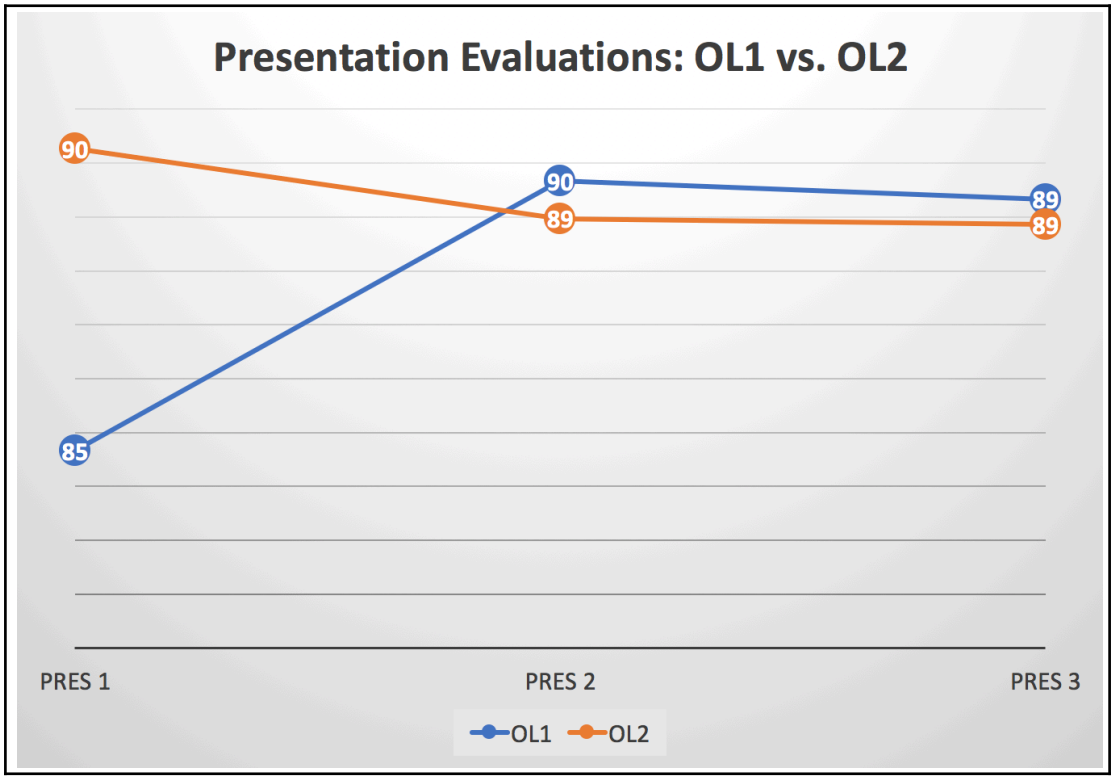


Figure 65: Presentation evaluation trends OL1 vs. OL2

There are several points to be made regarding the data in Figure 65. First, I told the students in both groups that I would be assessing them dynamically. That is to say that I warned them that with each presentation, I would be evaluating them more harshly in terms of content, delivery, and logical flow of their presentations. Second, I actually did "raise the bar", so to speak, for each presentation cycle. I expected them to do better, and adjusted my marking accordingly. Third, no matter how professional/unprofessional an evaluator may be, the scoring will inevitably be subjective at some level. So, the data presented here is more subjective than the other data presented in this thesis. Nevertheless, I have compiled the means of the presentation scores, and there are interesting deductions to be made.

So, with that caveat in place, what can be deduced from this data? *OL2* shows a flat line (90-89-89) while I was intentionally marking the students more severely each time. So, although I fully concede that evaluations such as these are subjective and personal (as discussed in Section 3.2), I am certain that I deliberately became tougher on them as the semester progressed. It was therefore a surprise that *OL2*

produced such a flat line across the semester. The fact that there is a flat line is unimportant, and is likely happenstance—but what *is* important is that the mean scores did not go down as I became progressively stricter with my marking. So, in the very least, we can deduce that the *OL2* students were improving in their presentation skills—as they themselves reported noticing (even though the marks produce a flat line). As for *OL1* we can see a sharp rise from the first presentation means (85) to the second and third presentation means (90 and 89, respectively). Everything that was said about the *OL2* data can be said about the *OL1* data, including the eventual flat line. But, there is one more significant issue—the first presentation from *OL1* was of much lower quality (the mean scores were much lower). Why might that be? I hypothesize that *OL1* did not prepare the students as well as *OL2* did for designing their presentations. But, by the second round, they were able to grasp what they needed to do, and greatly improved their presentations despite the fact that *OL1* did not do as good a job in preparing them for their presentations.

In retrospect, it is logical that this earlier edition of *OL* did not prepare the students as well as the second edition did. This implies that *OL2* was indeed an improved design over *OL1* in that regard. The provided data in Chapter 7 and Chapter 8 are in support of this hypothesis.

A second issue is of a very different nature, and is on a grander scale. Viewing final course marks alone, relatively speaking, the *OL1* and *OL2* students did just as well, per course. It was only via growth in complexity scores (Lectical Analysis) that we were able to to see a significant difference between the two groups' demonstrated cognitive growth. Therefore, it becomes obvious that "class grades" *per se*, can be terribly misleading, and downright erroneous.

What are the implications? On one hand, since both groups' students did what they were asked to do, and received good enough grades, we can say they both did well in their respective courses. On the other hand, the startling reality is that the *OL1* students did not grow as rapidly nor as effectively as the *OL2* students did.

OL1 data showed a small improvement in growth that is not "bad" per se, but

what can be deduced is that if the pedagogy is done *just right*, in 15 weeks, a course can bring on a *cognitive reorganization* at the university student level. If we can assume that a cognitive reorganization of this magnitude is beneficial for university EFL students, then the question is—*can we afford to not realize this difference in other courses?*

In short, perhaps it is time to rethink standard language evaluation. In most cases, the prime focus should be moved from "memorization" to "maturation". And, in my case, the extraordinary point is that, it was my full intent to make this switch when designing *OL1*, and yet, the Lectical Analysis data clarified that the intent to switch over to a more cerebral/deep thinking pedagogy alone is not sufficient – everything about it had to be laid out *just right*, and then implemented *just right*. *OL2*'s data contrasted with *OL1* shows this to be true. I will spend much of the remainder of this thesis unpacking data and making deductions that support this stirring discovery.

7.4.3 Discussion

It may seem that I have an infatuation with the phrase "cognitive growth as demonstrable in their L2s" –what do I mean by this, and why is this phrasing so important? As discussed earlier, I found the need to create a form of assessment/ evaluation format that could not be *taught to*, nor be *crammed for*. The assessment solution needed to compare the cognitive complexity of output from pre- and post-tests of each student, in their L2. Lectical Analysis-based tests cannot be directly studied for, nor can they be crammed for because of the way the test is designed—the participants are given a question of philosophical nature (a morality question), and then they are prompted to write an "elaborate-as-possible" justification for their current position on the proposed ethical conundrum. The complexity of each students' response (in this case, a typed response) is what is analyzed, coded, and ultimately given a score on the Lectical scale.

The *OL1* class showed a mean growth of 0.01/semester (or roughly 0.04/ year) while the *OL2* class showed a mean growth of 0.08/semester (or roughly 0.32/

year). According to Lectica, typical native growth speed per year is in the 0.05 to 0.08 range. To simplify, let us again just focus on the higher end of the range (0.08) as a good native speed benchmark. That would tell us that *OL1*'s mean growth rate was roughly 1/2 of that good native rate (0.04/year vs. 0.08/year), while *OL2*'s mean was roughly 4x the good native rate (0.32/year vs. 0.08/year).

How is it possible for *OL2* (used in a Japanese EFL context) to achieve a 4x growth rate of same age group students tested in their native language in the USA? That is a good question; these results can easily be misconstrued. Firstly, this is not an error in calculation. As Dawson puts it, "...(the) process is allowing students to express their real level of thinking more rapidly in English. The rapid rate is likely to be due to the fact that they already have higher level capabilities that they can now tap more effectively" (Dawson, 2014).

I have two hypotheses for this seemingly extraordinary growth speed of the *OL2* group: (1) as pointed out by Dawson above, considering that this was conducted in an EFL context, it is highly unlikely that the pre-test scores of the *OL2* students on the Lectical scale were equivalent to what they *would have scored* on the test had they taken it in their native language. Based on Lectica's data on American university students, and the scores that these EFL students provided, on their pre-test, many of these EFL students would likely have scored almost a full point higher on the Lectical scale had they been tested in their native Japanese. See Figure 66 for a graphic representation of this issue.

We can confidently presume that although the EFL students' brains had the potential for higher level thinking *natively*, they seldom had been pushed to, nor trained to, do so in English. And so, there is a real gap between their actual cognitive prowess, and what they are able to demonstrate in English (as seen in Figure 66). Pedagogically, what needs to be done is to "close the gap" between what they can do in Japanese and what they can do in English.

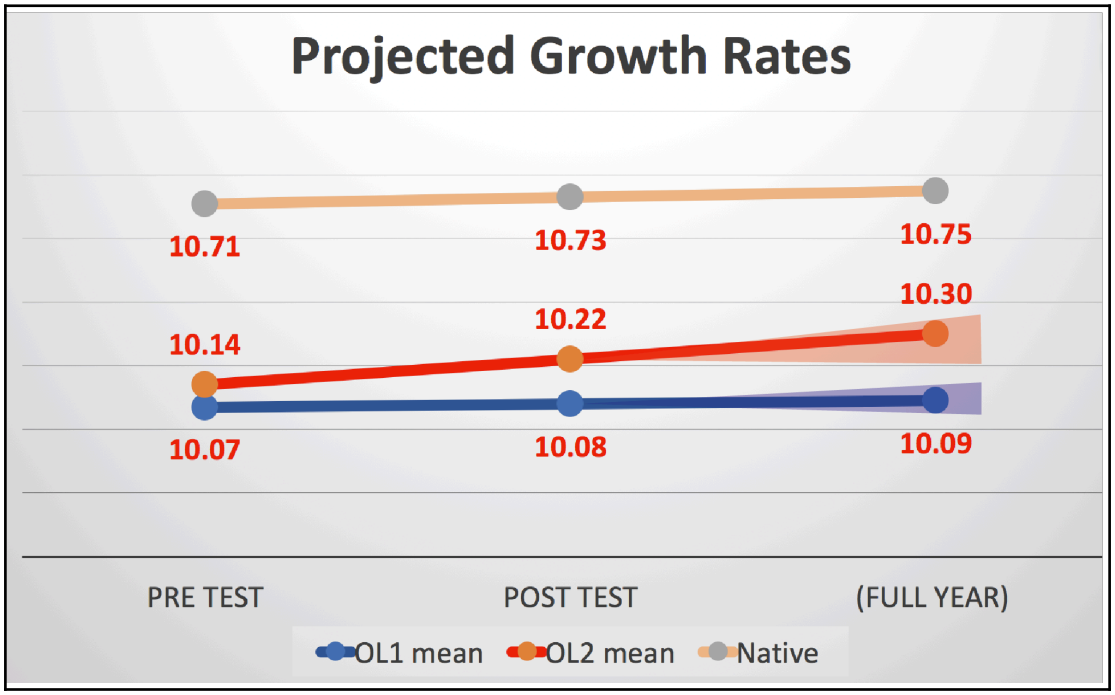


Figure 66: Projected growth rates for two semesters (full year), based on first semester growth

In Figure 66, I have used the real means from one semester of *OL1* and *OL2* (PRE TEST to POST TEST in the figure), projected a straight line of growth for a full year simulation, and then added a reasonable range of error to the bars. I did the same for the *Native* line with an "average Joe" university student (sans a range of error because this is a hypothetical student, but based on real data).

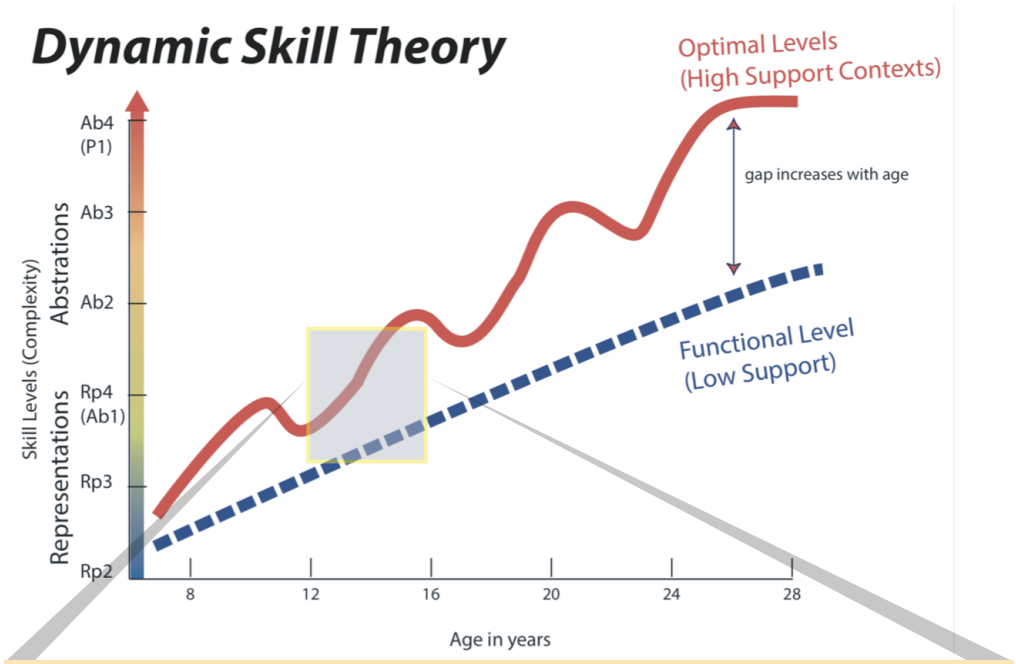
The graph depicts how *OL2* is closing the gap faster than *OL1*, while at the same time, expressing that although the "4x native rate" may sound extreme when viewed in numbers, if presented in graphic form like this, it becomes less extreme sounding; the large gap between the *native* trajectory and *OL2* trajectory help tell the story better. Simply put, the red line shows us what a good EFL coursebook *should* accomplish.

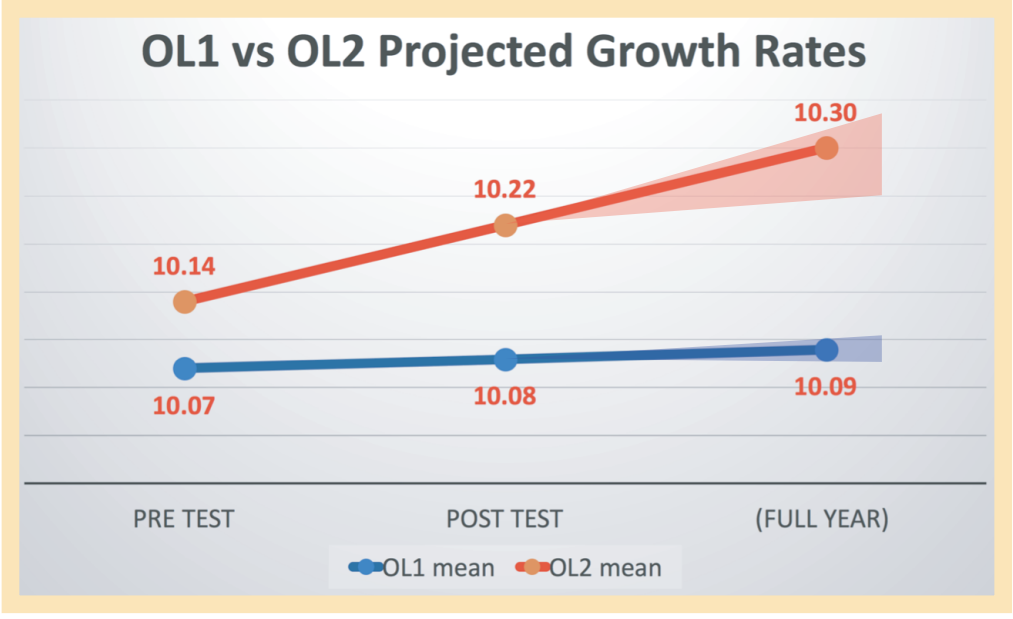
Note: One semester = 15 weeks + final exam. Two semesters is therefore 30 weeks + 2 final exams. "Full year" in this case is equivalent to 30 weeks of university teaching, not 52 calendar weeks. "Full year" projections are not based on existing hard data from the EFL context because there are none yet. It is imperative to have future studies that collect longitudinal growth data in EFL contexts that focus

on data accumulation for longer than single semester periods. This issue will be discussed further in the *Limitations* section of this thesis.

At this point in the thesis it is would be logical to circle back and reiterate Dynamic Skill Theory in light of these findings. Figure 67 compares cognitive development of "high vs. low support contexts" with *"OL1* vs. *OL2"*. Although the foundational intent of the *OL1* design was to create a high support context that brings on optimal levels (hence the coursebook title, *Optimal Levels*), this thesis' data makes it abundantly clear that, although the intent was established well-enough at the time, *OL1* did not do all that it set out to do.

It took several years of tweaking across this action research project to bring the *OL* series closer to the designated intent (in the form of *OL2*). Although this is a simple comparison, as can be seen in Figure 67, the resemblance between "high support vs. low support" and "*OL1* vs. *OL2*" is uncanny. Figure 67 shows the zone within Dynamic Skill Theory that was represented by the students' outcomes per book. From this blow up chart, we can see that *OL1* was essentially a low support coursebook when compared to *OL2*, which provided a high support context for the students.





Note: The data from the Japanese university EFL student range is 10.1 ~ 10.3; this range is typical of *native* English-speaking US high school grades 8~10. *OL2* students displayed growth from around *singular abstractions (Ab1)* to near *Abstract Mapping (Ab2)*, while *OL1* students were primarily stuck around *singular abstractions* (Ab1).

Figure 67: Reiterating Dynamic Skill Theory in light of OL1 vs OL2 growth trajectories.

Deductions and hypotheses

My first hypothesis is that "closing the gap" between demonstrable L2 and L1 complexity is *not as challenging as pushing students to increase their native complexity levels (their L1)*—most of the university EFL students' brains have literally already reached the capability of age-appropriate complex thought, albeit in Japanese. As we can see in Figure 67, the EFL students demonstrated L2 complexity on par with *native* US grades 8-10, even though these were first year university students. So, these students had to learn how to level-up their English competence in order to utilize their *already-existing-yet-never-used-for-English* university level thinking skills, and, this is easier for them to accomplish than to build higher level native skill networks.

I can say this because this part of the study shows that while the *OL1* class was closing the gap between their L1 and L2 performance at 1/2 the pace of American college student growth, *OL2* was capable of closing the gap at 4x the pace of American college students in their native language. Such a rapid growth rate would not be possible if *closing the gap* between L2 and native language was not easier than pushing current *native* complexity levels forward.

In other words, pushing the L1 in native students to gain points in demonstrable cognitive complexity is literally the hardest thing for a student to do (which is why growth rate is at 0.08/year at this age group, and why overall, even in optimal conditions, it takes over two decades for the human brain to cognitively mature). This study shows that playing *catch-up to the L1* is easier for students than it is to *grow in their L1*—that is, if we consider mean rapid growth of L2 complexity a good indicator of this.

This leads to point (2), my hypothesis regarding the pedagogical efficiency of *OL2*. The cycles of tasks embedded in *OL2* work synergistically to keep the students engaged because it puts cognitive demands on them that are based on their own passions but within a rigid cyclical framework of study, based on the COOPs protocol. So, the students are working toward higher levels of English proficiency with a growth rate efficiency that is attained by passion-based engagement in a

format that steadily trains them to make pre-emptive decisions based on the confidence they attain while in this unwavering pedagogical flow of events. The cyclical nature of the pedagogy enhances the efficiency of learning the more times the cycles are completed. As will be discussed later in this study (and nuanced in Figure 67), *OL2* students reported having considerably more diligence and having more out-of-class study time than their peers in other English courses. The majority of the data in this thesis strongly support these hypotheses.

If we frame the differences of affect (*OL1* vs. *OL2*) from the perspective of our brain's executive function (as in Meltzer, 2007), we can say that *OL1* students executive function in the prefrontal cortex is limited to lower level skills (especially concerning L2 developmental metacognition) while *OL2* seems to be encouraging more rapid executive function development in those areas. Higher executive function is naturally part and parcel of *cognitive reorganization.*

Figure 68 is an example flowchart from Meltzer et al. that can help in understanding portions of underlying complex development that *OL2* encourages--such as the solid nurturing of learning strategies. In retrospect, *OL2* does a much better job in these areas than *OL1*.

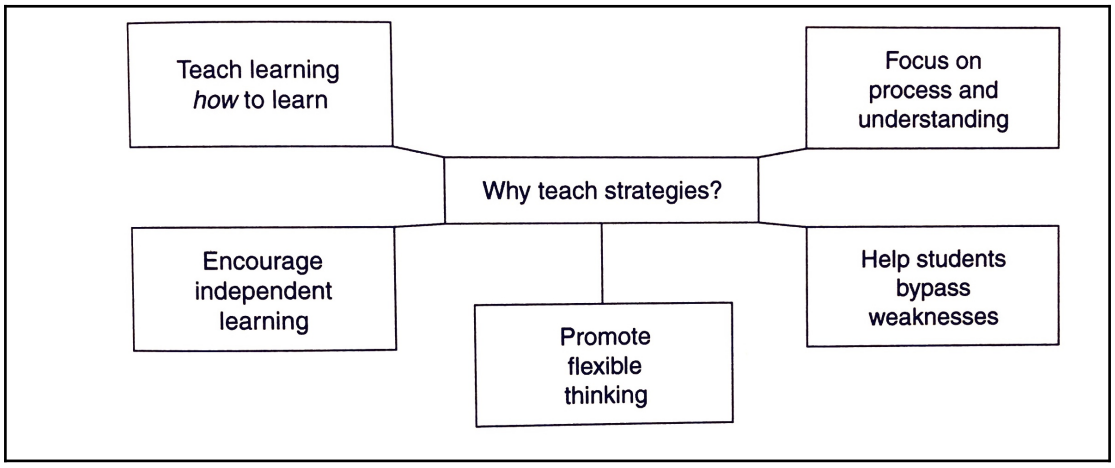


Figure 68: Why teach strategies? Example flowchart of how the development of executive function can deepen student understanding (Meltzer, 2007:167)

This leads to a pensive issue. Ultimately, only the *OL2* group demonstrated growth that can be recognized as *cognitive reorganization,* yet this recognizably and

arguably superior development in the students had little to no impact on their official semester grading at the university. This as an indication of a grading system that is much too simplistic; the evaluation system should be re-designed with demonstrated cognitive growth in the L2 as at least one of the evaluation points, if not the predominant point of evaluation.

All-in-all, these discussions culminate and spotlight the importance of establishing an understanding of *L2 Maturity*. Simply put, students' L2 competence should be evaluated with an understanding of each student's L1 competence--either as a group mean, or on an individual basis. Ideally, each EFL student's growth should be gauged in direct comparison with their personal L1 performance and growth.

Put in this way, knowingly asking an EFL student to do a cognitively complex task that is certain to be above their own native competence would be ridiculous--and the ridiculousness should be obvious at that point in time. Yet, it is easy to imagine that many of these ridiculous tasks are being assigned to EFL students across the globe (and causing unnecessary frustration), all because native cognitive prowess is hardly ever considered by the EFL teacher unless the student is noticeably cognitively impaired.

The need for an understanding of L2 Maturity, and how to teach EFL with L2 Maturity in mind, should be considered essential—but sadly, this far from the norm. Because it has become a prominent feature of this study, L2 Maturity will be discussed in detail in 10.1, as a mathematical equation, and as an Applied Linguistics construct--to make the concept easily tangible to all teachers and researchers.

7.4 Five years later: the affect of teacher rapport

As this was an on-going action research project, each year of research produced a new set of ideas for improved data collection--based on previous years' revelations. The data collection methods of each coming year were therefore modified to accomodate the newest emerging ideas. My final data collection sessions—regarding the affect of teacher rapport—were concluded during the 2018-2019 school year. Typically, quantitative data would be reviewed here in

Chapter 7. However, as previously discussed, this thesis is atypically mapped, where it is logically appropriate to do so. Regarding the 2018-2019 data set, the deductions and conclusions are more easily understood when the quantitative and the qualitative data are presented concurrently. I have therefore put them back-to-back in the next chapter, in Sections 8.11.1 and 8.11.2.

7.5 Conclusion

Several principal deductions regarding *OL2* can be made from the data presented in this chapter. The data presented here was collected from a wide range of categories: (1) comparisons with prominent coursebooks, (2) longitudinal comparisons regarding perceived proficiency, (3) longitudinal comparisons regarding envisioned future L2 usage, (4) comparisons between the primary study and the replication study, (5) students' perception of *OL2* vs. traditional coursebooks, and (6) the comparison of *OL1* vs. *OL2* growth rates as demonstrated in their L2s using Lectical Analysis. This section shall summarize the findings of each of those six categories.

(1) ***Comparisons with world-class coursebooks:*** A "department composite" was created with student questionnaire data spanning four prominent ELT coursebook series taught by the native English speaking teachers in the department where the primary study was conducted. When compared to the department composite means, *OL2* outperformed the composite scores in seven of the nine categories on the questionnaire. This shows that *OL2* can hold its own when compared to major league coursebooks, but there is still room for improvement – thoughts on points for revision will be discussed in Chapter 9. *OL2's* most prominent strengths in this portion of the study are regarding "Course content," "Own diligence," and "Study outside of class." From the collected data, *OL2* seems to be more engaging inside and outside of the classroom (than the composited means from the other coursebooks). My hypothesis is that the repetitive cyclical design of *OL2* nurtures automaticity via student-borne pre-emptive learning strategies that keep

their minds aligned with the course's goals for more hours than the other coursebooks. This hypothesis is substantiated by qualitative data from the students in Chapter 8.

(2) ***Longitudinal comparisons regarding perceived proficiency:*** Students using *OL2* in the primary study showed a marked increase in perceived proficiency across the year of usage. Descriptions of these perceptions (in the words of the students), will be reviewed in Chapter 8.

(3) ***Longitudinal comparisons regarding envisioned future L2 usage:*** Students using *OL2* in the primary study showed a decrease in confidence of *English usage in the future* at the end of the first semester, but by the end of the year (two semesters of *OL2*), the students reported confidence levels that surpassed their established "societal norm"—portraying a considerable boost in confidence by the end of two semesters of study with *OL2*.

(4) ***Comparisons between primary study and replication study:*** The replication study showed that the students in the replication study (conducted the year after the primary study year), reported their rise in confidence a semester earlier than the primary study group, and then their means generally plateaued across the second semester. I hypothesize that this quicker rise in confidence levels is largely due to the teacher's confidence *projected-unto-the-students* in the pedagogical strategies being employed. And, at the same time there may have been a manifestation of an acquired ability to play down the-then coming-to-light weaknesses of the coursebook. These intra-year trajectories are intriguing and worthy of more in-depth study. That said, I would like to call attention to the consistency of the year-end outcomes → the longitudinal 4-year study consistently showed plateaued responses across all nine categories across all four years of the study—*by the end of each of the years*. Therefore, it may be also be safe to assume that the results of the primary study and the replication study are satisfactorily representative of year-long *OL2* usage (even if there was variation in the mid-course trajectories).

(5) ***Students' perception of OL2 vs. traditional coursebooks:*** Other than the "department composite" means, I also administered year-end questions for students regarding the efficacy of *OL2* vs. traditional English coursebooks. In both the primary study and the replication study, the reported means indicate that the students reported that *OL2* had higher efficacy than traditional coursebooks. This dataset is in strong agreement with the dataset from the study that compared *OL2* with the department composite. Moreover, this data indicates that by the end of each of their terms, *OL2*-using students seemed to have embarked in developing a critical appreciation of coursebook design—and how *OL2* differs from the more traditional coursebooks that they had used in the past. (Chapter 8 illustrates this point in greater depth with several representative quotes from the students.)

(6) OL1 ***vs.*** **OL2** ***comparisons regarding cognitive growth rates as demonstrated in their L2s via Lectical Analysis:*** The demonstrated growth rates of students using *OL1* and *OL2* (as demonstrable in the L2) were compared via Lectical Analysis. In raw numbers, the growth rate demonstrated by *OL1* students was of a slower pace than the benchmark native student rate (roughly 1/2 the benchmark rate) while *OL2* students demonstrated faster growth than the L1 benchmark (roughly 4x the benchmark rate). Directly compared to each other, *OL2* outperformed *OL1* with roughly a 8x difference in growth rate. With inferential statistics, this growth was found to be statistically significant with a medium effect size.

The above deductions collectively portray *OL2* in a favorable light. This new book clearly outperformed the older book, and it held its own in comparison to world-class coursebooks. However, this data and the statistics can only supply half of the story. The next chapter—the qualitative data review—completes the story by reviewing student voices on their thoughts and realizations from *OL1* and *OL2* usage. Those voices shall also provide credibility to the claims made in this chapter regarding the *hows* and *whys* concerning the efficacy of *OL2* .

8. Evaluation of the coursebooks based on interviews and student essays

This chapter, along with the previous chapter (Chapter 7), offers an evaluation of two coursebooks (*OL1* and *OL2*). Whereas the previous chapter focused on *quantitative* data, this chapter will focus on student interviews and essays regarding coursebook usage. First, I will reiterate the primary data sources and the data types; I have chosen to use multiple data sources to help better piece together the complex puzzle that comprises the performance differences between *OL1* and *OL2* (as established in the previous chapter), and to learn from them. Therefore, in the earlier sections of this chapter, I will provide a comprehensive qualitative analysis of *OL2*—by compiling findings from those various data sources and systematically analyzing them. I will then provide a brief analysis of *OL1,* allowing for a comparison to be made between the two different coursebooks from a qualitative perspective. As discussed in Chapter 7, my primary studies were conducted on first-year students in the 2014-2015 school year, but I will also examine qualitative data collected from the longitudinal 4-year study (2014-2017).

8.1 Overview of data sources and data types

The majority of the qualitative data collected for this thesis was open-ended. During the time span of the primary study (2014-2015), there were seven courses taught per semester, with two semesters taught per year. One of the courses used *OL1*, and six of the courses used *OL2*. Because both coursebooks are from the same family of coursebooks—with equivalent goals—there was no "control group" *per se*. However, because of the way the study was established, the *OL1* group became a de facto control group, and the *OL2* group became the de facto experimental group. I recruited 4 volunteer members per course to form one focus group per course.

Therefore, there was a total of seven focus groups for the primary study. The protocol of the focus group sessions was to have one initiation ("Getting to know you") meeting at the beginning of the semester, then continue to have one meeting per coursebook Module. That made a total of four meetings in first semester. To my surprise, at the request from the focus group participants, we conducted more focus group meetings during the second semester; these second semester meetings were not in the protocol. It seemed that the Focus Group (FG) participants found the FG discussions during the first-semester to be cathartic and an enjoyable usage of time – and so they requested to have more of them. I hypothesize that being able to discuss classroom pedagogy in terms of their own development—as well as the development of their peers—seemed to be meaningfully satisfying for them. The meetings were held bilingually; I transcribed and then noted the significant points for every meeting. The meetings followed this embedded 9-question format (Figure 69):

1.Class progress: "How is the class doing?" (focus on peers)
2."Now let's discuss each cycle. Open up your books to Cycle x."
"What can you say about this Cycle?"
a. Cycle 1:
b. Cycle 3:
c. Cycle 3b:
d. Cycle 4a:
e. Cycle 4b:
f. Cycle 5a:
g. Cycle 5b:
h. Cycle 6:
i. Cycle 7a/b:
j. Cycle 8:
k. Cycle 9:
l. Cycle 10:
3."Which cycle is developing the best? Which is the worst?"
a. The best cycle(s)...
b. The worst cycle(s)...
4. Positive points about the course book:
5. More general comments:
6. Possible issues/difficulties with the coursebook:
7. Recommendations/suggestions:
8. Regarding the learning:
a. "Which tasks worked best in the class?"
b. "Which tasks could be further fine-tuned?"
9. Elicit ideas:
"Any ideas, comments, or final words…?"

Figure 69: The focus group interview question list (meeting protocol)

My other two forms of qualitative data collection were not interview-based—both of them were open-ended writing sessions regarding the coursebook. In both cases, I asked the students to write about the course with the following instructions: "I would like you to think about this term/year, and write your thoughts down. It could be an evaluation of the course, or anything that comes to mind about our time together. There are no wrong answers. You have 15-20 minutes." I collected these responses at the end of the 1st semester and at the end of the 2nd semester (end of year). The essays were read and then coded for positive and negative themes by a team of four readers—myself, as team leader, and three Japanese students (upperclassmen volunteers with a good command of English).

8.2 How the themes emerged

In the spirit of *grounded theory*, I did not prime my coding team to look for anything specific beforehand. I simply taught them how to "read the data" and while reading, mark off (hand-code) recurring themes that they noticed—using colored pens, post-it sheets, and notebook paper. This was a discovery session that lasted over 20 hours, spanning several months. Our team—myself included—initiated ourselves by first simply reading all of the focus groups' responses. From that reading experience, we noticed recurring themes in the focus group responses and listed them as our first codes. In this way, our primary list of recurring themes naturally emerged from the focus group data. With that primary list in hand, we then coded the first semester essays. Our coding team found several new recurring themes, and we added them to the theme list. Finally, we examined the second semester (end of year) essays. More new themes emerged, and we appended the list accordingly. This concluded the emerging theme list assembly.

Our finalized list of themes was composed of (A-B styled) symmetrical pairs of polar-opposite themes. That is to say that we created a polar opposite theme for every theme that emerged—even if we did not immediately find a corresponding "opposite" theme in our readings. Why establish a list of symmetrical pairs of themes? During our first coding session, we quickly realized the benefit of having

polar-opposite theme codes for each of the emerging themes. As an example, when "writing skills improved" emerged as a theme, we also established a "writing skills did not improve" theme code. By having this binary, polar-opposite coding system, the scanning process became more logic-based, and we were able to process the data with efficiency—and greater accuracy because we had more firmly established criteria in our minds as we scanned the data. This binary system helped sharpen our eyes for the "opposing themes" that we likely would have missed had we not established these polar-opposites as selectable code choices in the established list. Perhaps more importantly, this processing style allowed us to realize the significance of "zero-frequency" themes.

As a rule, regardless of which emotional valence emerged first, all positive versions of the themes were given odd numbers on the theme list, while the negative versions of the themes were given even numbers on the list. Each of the polar-opposites were kept together as a binary pair for easy reference. This binary symmetry in the list gave us the opportunity to work with a theme list that included zeros.

Why is this significant? Zeros as qualitative data are easily overlooked because in qualitative studies (such as in some grounded theory studies), researchers are typically looking for what has *emerged* from the data—what is missing from the data is not readily visible, nor is it typically on the agenda. I argue that these "zero-frequency" themes can have implications that are as significant as themes of high-frequency, and so in this portion of the study I have found ways to make good use of zero-frequency data as well as high-frequency data. Our team's finalized list of paired themes are listed in Figure 70.

A1. English skills improved
A2. English skills didn't improve
B3. Socio-cultural skills enhanced
B4. Socio-cultural skills lacking
B5. Presentation skills improved
B6. Presentation skills lacking
B7. Leadership/management skills improved
B8. Leadership/management skills lacking
B9. Logical thinking improved
B10. Logical thinking lacking
C11. Motivation/affect improved
C12. Motivation/affect lacking
C13. Metacognition/deeper thinking developed
C14. Metacognition/deeper thinking lacking
D15. Book design, positive
D16. Book design, negative
D17. Positive outlook regarding course
D18. Negative outlook regarding course

Figure 70: The finalized list of paired themes for qualitative coding.

8.2.1 Generation of the frequency scores

In qualitative data discovery sessions, it is not uncommon to uncover a theme/response and to then intensely work at deducing its significance. For this study, I also dug into individual responses in long deduction sessions. I also decided to add a variation to the discovery process; I worked with my coding team to produce theme emergence *frequencies*. We read through all the occurrences of each theme and collected them into a single chart, producing quantitative data from what was originally qualitative data. For example, if the theme "English skills improved" was mentioned four times in one interview, that interview would gain a frequency score of "4" for that particular theme. This frequency chart became our reference for the cline of popular to unpopular themes. This cline was also used extensively throughout the writing of this chapter to help make deductions regarding the efficacy of, and differences between, *OL1* and *OL2*.

8.3 Positive and negative themes of OL2 from the focus groups

For this portion of the study, I closely followed the same protocol for both groups (*OL1* and *OL2*) but I found a significant difference in response outcomes. In typical qualitative analysis, comparisons are often avoided because of the non-representative samples; however, in this case I found it useful to view the ratios between positive and negative outcomes of the two coursebooks. As illustrated in Figure 71, the *OL1* (*Original)* group produced a considerably larger percentage of negative comments than the *OL2 (NewOL)* groups did. *OL1* showed a 70:30 ratio between positive and negative themes. However, the *OL2* groups' tally came to a 93:7 ratio between positive and negative themes.

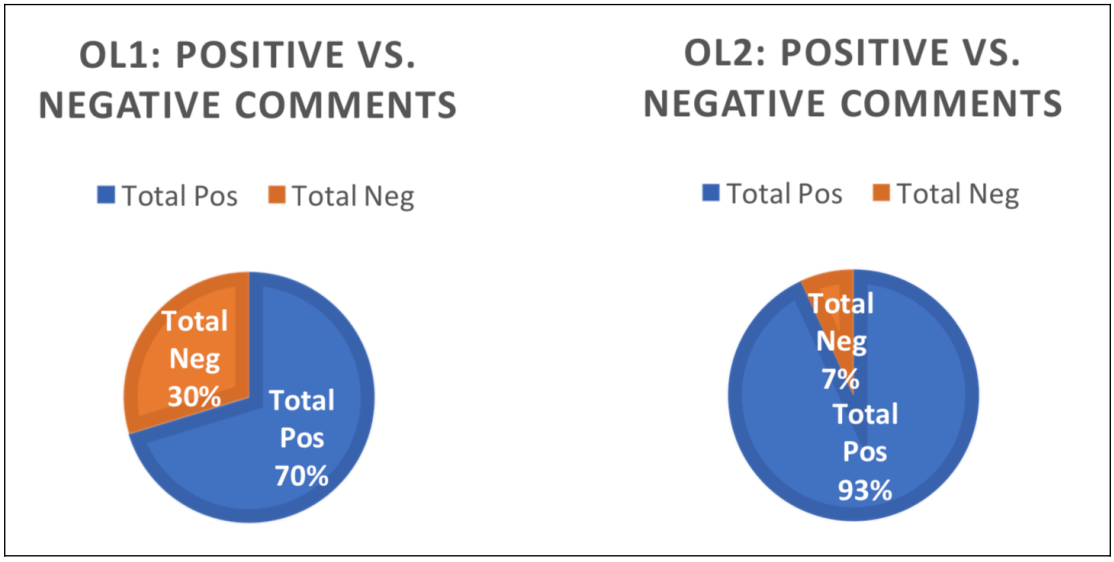


Figure 71: Positive vs. negative theme ratios, OL1 (Original) and OL2 (NewOL) focus group meetings

Both groups' themes (from *OL1* and *OL2*) were combined into a single list. The entire grouping is listed in Figure 72.

A1. English skills improved
A2. English skills didn't improve
B3. Socio-cultural skills enhanced
B4. Socio-cultural skills lacking
B5. Presentation skills improved
B6. Presentation skills lacking
B7. Leadership/management skills improved
B8. Leadership/management skills lacking
B9. Logical thinking improved
B10. Logical thinking lacking
C11. Motivation/affect improved
C12. Motivation/affect lacking
C13. Metacognition/deeper thinking developed
C14. Metacognition/deeper thinking lacking
D15. Book design, positive
D16. Book design, negative
D17. Positive outlook regarding course
D18. Negative outlook regarding course

Figure 72: List of coded and paired themes from OL1 and OL2 (focus group data)

With regard to the positive themes (Table 33), the top three themes are of particular interest. They are "Book design, positive", "Motivation/affect improved", and "Metacognition/deeper thinking developed." With over 20 years of teaching experience in Japan and more than 3 decades of living in Japan, I can say with conviction that these top themes are not what I would imagine to be typical Japanese EFL students' highest three remarks regarding their own English coursebooks.

Table 33: The positive themes that emerged and their frequency.

OL2 positive themes (focus groups)	Frequency	Illustrative example
D15 Book design, positive	221	"With this book we could get new vocabulary… we can raise our presentation skills… and by assessing and reviewing, it became clear where need to improve our English…" (Focus Group 4, 2nd meeting; May 2014)
C11 Motivation/affect improved	175	"…as we enjoyed using the book, it gets better. Strong bonding. When we watched a movie together, we bonded - even outside of class. The book made us bond. Bonding was strong." (Focus Group 3; 3rd meeting July 2014)
C13 Metacognition/ deeper thinking developed	168	"…There were many stages - had me think deeply more than usual... the learning does not end quickly… they (the learning) build up, then go to the presentation. It is the first time for me to think so deeply even in Japanese." (Focus Group 6; 2nd meeting, May 2014)
A1 English skills improved	76	"Easier to talk to people now. Conversation flows." (Focus Group 7; 3rd Meeting, June 2014)
B9 Logical thinking improved	33	"It became clear what to do by next time. I felt clarity, having the discussions with partner I could understand others opinions better -was helpful, it was helpful hearing other ideas - plans, I can actually do the plans that we came up with." (Focus Group 4; 2nd Meeting; May 2014) "…choosing the theme and then adding the words lets us complete the presentation easily (it seems to write itself). Our group completed this task (Cycle6) in correct order, so this worked VERY well for our group." (Focus Group 5; 2nd meeting; May 2014)
B5 Presentation skills improved	30	"I can remember more than in high school. I learned how to construct good presentations. I try to create presentations that are different from other groups. It is fun and everyone is getting better! Confusion turns into learning - more motived too. We have 'forward planning' now." (Focus Group 2; 3rd meeting; June 2014)
D17 Course, positive	30	"(Regarding the course) Because we totally make this on our own, we get what we put in... not top-down. It is student-centered... felt like proper university-level thinking. The peer assessment makes us work harder. It teaches us to be more unique with our designs and output. We now want to do better than the other groups... group rivalry. We learned how to have better group work... these skills are MORE THAN English.... learned much more than that." (Focus Group 2; 4th meeting; July 2014)
B7 Leadership/ management skills improved	13	"Leadership has become more common! It seems more fun… smiling and relaxed." (Focus Group 1; 2nd meeting July 2014)

B3 Socio-cultural skills enhanced	3	"I learned how to add interesting points... learned how to work with non-friends (in group work, etc.) I learned how important teaching is - thinking about how to make tests, power points, eye contact... I've never had teaching experience, so it opened my eyes to teaching methods, and presentation methods. I am standing on the other side of the desk now... learned a lot from this." (Focus Group 2; 3rd meeting; July 2014)

The fourth theme on the list (Table 33) is "English skills improved." An immediate reaction to this might be that seems to be a blunder – shouldn't "English skills improved" be on the top of the list if it is an excellent English coursebook? This is a particularly intriguing key issue that I will be methodically shedding light on across this chapter.

I hypothesize that the *OL2* group produced that faster rate of demonstrable L2 cognitive development (as discussed in Chapter 7) specifically *because* the design of *OL2* allows for the development of motivation and metacognition to be prioritized over the development of specific English skills. Moreover, it should not be overlooked that these first-year students at the university (none of them being Applied Linguistics majors), formed these specific top three "deep" themes on their own—having only their usage of *OL2* as their prompt. This in itself may be a vital piece in the puzzle for understanding the affect and efficacy of *OL2's* pedagogical design.

All of these "deep" themes at the top of the list do not concern anything that would be considered to be clearly tangible to the students. It is also remarkable that they do not represent anything that is commonly evaluated by the students on a "performance" or a "skill" scale either. The top themes that emerged are decidedly abstract concepts for these first-year students, yet they surfaced as the most commonly discussed themes. Only when we look down the list to slots 4 and 6 do we finally see the more concrete performance-based observations such as "English skills" and "Presentation skills". These concrete skills are what I would have (and still do) assume that typical Japanese students are more likely to naturally fixate on when asked about their coursebooks. Therefore, it is interesting that the theme of "Metacognition/deeper thinking developed" seems to be displaying (reciprocating?)

its own affect on the students by being high on their list. This is an enlightening and amusing phenomenon; it hints at the possible existence a self-perpetuating virtuous cycle of events taking place.

Concerning the negative themes (see Table 34), the first important observation is that there were considerably fewer negatives than there were positives. There were 749 positives vs. 56 negatives. The negatives represent 6.9% of the themes. As I did with the positive themes, I established a theme frequency table, but because the frequencies are so low, it would be misleading to say that there was a "top three"*, per se.* The lone double digit list-topper on this negative list is somewhat at odds with the positive list; it is "Book design, negative." This signifies that there were some unresolved (negative) issues with the coursebook. What those points were, and how they were resolved, will be discussed below and in Chapter 9.

Table 34: The negative themes that emerged, their frequencies, and illustrative examples.

OL2 negative themes (focus groups)	Frequency	Illustrative example
D16 Book design, negative	30	"5b not useful." (Focus Group 1; 3rd meeting; Aug 2014) "5b skipped." (Focus Group 2; 3rd meeting; July 2014) "Cycle 5b may not be necessary. We skipped it." (Focus Group 5; 3rd meeting, July 2-14) "5b not useful." (Focus Group 1; 3rd meeting; Aug 2014)
D18 Course, negative	9	"…because it's our own assessment, sometimes we are not sure if we are totally correct. I wonder if the grammar of the other teams' presentations are correct. I want the teacher to correct our content more. I can't tell if my own sentences are getting through." (Focus Group 2; 3rd meeting; July 2014)
C12 Motivation/ affect lacking	8	"Some members of my group got lazy. I wanted to raise the quality, but this time we didn't meet after class to practice like we should have." (Focus Group 7; 2nd meeting; June 2014)

In Table 34 I have tabulated the three highest frequency negative themes. The non-tabulated negative themes (including zero-frequency themes) are: Presentation skills lacking (5), Metacognition/deeper thinking lacking (3), English skills didn't improve (1), Socio-cultural skills lacking (0), and Leadership/management skills

lacking (0). The adjacent bracketed numbers are the respective frequencies. This data from the students was helpful in uncovering the lingering issues with *OL2*'s design. In particular, Cycle 5b was clearly not appreciated by the students as can be seen by the four illustrative quotes. The zero-frequency negative theme data were regarding *not developing leadership skills* and *not developing socio-cultural skills*. The single-frequency theme was regarding non-English skill improvement. I believe that it is important to note that zero to one of the students thought that these areas did not improve; this should not be conflated with the data on the positive theme list.

8.4 Student appraisals of the sections of OL2 from the focus group

In order to better understand students' overall appraisal of *OL2*, I also asked the focus groups to discuss which specific "Cycles" (sections) in the coursebook they liked the most, and why.

Each of the sections within *OL2*'s Modules are officially titled and referred to as "Cycles" for the benefit of the students—to highlight the cyclical nature of the syllabus design for the students. Moreover, as discussed in Chapter 7, each Module of the coursebook unwaveringly follows an identical pedagogical flow, thus creating cycles within cycles, not unlike the design of a Mandelbrot fractal (Figure 73).

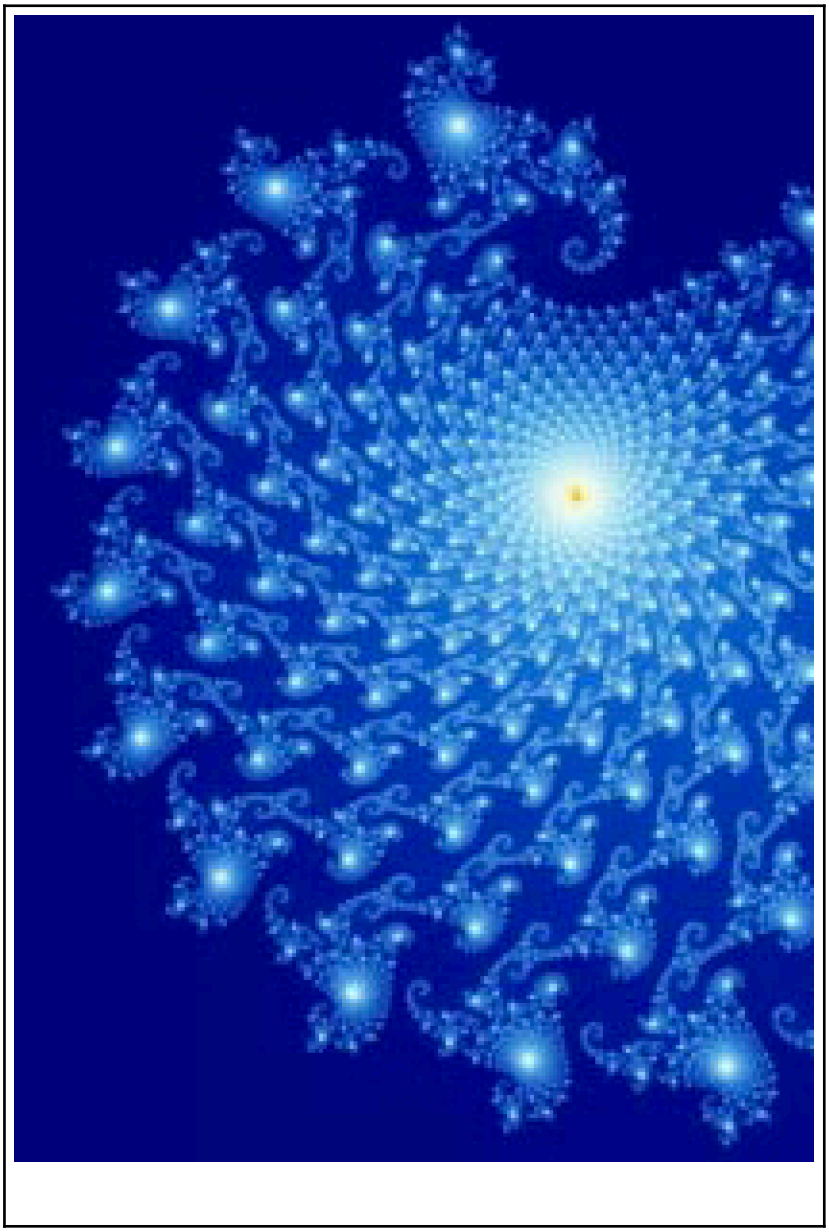


Figure 73: Fractals as an image of cyclical pedagogical flow and development

I wanted this fractal-like aspect of the learning to be explicit to the students (and other teachers using the coursebook) because it is so central to the pedagogical theory of *OL2—and where OL2 most dramatically diverges from OL1's design.* By the end of the second semester (one year of teaching), the students experienced six cyclical learning modules, all of which had 10 of these smaller cycles embedded within each of them, hence the Mandelbrot fractal as an effective analogy.

8.4.1 The most popular cycles

*Table 35: Focus groups' chosen best cycles' frequencies. *Highest are highlighted*

Cycle number	Frequency "chosen best"	Percentage
1	1	2.4%
2	0	0.0%
3a	2	4.8%
3b	2	4.8%
***4a**	**7**	***16.7%**
4b	3	7.1%
5a	3	7.1%
5b	0	0.0%
***6**	**10**	***23.8%**
7a	2	4.8%
7b	3	7.1%
***8**	**6**	***14.3%**
9	0	0.0%
10	3	7.1%

As can be seen in Table 35 and Figure 74, the most popular cycles were Cycle 4a (16.7%), Cycle 6 (23.8%), and Cycle 8 (14.3%). Of those top three, Cycle 6 was the most popular (23.8%). This data helps shed some light on the positive factors of the coursebook, and its efficacy. The first question is, what made these three cycles the most popular? And, how do they differ from the other cycles—especially their efficacy?

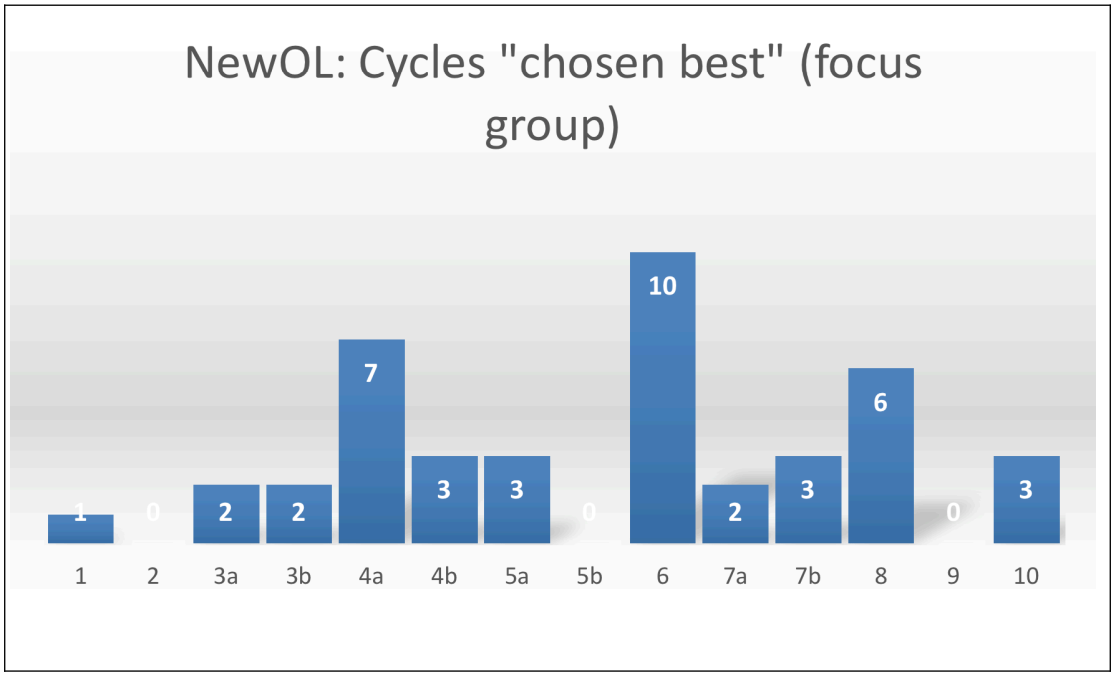


Figure 74: Bar graphs of the "chosen best" cycles (focus groups).

Cycle 6 Discussion: For the students, although Cycle 6 was located in the middle of the module's line up, Cycle 6 is the *perceived climax* of the module (Figure 74 and 75). It is where each student's work up to this point suddenly takes a seriously tangible (concrete) form. This tangible outcome had been nurtured by bursts of creativity and autonomy on their part until this point, but had no real outlet until Cycle 6.

The yellow sidebar next to the *Slide Design boxes* on the Cycle 6 pages (Figure 75) logically require the students to title an apt and specific theme per slide, then it requires the students to transfer the most appropriate lexis from their Master Lists, to be combined in ways of their own choice—and to be used as speaking prompts for when they present the slides to the class during their group presentation. In this way, all of their work until now manifests into a real product (a presentation format) that will be performed in front of their peers and the teacher—for official course assessment, but also with an element of "edutainment" embedded into the presentation for the sake of the viewers.

Considering how all the learning until this point gives the students what they need to perform well in front of their peers, and that this cycle gives them controlled autonomy regarding how they will design the implementation of it all, it can be logically predicted that this cycle (Figure 75) would naturally be the "chosen best" cycle. What did the students have to say in support of this?

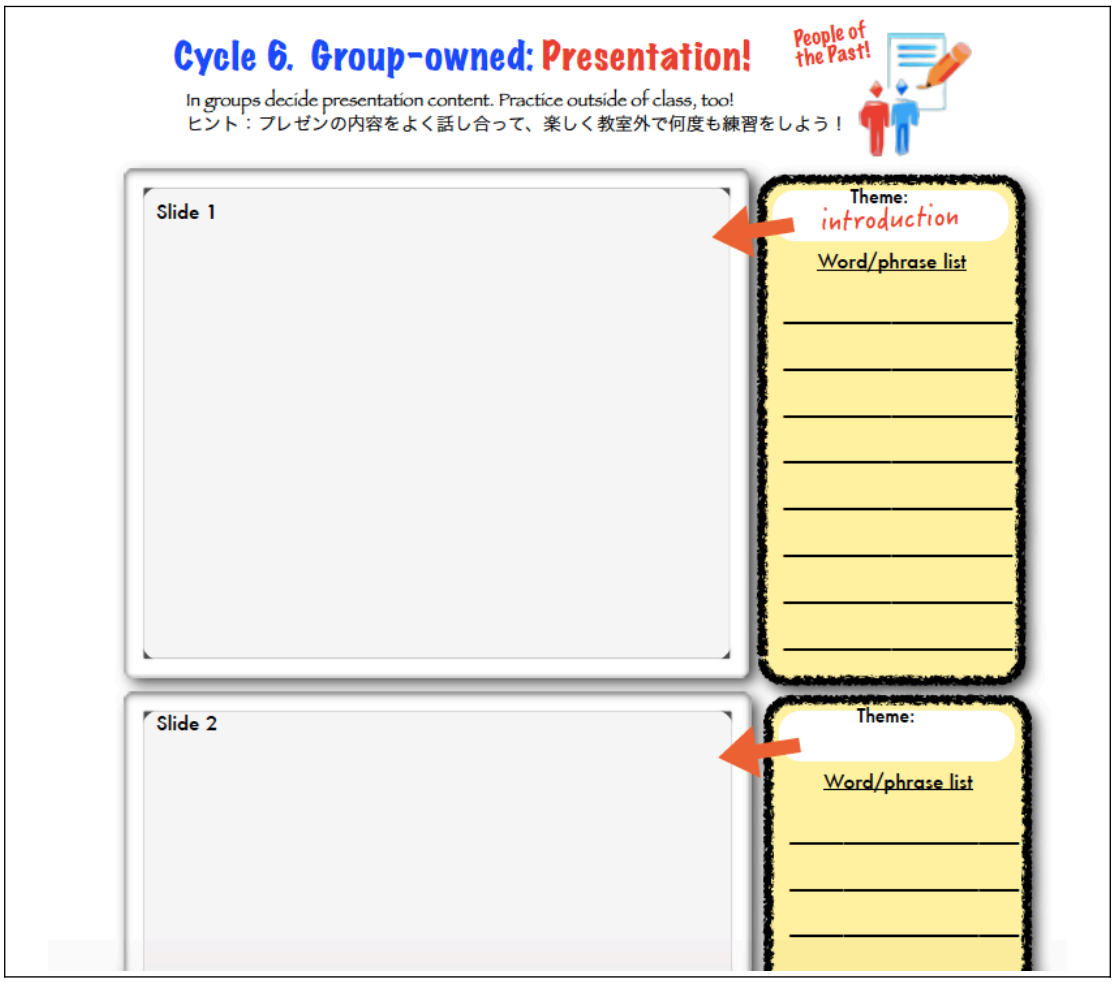


Figure 75: Cycle 6: "Group Owned: Presentation!"

> *"...the usefulness of the sidebar became apparent during the presentations! Actually, for this first presentation my group didn't make a clear flow and didn't practice enough, but now we know better!"* (Focus Group 5; 1st meeting; May 2014)

Being a radically different approach of English study for most of these participants, we can see that the students may have started off being unsure/uneasy about the overall pedagogical design, and in this case how to use the sidebar (initially), but the metacognition/realization was made apparent during the presentation— this is a pivotal time for such an "aha moment" to occur.

> *"…last time I used the sidebar well. Six slides is the perfect amount. The sidebar is useful because we can remember the words more easily. For the audience, it is also very good."* (Focus Group 5; 2st meeting; June 2014)

Similar to the first quote, this second quote and the third quote (below) talk highly of the sidebar and how they found it useful during the presentation later on ("very good for the audience"). They also talk about how it makes the lexis more memorable and how it is helpful for the group and for their peers.

> *"…choosing the sidebar theme and then adding words lets us complete the presentation easily… it worked very well for our group."* (Focus Group 6; 2nd meeting; June 2014)

Cycle 4a Discussion: What made Cycle 4a (Figure 76) another student favorite? Somewhat ironically, it seems that the *top-down* approach of Cycle4a has much to do with its popularity. On paper, this cycle may not look very interesting. It literally is a "lecture" from the teacher with the "gift" of lexical items for the students to choose from if they want to. The students are expected to take notes on this designated page in their coursebook—regarding points of their choice from the teacher's lectured content. The mystery behind this cycle's high popularity becomes clear when it is *experienced as a part of the continuum* from earlier cycles. This cycle *is* primarily about students listening to a lectured mini-presentation from the teacher—but it is more effective than the average lecture because of its strategic

placement within the module. Had this "lecture" come at the beginning of the module, it would be indistinguishable from other mainstream coursebooks' teaching strategies and would surely have put some of the students to sleep (this is common in Japanese classrooms). But the placement of Cycle 4a is what made the difference. The rationale is discussed below.

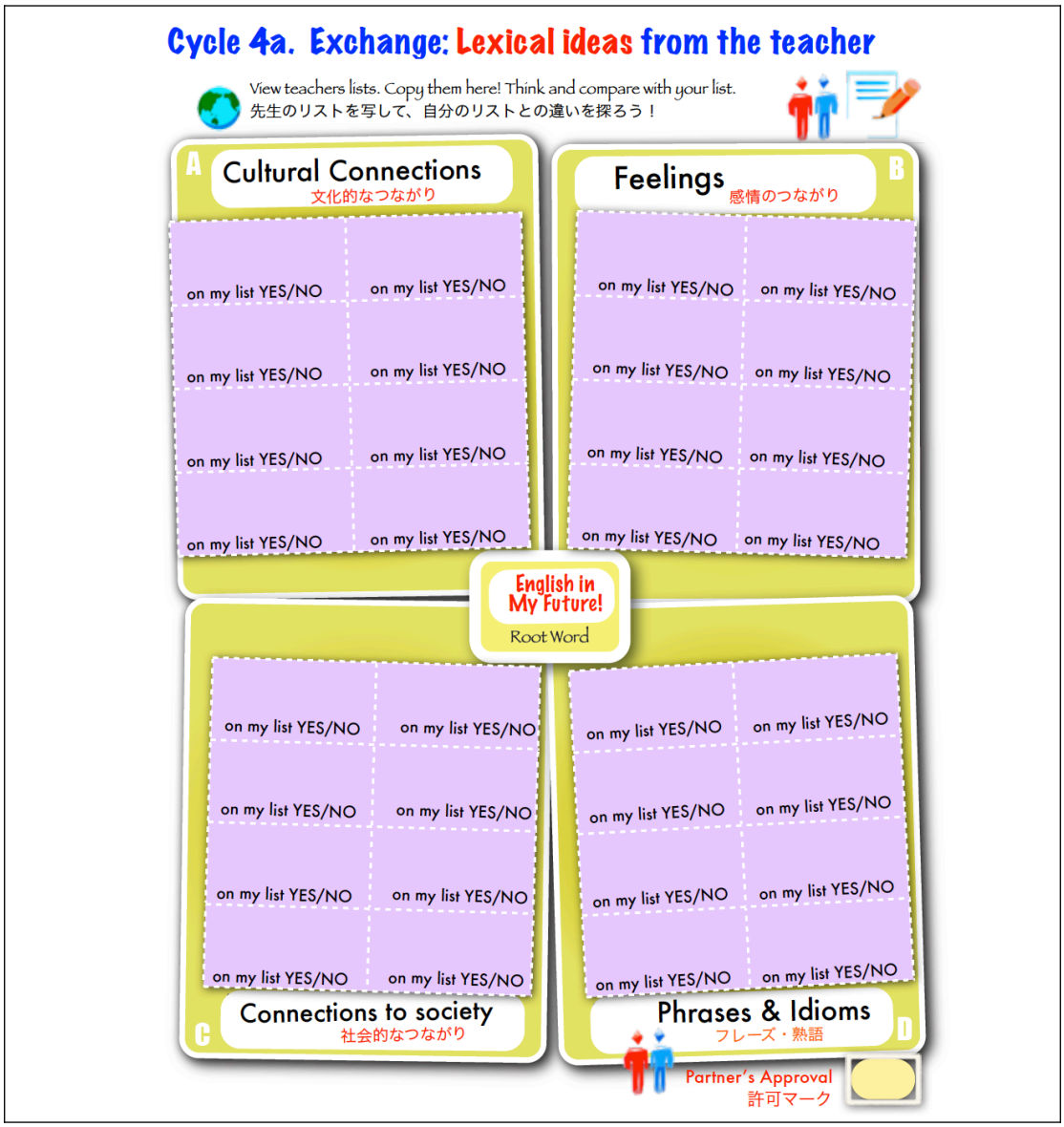


Figure 76: Cycle 4a: "Exchange: Lexical ideas from the teacher"

Digging deeper: Cycle 4a comes immediately after the students had a go with their own *bottom-up* thinking and writing in the earlier cycles. At this point in

the Module, keen students in the class will begin to viscerally realize (they manifest a "gut feeling" need) for better lexical items and more appropriate grammar to express what is building up in their heads. That feeling of this growing frustration creates a very real need for linguistic help in the students' minds –be it on a conscious or a subconscious level.

As the students have this burning (and ever-growing) desire for better lexis and grammar patterns to use in their upcoming presentation, Cycle 4a seemingly arrives out of the blue, but with perfect timing for their salvation. In that moment, the teacher entertains the students with lectured content that "magically" incorporates lexical items that jump out at the students as being *immediately useful* for them. This "magical" property is of course not magical at all; this teacher's mini-presentation has been pre-designed to dynamically match the specific needs of the students. So, it is inherently a gratifying presentation that satisfies the students' growing appetite for appropriate lexical items and grammatical forms. It may not be magic, but it does wow the students because the pre-planned content delivery is feigned as an ad-libbed talk from the teacher.

Furthermore, the teacher's inspired "ad-libbed" delivery may nurture aspirations toward better presentation skills (another naturally growing need within the students) –and, it also may nurture aspirations of better English fluency (another naturally growing need within the students).

For this activity to work well, the teacher should not come across as being preachy nor pedantic. Although the teaching content is being provided *top-down* at the time, the students are being asked to do *bottom-up* work with the content (personally choosing what they feel will be appropriate for them with their upcoming presentation). This creates an intriguing dynamic—the teacher's feigned "ad-libbed" content is appreciated on a practical level as well as at an emotionally appreciative level—which is unlike that of the typical lecture-styled English lessons that they are more used to experiencing. In those more typical lectures, the lecture content is typically presented to the students before the the students have viscerally felt the need for that specific content. Several quotes from the focus group have been chosen

that illustrate the complex dynamics, affect of the timing, and the students' feelings regarding this cycle:

> *"…this was central to our work…made me very happy! It was core and very useful and to see the real usage—great new words…helpful."* (Focus Group 1; 2st meeting; July 2014)

> *"Because of this cycle, I felt deep down that I needed to raise my vocabulary level. I knew some of the words that the teacher demonstrated, but I never thought of using them until now."* Focus Group 2; 2st meeting; May 2014)

> *"…lots of useful phrases. They were easy to apply. The phrases were useful."* (Focus Group 2; 3rd meeting; May 2014)

> *"…allowed me to expand my mind and seek out other idioms. Made me happy."* (Focus Group 6; 3rd meeting; May 2014)

The students make it clear that there is emotional appreciation regarding the flow of events surrounding this cycle—even at a metacognitive level. A highly desirable *package of emotions*—euphoria combined with the motivation to be more diligent—is apparent from the quotes. This "package of emotions" will be further explored below.

Mid-summary -- the "magic" of cyclical pedagogy

The built-in cyclical support system – the flow from Cycle 1 to Cycle 6, inclusive of the Master List usage, culminates here in Cycle 6. Eventually, the quicker students begin to recognize that some sort of ingenuity has been built into the flow of the coursebook, and how, as a package, it all helps them do a better job. Their growing appreciation of the pedagogical design comes in the form of *"aha moments"* during their presentation preparation, and during the actual presentation.

The culmination of these aha moments with the range of emotions that are inherent with presentation accomplishment is probably what makes Cycle 6 so popular with the students.

Further thinking: Had this instead been a task without that yellow sidebar for notation, nor had it been a task that required the group-led *conception* of an appropriate theme title per slide, nor had it been a peer-to-peer assessment activity, it may not have culminated with affective aha moments and it may not have stood out as being a particularly helpful page. But because of the dynamic and viscerally affective culmination, this is where affect, and the students' metacognition of that affect, go hand-in-hand in helping to produce even more student maturity while nurturing dynamic growth internally—while helping the students realize their own *needs* regarding the growth of the L2 within themselves. From the data it seems that these realizations are pleasant and perhaps even cathartic for some of the students.

8.4.2 The least popular cycles

Figure 77 is a stacked bar graph of *OL2's* popular vs. unpopular cycle data. I have stacked them here to show discrepancies and/or to see if there can be any logical deductions. Although the numbers were much lower, there were a few "Chosen worst" cycles (orange). Oddly, Cycle 3b was both popular and unpopular; moreover, the numbers were low so I decided to regard the 3b data as an outlier. I put appropriate weight on the two significantly negative cycles Cycles 5b and Cycles 9. Putting the spotlight on the significantly negative Cycles should help shed more light on the efficacy of the coursebook, and what improvements should be made. What caused these two Cycles to be the least popular?

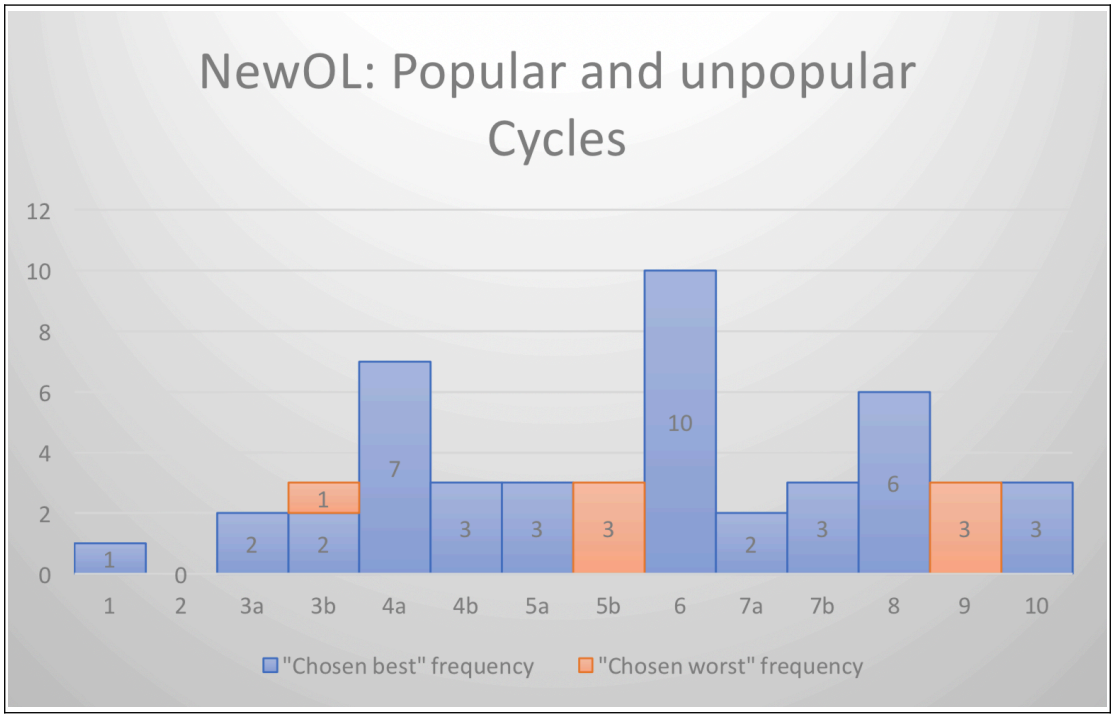


Figure 77: Popular vs. unpopular cycles

The six focus groups' determined "worst sections" have been tabulated with percentages in Table 36. The least popular sections were Cycle 5b (42.9%) and Cycle 9 (42.9%). Cycle 3b (the outlier) was a distant third (14.3%).

*Table 36: Focus groups' chosen worst cycles' frequencies *Highest are highlighted*

Cycle	Frequency "Chosen worst"	Percentage
1	0	0.0%
2	0	0.0%
3a	0	0.0%
***3b**	**1**	***14.3%**
4a	0	0.0%
4b	0	0.0%
5a	0	0.0%
***5b**	**3**	***42.9%**
6	0	0.0%
7a	0	0.0%
7b	0	0.0%
8	0	0.0%
***9**	**3**	***42.9%**
10	0	0.0%

When viewed in this way, the negative data is salient. What are the causes of this very clear unpopularity? The following two quotes help elucidate the students' attitudes toward these Cycles:

> *"We do not have enough time for this—we just rush and write them…"* (Focus Group 4; 3rd meeting; June 2014)

> *"…I used only about half of the templates—one group didn't use any from 4b."* (Focus Group 6; 1st meeting; May 2014)

The problems with Cycle 5b

The problems with Cycle 5b (Figure 78) are multifaceted—likely because Cycle 5b was based on what I must admit was a "half-baked" idea. This cycle was intended to be a secondary page to be used after that feigned ad-libbed "magical" talk from the teacher (Cycle 4b). On this *additional* page, the students were supposed to write down similar sentences (Figure 78) for their own benefit. The major issue seems to be, in retrospect, that they had *already* taken notes regarding possible grammar points for their upcoming presentations on that "magical" Cycle 4b page. So, while from the teacher's perspective this additional page may seem to be useful repetition and a place for further growth with grammar points via extra writing practice, from the students' perceptive, it was just a simplistic and uninspiring redundancy. This cycle was a redundancy for the students specifically because there was no novelty embedded in the task, nor any inspiring autonomy built into it. Hindsight is 20/20—and this cycle in hindsight is certainty a "dud" of a cycle. This cycle's content is not on par with the theoretical integrity of the adjacent cycles in the coursebook. The timing (chronological placement of the task) is also not on par with that of the rest of the book. From this, I have learned that the timing of the cycles in *OL2* can be just as important as the content of the cycles –and that ideally, the content and the timing need to be worked "just right" for the course to realize its full potential at teaching content. Below I have focused my thoughts on the issues of

timing and lack of in-depth thinking.

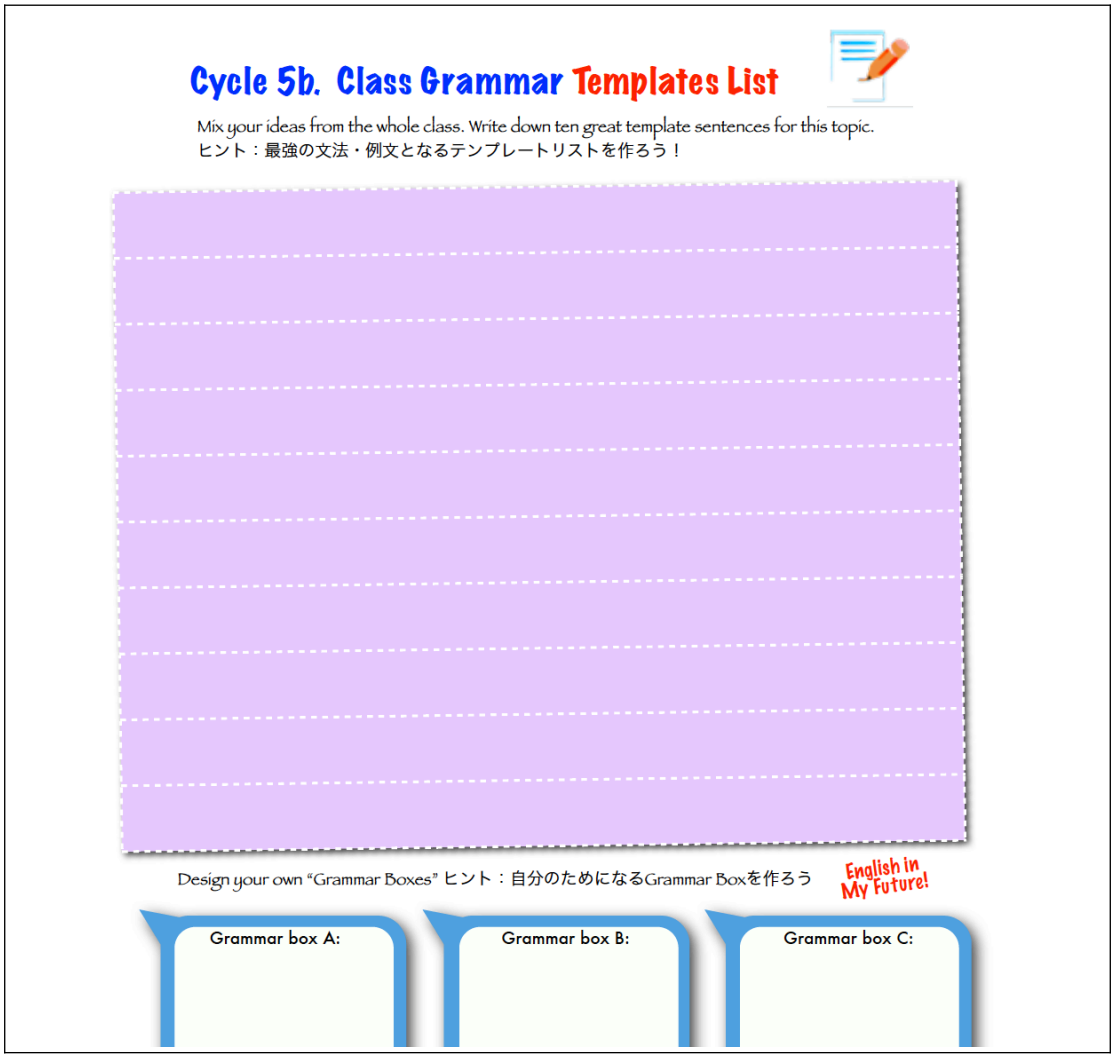


Figure 78: The unpopular Cycle 5b

Timing: Firstly, this even before we get into the chronology of the cycle, I must say that this cycle was also much too rushed. Cycle 5b is supposed to be building on the foundation of 4b, but the flow from Cycles 4b to 5b seems to be overly involved, and in the end, this is not something that can be completed at the coursebook writer's intended speed. Therefore, the students seem to be only making tenuous connections. This reduces the task to being a superficial activity. The superficiality and hints of bad timing were clearly illustrated by the above quotes; the quotes clarify where this internal integrity is lacking in the flow of *OL2*.

Lack of in-depth thinking: A staple of my coursebooks (be it *OL1* or *OL2*) is in-depth thinking in English (their L2). However, in retrospect, it is evident that Cycles 4b and 5b cannot be thoughtfully connected within the given timeframe and without proper support. As a natural consequence, in-depth thinking was lacking. Fascinatingly, the most diligent students (who were *thoroughly* engaged with the goals of the Module and the course itself) created a pragmatic compensation strategy for this shortcoming of the coursebook. They chose to ignore or play-down the importance of Cycle 5b, making it one of the most ignored/irrelevant cycles for them. It is fascinating, and educational for me, that it is was the most diligent students who used this strategy to move forward—their diligence toward the set goals allowed them to—on some level—regard this Cycle as an impediment to the goal instead of being good sustenance toward the goal, and therefore they circumvented it. (This specific issue in *OL2* was rectified, as discussed below, and in Chapter 9.)

The problems with Cycle 9

Cycle 9 is an assessment page that is used to assess each groups' group-designed test. This cycle was unpopular for two interconnected reasons: (1) aesthetic reasons, and (2) lack of student integrity toward the material. Regarding the aesthetic issue, because of a dark purple color used on this page, the students were not able to easily see their own penciled-in circles that they wrote down during the assessment task, making the task far more difficult than useful. As for the student integrity issue:

> *"…the class chose a bad test. Maybe they were not following the task (taking scores properly) on this page."* (Focus Group 5; 2nd meeting; May 2014)

Here I assume that the bad color choice for this page in the book was an obstruction for the students, whether they were explicitly aware of it or not. A context in which it is hard to read your own writing must have led to a lowering of

motivation to stay on task at some level in many of the students. This is a plausible flow of events—which is nuanced by the quote, albeit in not so many words. The bad color choice was rectified in the next version, as is discussed in Chapter 9.

8.5 Written student appraisals at the end of the first semester

Students in all seven of the courses were given blank papers to write open-ended paragraphs for reflection of their course—as appraisals of the course. This section shall examine the emerging themes from those reflections. Here I have provided the-then newly established theme table (Table 37). To make meta-assessment possible, the-then newly emerging themes were added to the previously established coding system from the focus group discussions.

Table 37: Assessment paragraphs' emerging themes

A1. English skills improved
A2. English skills didn't improve
B3. Socio-cultural skills enhanced
B4. Socio-cultural skills lacking
B5. Presentation skills improved
B6. Presentation skills lacking
B7. Leadership/management skills improved
B8. Leadership/management skills lacking
B9. Logical thinking improved
B10. Logical thinking lacking
C11. Motivation/affect improved
C12. Motivation/affect lacking
C13. Metacognition/deeper thinking developed
C14. Metacognition/deeper thinking lacking
D15. Book design, positive
D16. Book design, negative
D17. Course, positive
D18. Course, negative

This data collection method was significantly different from the focus group's data collection method: (1) They were hand-written, (2) they were not openly

discussed in groups, and (3) because there was no discussion, students were not able to influence each other's thinking. `

In Dynamic Skill Theory terms (Fischer, 1980), this context would be deemed a *low support context* (see Section 5.3.4). For contrast, the elaborate discussions that ensued with the focus groups is a *high support context*. High support contexts invariably produce more in-depth responses to inquiries compared to responses received from low support contexts. I therefore expected that these open-ended paragraphs would provide significantly different results (responses with less detail, less depth, and less complexity) compared to the focus group data. However, as can be seen in Table 38 (the table of first semester themes from students' paragraph length commentaries about the semester), the *top five* frequency themes are the same as the focus groups', with only the slot numbers shuffled. This outcome was a genuine surprise and it is highly noteworthy.

Table 38: The positive themes, their frequencies, and illustrative examples (first semester paragraphs)

OL2 positive themes (paragraphs)	Frequency	Illustrative example
B5 Presentation skills improved	136	"Because we did so many presentations, I feel more comfortable speaking in front of people. Thank you very much!" (Sem1, 001, 2014) …I was very glad to be able to talk about them in this class. Having the freedom to speak about myself…was an enjoyable pedagogy for me. I think the presentations were effective too. Thank you very much!" (Sem1, 002, 2014) "I was able to come up with new vocabulary and fit them into my presentation -- and it was fun doing this." (Sem1, 003, 2014) "Unlike this class, the majority of the English classes that I take are about memorizing grammar. But as for this class, having real English usage for presentation raises our speaking ability and our presentation ability very well." (Sem1, 012, 2014)
C11 Motivation/ affect improved	115	"I have never taken such a test with so much freedom in the design. I feel it is fun to be able to write freely." (Sem1, 004, 2014) "We still have some pages left in the textbook. I'd like to complete them. Thank you for this half year". (Sem1, 013, 2014) "I enjoyed engaging in a single theme from multiple angles and multiple activities to come up with new ideas. I now feel like doing not only things that are close to me, but presenting on social problems as well." (Sem1, 014, 2014)

C13 Metacognition/ deeper thinking developed	66	“Thank you for such a new teaching style. I've realized that through long-term stewing a single theme, we can gain new knowledge and deeper thinking. While stewing the theme of Justice for my presentation, I went to the library and borrowed books by Michael Sandel and then asked around my group of friends what their thoughts were. It was a fantastic stewing opportunity. I look up to people with lots of knowledge and people who have interests in a wide variety of fields. From now on I won't just "speak English", I will try to use my English to exchange ideas with foreigners so that I can stew my ideas with them!! Thank you very much.” (Sem1, 005, 2014) “I normally don't think of things very deeply. This class made me think deeply about things that I used to take for granted.” (Sem1, 006, 2014) “I typically do not think about my family or hobbies very deeply, so this class was an eye-opener. In enjoyed unearthing my buried memories. Thank you for this semester!!” (Sem1, 007, 2014)
D15 Book design, positive	51	“I believe this book (course) was the ultimate way to raise personal English expression levels.” (Sem1, 008, 2014) “Reading the forward in your book, I realized that you are using pedagogy that I recently studied about in my Pedagogic Theory class and thought it was incredible that you had implemented the pedagogic theory into your own textbooks.” (Sem1, 009, 2014) "It makes me realize the difficulty of putting my thoughts into English, but at the same time, I felt how fun English could be. Your book was unique and made this class enjoyable. I love this class!” (Sem1, 017, 2014) “I was able to realize my mistakes in spelling, grammar, and other areas because we worked on a single theme from many directions.” (Sem1, 021, 2014)
A1 English skills improved	45	“This class was very fun, because we can choose our presentation topic very freely. Thanks to this, it is easy to talk in English.” (Sem1, 010, 2014)
D17 Course, positive	26	“Up until now, English classes have been focused on Writing and Reading. However, in this class we get to think and talk about ourselves in English -- in front of everyone. This was a very new pedagogy for me and it was enjoyable! I'm not that good at public speaking, and I was therefore very nervous, but it served as a good training for my future. Moreover, I found it interesting to listen to other people's ideas, their likes... I learned new things about them, and it was enjoyable.” (Sem1, 011, 2014) “Rather than the typical "listening mode" we are usually put in here in Japan, your style has us think for ourselves, and therefore it was fun! Thank you very much.” (Sem1, 015, 2014)

The non-tabulated data is: Logical thinking improved (3), Socio-cultural skills enhanced (1), and Leadership/management skills improved (0). The frequencies are in the adjacent brackets, respectively.

Having identical top themes between the focus groups’ discussions and the

entire student body's assessment in paragraph form was an entirely unexpected finding—because the context, modality, and genre of data collection were all completely different. This was therefore a substantial discovery. In sum, the most attention-grabbing points from this section are:

- The top slot is now a *non*-abstract theme; it is performance-based ("Presentation skills")
- "Motivation" and "Metacognition" themes are still in the top three
- "English skills improved" is now at slot 5 instead of 4 (but still in the top five)
- And, all top five themes from the paragraphs are identical to the focus group's top five

Although the five themes were identical, they were of different frequencies. The student body collectively put a concrete theme at the top. It may be deduced that when left up to their solo-thoughts (without any priming from the teacher), the majority of the students tend to focus more on what is non-abstract and what is visibly apparent ("Presentation skills improved", N=136). What was most surprising was that slots 2 and 3 were of the *non*-tangible sort metacognitive abstractions ("Motivation/affect improved", N=115); "Metacognition/deeper thinking developed", N=66). Two slots down from those themes we finally find "English skills improved." As with the focus group data, this seemingly *most-important-theme* is not in the top three. A pattern is forming with these findings. What is the significance of this emerging pattern?

From my own long-term work (spanning over a decade) with high support contexts (Murphy 2008, 2014), and seminal works such as Fischer's (1980), it is clear that high support contexts *nurture* deeper thinking processes in the brain. Students' demonstrable thinking is often *much* more profound when they are put in these nurturing high support contexts (as opposed to when stuck in low support contexts); within Dynamic Skill Theory research, this is a globally consistent

finding. My own research in this area was primarily conducted in Japanese EFL contexts (Murphy, 2009); those results were found to be consistent with published results from South Korea, Taiwan, and the United States.

As so, at first glance the *OL2* data discussed above appears to be utterly incongruous with the previously published data. This "incongruity" does have a logical explanation—in short, it is not an incongruity. Nevertheless, the findings were unexpected, and the findings became a prodigious point of interest in this doctoral project. I shall discuss the rationale behind these claims in depth within the next section.

8.5.1 Positive themes

Let us have a closer look at the whole range of positive themes (listed in Table 38). For this task, the students provided responses to an open-ended question regarding appraisal of the course, anonymously. Because it was conducted anonymously, the students had no pressing reason to adjust their commentaries. I should also emphasize that the students themselves popularly reported that, "Metacognition/deeper thinking" improved during the course (N=66). This was surprising when first uncovered, but, by the time of this writing, these findings have collectively transformed into a primary (recurring) theme for this study.

Evidently, and what was highly intriguing was that the "un-primed" students—the students whom had not attended any of the focus group meetings nor had ever had any specific discussion time regarding these pedagogical concepts—seemed to have been *altered* in a way that was conducive of higher-level thinking regarding: (a) the affect of the pedagogy, (b) their own L2 growth rates, and (c) the connection between the pedagogy and their L2 growth rates. Collectively, the "un-primed" students demonstrated a level of thinking that was on par with the focus groups.

Upon deliberation, my hypothesis (based on Fischer's work) was that long-term *OL2* coursebook usage may be nurturing an affect on the users that lasts long enough to create a good amount of automaticity for the students to spontaneously

reply to these questions at levels of depth that are indistinguishable from "just-now-primed" high support responses.

In sum, there never was an incongruity with the data and past dynamic skill theory studies; it was the coursebook usage that had nurtured them to develop automaticity skills that are necessary to be able to quickly slip into a high support response mode. This is, therefore, one of the most significant findings of this study

Due to its methodical and unwavering cyclical design, long-term usage of *OL2* seems to be able to produce a condition within the students (nurturing a *purpose-designed* network of skills) that enables them to quickly demonstrate in-depth thinking in their L2. Fisher calls this biological phenomenon a *cognitive reorganization* (2008).

Fisher does not use this term lightly; he reserves usage of "*cognitive reorganization"* to denote (context-influenced) massive changes in a subject's neuronal configurations allowing for deeper level thinking that was not previously possible for the individual. "According to the network-growth hypothesis, the changes in energy…arise from developmental changes in neural networks in the brain, developments that come about through a cyclical process of rewiring and retuning networks" (2008:139). Fischer goes on to explain, "According to the hypothesis, the network cycle corresponds to periods when particular types of learning and developmental changes occur, such as spurts in a major skill" (2008:140). Although I was versed in Fischer's work in this area, I had not expected one-to-two semesters of *OL2* to be able to produce an affect of this magnitude. And so, Fischer's model of "cognitive reorganization" was not on my mind while conducting this particular area of research. And, this is why I was puzzled by the quality and quantity of the high support responses that I was receiving from what I had assumed to be low support contexts.

This explanation (the acknowledgment that these students' results are representative of *cognitive reorganization*) goes far in helping answer questions that were raised in Chapter 7, and is compatible with its findings. From the perspective of Chapter 7 (particularly the *Microdevelopment* discussion), a differently worded

explanation would be that the persistent cyclical design of *OL2* had biased the students enough to be able to effectively "rise" to the called upon task; the students were able to leave the novice stage behind and move on to a more stable level of displayed understanding. The gists of these differently-worded explanations are complementary. They both point to the conclusion that *OL2* somehow embodies a type of high-efficiency procedure that leads to a rapid and stable display of L2 Maturation.

Table 38 (p.298) elucidates hints regarding the potential affect of high-efficiency teaching and learning. The illustrative quotes in the "Metacognition/ deeper thinking developed" are particularly supportive of the hypotheses regarding the occurrences of affective contexts leading to high-efficiency learning. Quotes from the adjacent themes also overlap and strengthen this idea of efficiency and the foundation for high-level thinking (particularly Fischer's notion of *cognitive reorganization)*. I say this because the quotes show that these results are recognizable to the students, *and* that is these higher-level skills gradually become second-nature when required by the students' context. This qualitative data (along with the quantitative data), therefore helps define the properties of—and the affect of—*OL2's* design.

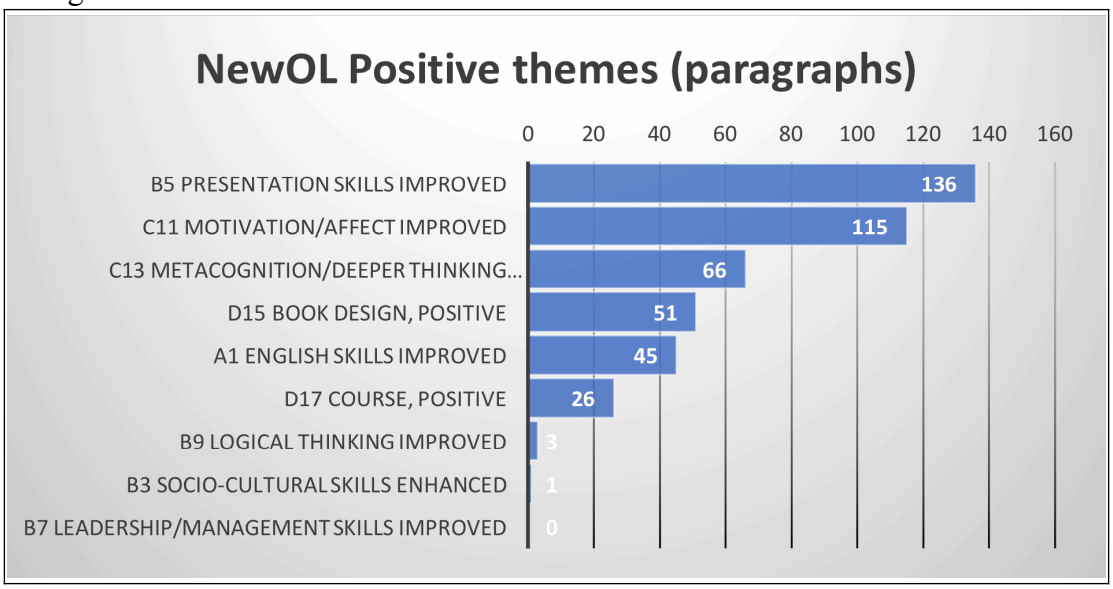


Figure 79: The frequency of the positive themes from the assessment paragraphs, in graphic form.

The data in graphic form (Figure 79), provides visual clues regarding the occurrence of these themes. The top two slots—the themes regarding *Motivation being improved*, and *Metacognition/Deeper thinking*—are likely to be naturally reciprocal. One would probably not be noticed (nor exist) without the other; they seem to have a symbiotic relationship. Moreover, these two themes are also prominent in the focus group frequency list (helping fortify the validity of this data collection).

For the *OL2* courses, courses that rotate around student-designed presentations and student-designed tests, it would also seem that the themes "Presentation skills improved" and "Motivation improved" also have a sort of symbiotic relationship. Viewed in this way, it becomes understandable why the top five positive themes were identical across the focus group list and the paragraphs list—even though those data collection modalities were vastly different.

Perhaps more importantly, when flipped around, it can be said that the five themes that emerged can be flagged as prerequisites for high-efficiency EFL coursebook design. (If such a list were to be used to establish EFL coursebook writing protocols, the top slot would need to be transferrable with whatever primary task satisfies the curricular needs of that particular course—but all else would seem to function well as a template for protocol production.)

8.5.2 Negative themes

Table 39 provides the tabulated negative themes from the first semester paragraphs, along with their frequencies, and illustrative examples from the paragraphs. Almost like a twin of the focus groups' list, there was also only *one* significantly larger frequency number at the top. Moreover, that top theme was the *same theme* "Book design, negative." This data, therefore, also strengthens the hypothesis that at least some of the students taking the *OL2* course may have indeed been in a cyclical state of growth within a context that was able to nurture skills of high support context caliber. With some of these students, this seemed to be happening to such a degree that even when given simplistic low support tasks such

as this one, they went on to produce output at levels on par with high support-based output.

Table 39: The negative themes, their frequencies, and illustrative examples (first semester paragraphs)

OL2 negative themes (paragraphs)	Frequency	Illustrative example
D16 Book design, negative	13	"…class was fun, but I would have liked it to be a bit more rigorous. I think after a simple theme in the beginning we could move on to more difficult themes such as ecological problems. I would have like that challenge. I would have also liked to hear more concrete points on how to make better presentations." (Sem1, 020, 2014) "It is possible to do presentations without doing the exercises in the textbook so I didn't see the purpose of having the book. But I liked your class because it was fun." (Sem1, 016, 2014) "I wish we could have used peer-to-peer assessment in more ways." (Sem1, 018, 2014)
A2 English skills didn't improve	6	"I would have liked to have seen more details in the assessment (more good points and bad points)." (Sem1, 017, 2014) "This class is so hard for me..." (Sem1, 019, 2014)

The non-tabulated negative themes were: Presentation skills lacking (3), Metacognition/ deeper thinking lacking (2), Socio-cultural skills lacking (0), and Leadership/management skills lacking (0). Focusing on the zero-frequency negative themes—it is important to note that none of the students reported that they felt that socio-cultural skills development was lacking, nor that their development of leadership/management skills was lacking. As with the previous section's data, I would like to emphasize that zero-frequency data on negative themes is not equivalent to having positive data regarding the positive version of the theme; they should be considered separately as I have done here.

To end this section, below is the side-by-side view (Figure 80) that displays how the positive (red) bars' frequencies far outweigh the negative theme frequencies. It also shows where the zero-frequency themes exist.

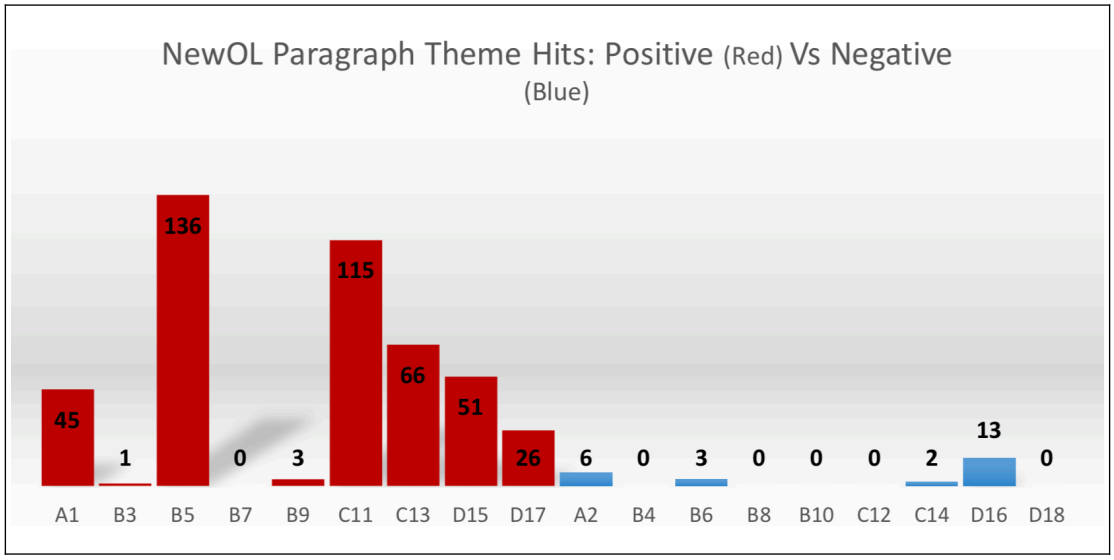


Figure 80: Side-by-side of positive themes and negative themes that emerged from the paragraphs.

8.6 Written student appraisals at the end of the year

Table 40 is the list of established themes from the students' final essays, received at the end of the second semester. This data is significant because, by the time the students had written these essays, they had completed *two* 15-week *OL2*-based courses, over the span of a full school year. They were now intimately familiar with the pedagogical flow and cycles of the coursebooks. How much have their insights changed/developed? Here I shall examine the results. For compatibility and alignment purposes, the emerging theme list format for the second semester is identical to the first semester's format.

Table 40: Established themes list from end of year final essays

A1. English skills improved
A2. English skills didn't improve
B3. Socio-cultural skills enhanced
B4. Socio-cultural skills lacking
B5. Presentation skills improved
B6. Presentation skills lacking
B7. Leadership/management skills improved
B8. Leadership/management skills lacking
B9. Logical thinking improved
B10. Logical thinking lacking
C11. Motivation/affect improved
C12. Motivation/affect lacking
C13. Metacognition/deeper thinking developed
C14. Metacognition/deeper thinking lacking
D15. Book design, positive
D16. Book design, negative
D17. Course, positive
D18. Course, negative

8.6.1 The positive themes

Table 41 lists the frequency of the positive themes in the second semester and the end of year essays, along with illustrative examples. Here we see a considerable deviation from first semester themes. The theme, "English skills improved," which was always lower in the list during the first semester (4th slot for the focus groups; 5th slot for the assessment essays), was now the number one frequency theme. After that, the list is almost identical to the first-semester frequency list.

Table 41: The positive themes, their frequencies, and illustrative examples (second semester essays)

OL2 positive themes (final essays)	Frequency	Illustrative example
A1 English skills improved	134	"I think my speaking skills went up after presenting in English… I think presentations are good for listening, speaking, and writing." (Sem2, 101, 2015) "These presentations helped me to progress in my English writing and speaking skills. It gave me a chance to speak in the presence of others." (Sem2, 102, 2015) "(because of this class) I can use English when I want to use it now. So, I think this class was a very nice experience to study English more than (previous classes)." (Sem2, 103, 2015)
D15 Book design, positive	129	"(because of the book's design) I much think of the theme in various viewpoints. And find various sub-themes to the main theme. Secondly, I should use more kinds of words… so we can think of each idea more deeply. I think some of these things come together to make my presentation better." (Sem2, 104, 2015) "(regarding the book) … talking to everyone is one of the important things in our life, I think. I feel pleasure because I get these skills through this (book's process) in this class. In addition, I could design the passages of the presentation. So, I also get design skills!! It's very fun." (Sem2, 105, 2015)
C13 Metacognition /deeper thinking developed	88	"I think deeply due to presentations. I don't think deeply about English usually. So, (this class) is a good chance for me to think more deeply. Also, I like listening to others' presentations. I can get very nice and a lot of ideas. Other' presentations are very interesting." (Sem2, 108, 2015) "I think that the power to write sticks to the body while preparing for presentations. I feel it because this does not happen when we just write." (Sem2, 109, 2015) "…and it is important to say one's own opinions in presentations." (Sem2, 110, 2015) "Maybe I don't give the best advice, but due to many practices, I have gained confidence, and I may be able to be consulted now." (Sem2, 111, 2015) "I think the most important thing is thinking. Through (this class's process) I can deeply think. And, when I listen to my friends' presentations, I make an effort to understand the content." (Sem2, 112, 2015)

C11 Motivation/ affect improved	91	"At first I was shy to stand in public. But now it is a pleasure for me… and to watch everyone's faces." (Sem2, 106, 2015) "But, I spoke in front of everyone any number of times, so I'm not nervous anymore! I can tell my ideas to everyone in English. I became to like English a little." (Sem2, 107, 2015)
B5 Presentation skills improved	84	"At first I took much time and thought a lot. I now realize that clear talking is important too. My presentations became better!" (Sem2, 113, 2015) "Although my sentences may be still simple and childish, I can now write my presentations with the words and phrases that I learned (from the coursebook's activities)" (Sem2, 114, 2015)
D17 Course, positive	30	"I am the worst at public speaking. For me, this presentation course is a good opportunity for overcoming… English used in our class is easy to understand." (Sem2, 115, 2017) "(this course) is an effective was to study English. The presentations make me review English. The presentations are pleasant…" (Sem2, 116, 2015)

The non-tabulated positive themes were: Logical thinking improved (6), Socio-cultural skills enhanced (2), and Leadership/management skill improved (0).

Regarding the elements to be drawn from the positive themes—at least four details can be deduced from the movement seen in this frequency list. Firstly, the top five themes have gone relatively unchanged all year, regardless of the modality, and the time of year—this further strengthens the trustworthiness of the data collection strategies implemented for this study, as well as the gamut of hypotheses discussed in the above sections.

Secondly, the themes regarding "Motivation" and "metacognition" were always in the top five of each list, no matter what the modality was—again, it can be deduced that the students seemed to realize on some level that they were learning from this course in a somewhat novel way that maintained their motivation for continued in-depth thinking of the presentation content. The relationship between these two themes appears to be logically intertwined and perhaps even symbiotic in nature.

Thirdly, it seems that the design of *OL2* did not prompt much student burnout; conversely, *OL2* seems to have been able to sustain and perhaps even

support a continued flow of interest that helped nurture an internal desire to maintain the flow of their own skill improvement. This is a multi-layered phenomenon; the students' accumulative skill-building included their expanding abilities to meta-cognitively notice *their* own cognitive growth, as it was happening in realtime, and during their moments of retrospection—these are different skill sets that can later gain practical parity that would likely produce *aha moments* as they form a bigger picture for the students. As discussed above, a fitting image for the cyclically embedded complexity of the multiple levels of motivation and learning being described here is the Mandelbrot.

And fourthly, as the end of the year was upon them, the students seemed to (finally) realize to what degree their own English capabilities had improved over the course of the year; this deduction stems from the fact that the theme "English skills improved" suddenly climbed to the top of the frequency slot—after always being in the 4-5 place in the frequency list until this point in time (the very end of the year).

8.6.2 The negative themes

Table 42: The negative themes, their frequencies, and illustrative examples (second semester essays)

OL2 negative themes (final essays)	Frequency	Illustrative example
A2 English skills didn't improve	5	"I didn't go to the trouble of using difficult vocabulary and grammar." (Sem2, 117, 2015) "I still can't communicate with each other…" (Sem2, 118, 2015)
C14 Metacognition/ deeper thinking lacking	5	"I use words and expressions as plain as possible." (Sem2, 119, 2015)
D16 Book design, negative	5	"When I was a high school student, I was writing a lot of English. But, I do not do a lot of writing after I became a university student (in this class now)" (Sem2, 120, 2014)

The non-tabulated negative themes were: Presentation skills lacking (3), Socio-cultural skills lacking (0), Leadership/management skills lacking (0), logical thinking lacking (0), Motivation/affect lacking (0), and Course, negative (0).

The negative frequencies are so low at this point (Figure 81 and Table 42) that although I have not entirely ignored them, it would likely be more suitable to observe them as outliers—especially considering their paradoxical position in comparison to the vast majority of the data. The outliers were "English skills didn't improve" and "Metacognition/deeper thinking lacking". Considering the quantity and quality of the positive data that came before this, perhaps an apt summation for this data is, "No matter how hard you try, you simply cannot please everyone." That said, I have explored the possible rationales behind these responses, and have provided my conclusions below.

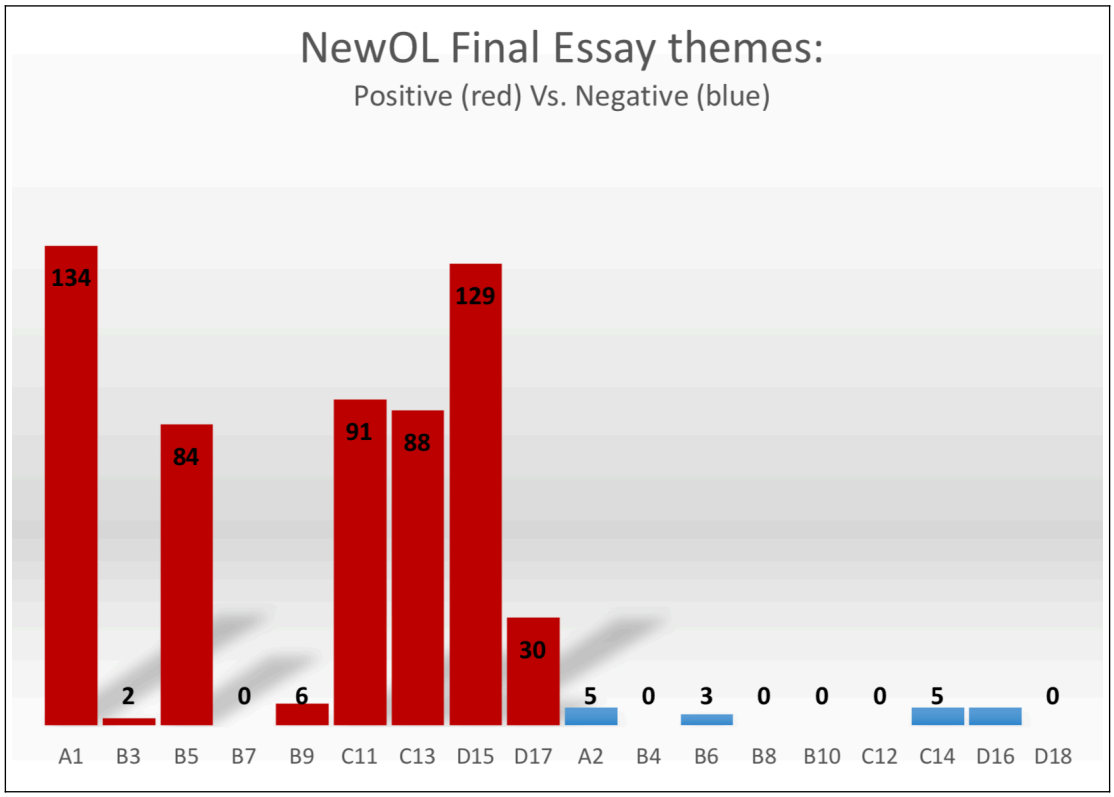


Figure 81: Side-by-side of positive themes and negative themes that emerged from the final essays.

The students who raised those negative comments were mostly students who missed a higher number of class hours than their peers (they averaged three to five missed classes, and/or were significantly tardy on several occasions). They were not particularly "bad students"—but they were apathetic students. Their marks did

indirectly suffer from missing the classes because they did not participate much in class—they suffered much from their lack of group work hours. Because of this, they missed out on some of the positive psychological factors of the course (factors such as *student-bonding* and *mutual encouragement* that occurred across the majority of the student body). Naturally, their portions of their group presentations were not of a high standard either. The way *OL2* is designed to encourage group work and mutual growth, it is not easy for truant students to catch up with peers who have already bonded, and are working together enthusiastically on their various tasks. It can be said that these "truant" students (those whom provided the negative remarks) were being quite truthful when they reported that their English skills did not improve. This data highlights a downside of the *OL2* pedagogical strategy because the outcomes of the course are wholly dependent upon how well the cycles establish L2 Maturity (inclusive of peer-to-peer nurturing and the growth of spontaneous language usage), if a student outright falls out of the loop due to truancy or perhaps circumstances such as a hospitalization, it may be difficult or impossible for that student to catch up with their peers during that particular term. Such students may have to re-start their cycle of growth with a newer group of students in an upcoming term.

Returning to the main point of this section (Positive vs Negative theme frequencies), the side-by-side bar graph (Figure 81) puts this data into visual perspective; by the end of the year, the positive theme frequencies greatly outweigh the negative themes. The blue bar themes were provided by the students with the least amount of classroom participation.

8.7 A composite of all the data sets from the primary 2014-2015 study

Although the modalities differed, the theme coding system was kept consistent across the modalities. It was therefore possible to compile the data directly into a single data set and make deductions from the compilation. Figure 82 is a stacked bar graph of non-adjusted data from the positive theme frequency data sets.

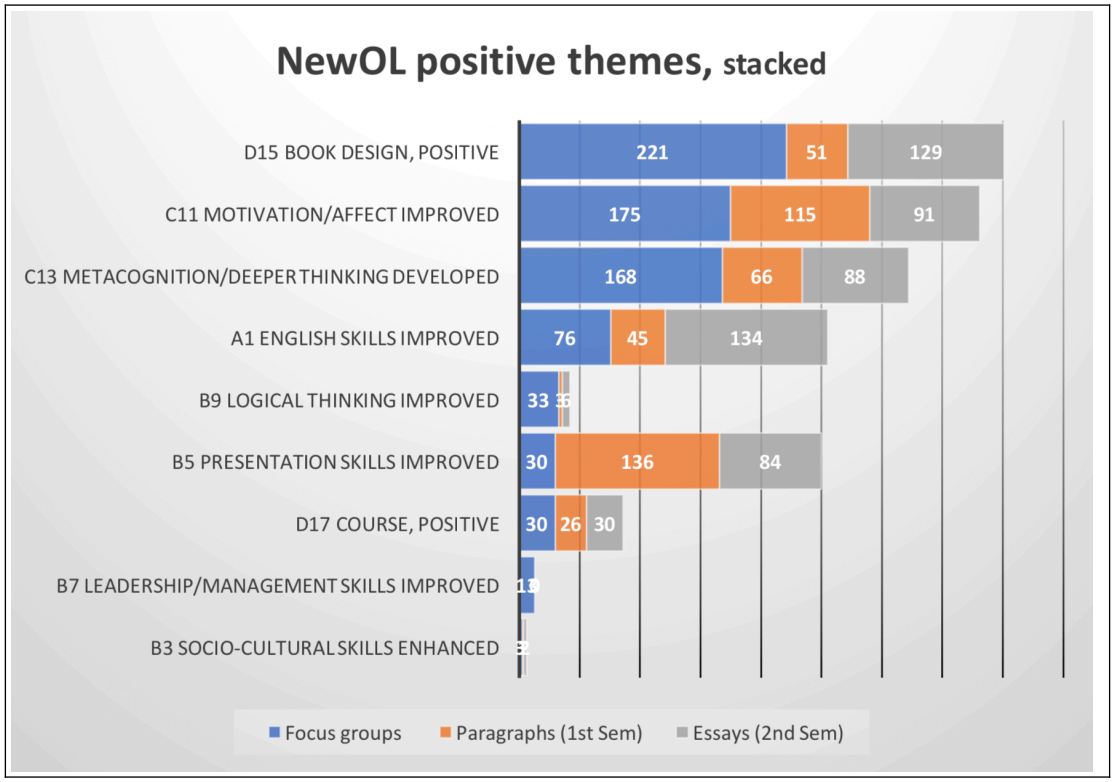


Figure 82: Semester one and semester two, positive themes stacked in graphic form.

We can see the top positive themes in each of the three modalities in Figure 82. I have color coded them for easy reference. Viewing the data from the left (focus group list), at the top of the graphic list is Book design, positive (221). The adjacent bracketed number is the frequency. This theme remains in the top five themes, across the board. The same can be said for Motivation/affect improved, and Metacognition/ deeper thinking developed. Two prominent deductions were made from the compilation of positive theme data:

"Presentation skills improved" showed high frequency at the end of semester one with the assessment paragraphs (number one slot; *freq*.136), and remained a high-frequency theme throughout the year. Similarly, the "Motivation" and "Metacognition" themes were also consistently at the top—likely due to the combination of: *OL2*'s autonomous ("ownership" feeling inducing) student-centered, workbook-like design, and the continued peer-to-peer assessments and student-designed test designing activities.

The most significant change by the end of the year was "English skills improved"—this moved up to the number one slot in the frequency list. From the students' perspective, their L2 skills had improved; they were regularly self-assessing themselves. This is a reassurance that the *OL2* design was not only about increasing in-depth cognition—but also about the realization of English language improvement with this concurrent and interrelated expansion of deeper thinking skills.

There are two caveats: (a) theme frequency counts should not be construed as indicators of intensity of that theme, nor should they nuance statistical significance. They do, however, directly reflect how prominent the themes were in the discussions and in the students' writing, so this form of quantification of qualitative data does have merit. And therefore, (b) the compiled data set in Figure 78, being a raw data set of frequencies of multi-modality, is not intended to depict statistical significance. Rather, Figure 78 is supplied to provide a visual perspective of the raw data. Moreover, quantitative data and statistical analysis (using different methods) have already been provided and discussed in Chapter 7.

Regarding the negative themes, the raw data, side-by-side, is provided in Table 43. It is noteworthy that as the year progressed (left to right), the negative theme frequencies all declined, except for one theme. The most prominent change is in "Book design, negative" (N=15 → N= 5).

*Table 43: First and second semester, negative themes side-by-side. (*Largest change)*

OL2 negative themes (both semesters)	Focus groups	Paragraphs (1st Sem)	Essays (2nd Sem)
A2 English skills didn't improve	1	6	5
B4 Socio-cultural skills lacking	0	0	0
B6 Presentation skills lacking	4	3	3
B8 Leadership/management skills lacking	0	0	0
B10 Logical thinking lacking	1	0	0
C12 Motivation/affect lacking	6	0	0
C14 Metacognition/deeper thinking lacking	2	2	5
***D16 Book design, negative**	**32**	**13**	**5**
D18. Course, negative	7	0	0

From this data it appears safe to claim that during their studies during their second semester, student attitudes changed regarding the book design. Since this change correlates with the sharp rise in frequency of "English skills improved," as discussed above, perhaps this is not a mystery—the simple deduction is that the students' attitudes toward the book changed as they realized the benefits of the book as they began to realize their own language improvement (be it their own improvement, or their peers' improvement). Chapter 9 will feature prominent exploration of these book design issues and noteworthy learning points from this study.

8.8 Student appraisal of OL1

To better understand the efficacy differences between *OL1* and *OL2*, it is necessary to review and compare the design of *OL1* at the cycle level as well. Note that while the sections within each chapter of *OL2* were deliberately and explicitly referred to as "Cycles," this was a deviation from *OL1's* design. In *OL1*, the chapter sections were more simplistically referred to as "Sections." Table 44 presents the "chosen best" Sections in *OL1*.

8.8.1 Student evaluation of the different sections

An interesting recurring phenomenon from the comparison of *OL1* and *OL2* is that the amount of data that is produced by the *OL1* group is consistently less than that of the *OL2* group. The *OL1* group were certainly not a grouping of "bad students" (they did well in their overall marks). Yet, in comparison with *OL2* students, it seems that *OL1* students had less *willingness* to express their thoughts. The frequency data below (Table 44) supports this observation.

Table 44: The "chosen best" Sections in OL1.

Section	Frequency "Chosen best"	Percentage
1	2	33.3%
2	1	16.7%
3	1	16.7%
4	0	0.0%
5	0	0.0%
6	1	16.7%
7	1	16.7%

Because the numbers are low, it is hard to make more substantial claims. However, the zeros in Section 4 and Section 5 do stand out. Moreover, this is fully congruent with all other data pointing to the conclusion that the middle of each chapter in *OL1* is the low point of the chapter.

Table 45: The "chosen worst" sections in OL1.

Section	Frequency "Chosen worst"	Percentage
1	1	33.3%
2	0	0.0%
3	0	0.0%
4	1	33.3%
5	0	0.0%
6	1	33.3%
7	0	0.0%

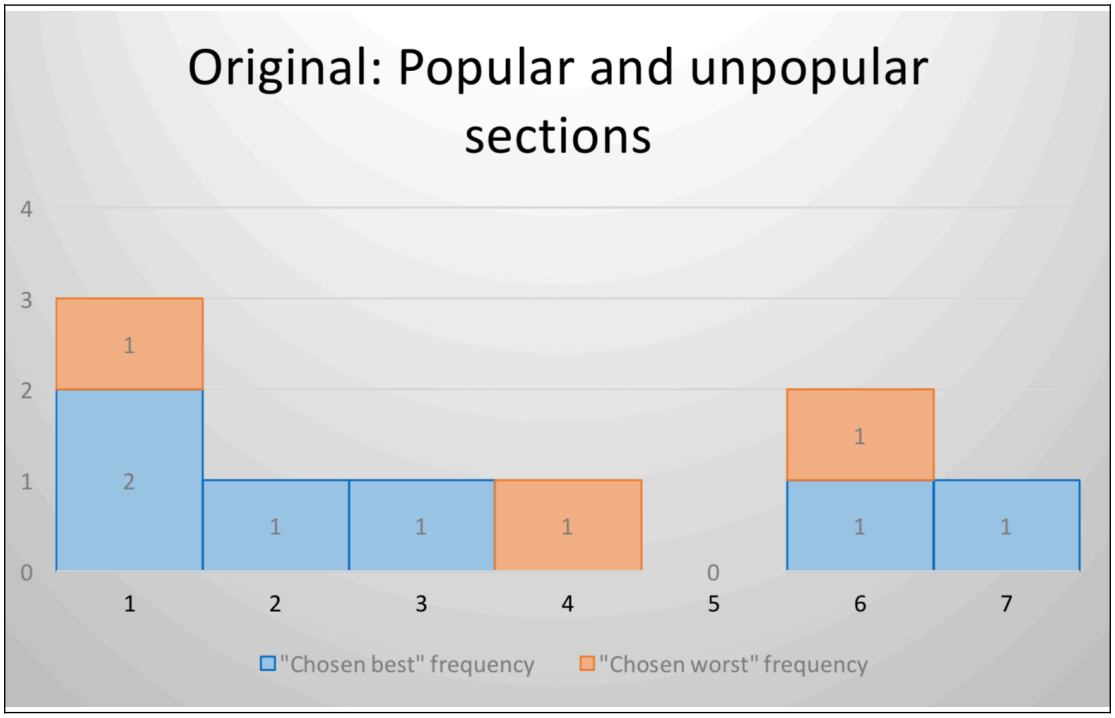


Figure 83: Popular vs. unpopular cycles

Figure 83 is a stacked bar graph version of the popular vs. unpopular sections' data (Table 44 and Table 45). I have stacked them here to show discrepancies and/or to see if there can be any logical visually-based deductions. At first glance, there was little to deduce here; the data seems to be incongruent. However, even considering the incongruences, apparently Section 4 and Section 5 are the low points of the chapter; the chapter begins well-enough, then it gets boring for the students in the middle, then slightly recovers by the end.

In essence, the *OL1* chapter created an inverted bell curve. Theoretically, this may denote a "critical break" (Murphy, 2017) in the flow of the chapter. A critical break is when there is a break in the learning flow from which the learning process' flow can never fully recover from. If a critical break has been designed into the middle of a chapter, it is catastrophic—this may be the most vital evidence to help explain the vast efficacy variance between *OL1* and *OL2*.

8.8.2 Data from the Focus Group

Table 46: Positive themes from the OL1 focus group sessions.

***OL1* positive themes (focus groups)**	**Frequency (high-low)**
A1 English skills improved	10
C11 Motivation/affect improved	7
B7 Leadership/management skills improved	3
D15 Book design, positive	3
B3 Socio-cultural skills enhanced	1
B5 Presentation skills improved	0
B9 Logical thinking improved	0
C13 Metacognition/deeper thinking developed	0
D17 Course, positive	0

As can be seen in Table 46 there were only a few positive themes from the *OL1* group, and they substantially deferred from the *OL2* groups' themes. The most prominent difference is that the "Metacognition" frequency is zero for the *OL1* group. This is by far the most alarming difference; in contrast, "Metacognition" talk was consistently of very high frequency with the *OL2* groups. In *OL1*, "English improved" was at the top in frequency, which sounds normal, but considering the discussion thus far, this discovery is seems rather lackluster; it seems as though *OL1* students are not thinking very deeply. This is not the students fault – it is likely a fault in the *OL1* design. "Leadership/management skills improved" is a new theme for this discourse. Also to great surprise, no student *ever* discussed "Presentation skills improved" although they all organized and accomplished three group presentations across the 15-week course (as did the *OL2* students).

The *OL1* students' presentation teacher evaluations do not indicate a lack of growth of presentation skills – on the contrary, the data shows improvement, as discussed in Section 7.4.1. And so, I must conclude that the *OL1* group could not demonstrate an understanding of their own skill improvement.

Although the frequencies were low, the reported negative themes support all of the above conclusions. From the top, "Motivation was lacking", "English skills didn't improve", and "Presentation skills lacking" (Table 47). Again, theme

frequency should not be construed as an indication of theme intensity or significance, so even though the frequencies may be low, the emergence of these themes should not be taken lightly.

Table 47: Negative themes from the OL1 focus group sessions.

***OL1* negative themes (paragraphs)**	**Frequency (high-low)**
C12. Motivation/affect lacking	6
A2. English skills didn't improve	2
B6. Presentation skills lacking	1
B4. Socio-cultural skills lacking	0
B8. Leadership/management skills lacking	0
B10. Logical thinking lacking	0
C14. Metacognition/deeper thinking lacking	0
D16. Book design, negative	0
D18. Course, negative	0

8.8.3 Written data

Considering the trends found in the *OL1* data production, as may be expected, the open-ended writing task from the *OL1* group produced less data than that of the *OL2* group (Table 48). Although the frequencies ae not very high, the content of the written data from the *OL1* group is congruent with the *OL1* data from above.

Table 48: Positive themes from the OL1 class paragraphs

***OL1* positive themes (paragraphs)**	**Frequency (high to low)**
A1 English skills improved	11
B9 Logical thinking improved	4
C11 Motivation/affect improved	4
B3 Socio-cultural skills enhanced	2
C13 Metacognition/deeper thinking developed	1
B5 Presentation skills improved	0
B7 Leadership/management skills improved	0
D15 Book design, positive	0
D17 Course, positive	0

Table 48 shows that perhaps the *OL1* students "appreciated" the class, but there is no evidence of a *rich experience* (such as expressed by the *OL2* groups). I hypothesize that almost any mediocre EFL coursebook could produce results similar to those in Table 48. In other words, while the *OL2* data showed there was something rich about the *OL2* teaching experience, the *OL1* data contrasts *OL2* data by indicating dull "generic" responses from the students.

Table 49: Negative themes from the OL1 class paragraphs

***OL1* negative themes (paragraphs)**	**Frequency (high to low)**
A2. English skills didn't improve	6
D18. Course, negative	4
B4. Socio-cultural skills lacking	3
B6. Presentation skills lacking	1
D16. Book design, negative	1
B10. Logical thinking lacking	0
B8. Leadership/management skills lacking	0
C12. Motivation/affect lacking	0
C14. Metacognition/deeper thinking lacking	0

The prominent points on the negative table (Table 49) were "English skills didn't improve" (most frequent), followed by "Course, negative," and then a new key theme "Socio-cultural skills lacking" in slot number 3. My deduction: the fact that *OL1* had an inverted bell curve that may be creating a critical break in engagement, and with the emergence of this new "socio-cultural skills lacking" theme, I hypothesize that the students may have felt *lost* from time to time during the course. This may be part and parcel with: (a) the negative affects that stem from having "critical breaks" in the middle of each chapter, and (b) the lack of design features that are meant to carry students though a virtuous cycle of development. In short, *OL1* literally and figuratively seems to be lacking the "Cycles" of deeper growth pedagogy that *OL2* embodies.

8.9 Comparison of OL1 and OL2

This next section shall take this discussion one step further. It will look at the real differences in the designs of *OL1* and *OL2*. It will then directly compare the results of *OL1* and *OL2*. This is necessary to help deduce the reasons behind the significant discrepancies between the *OL1* and *OL2* data that was produced by relatively similar students.

To better understand the difference in the flow of the two books, below is a side-by-side representation (Figure 84) of *OL2's* popular cycles (left) with the explanations (right).

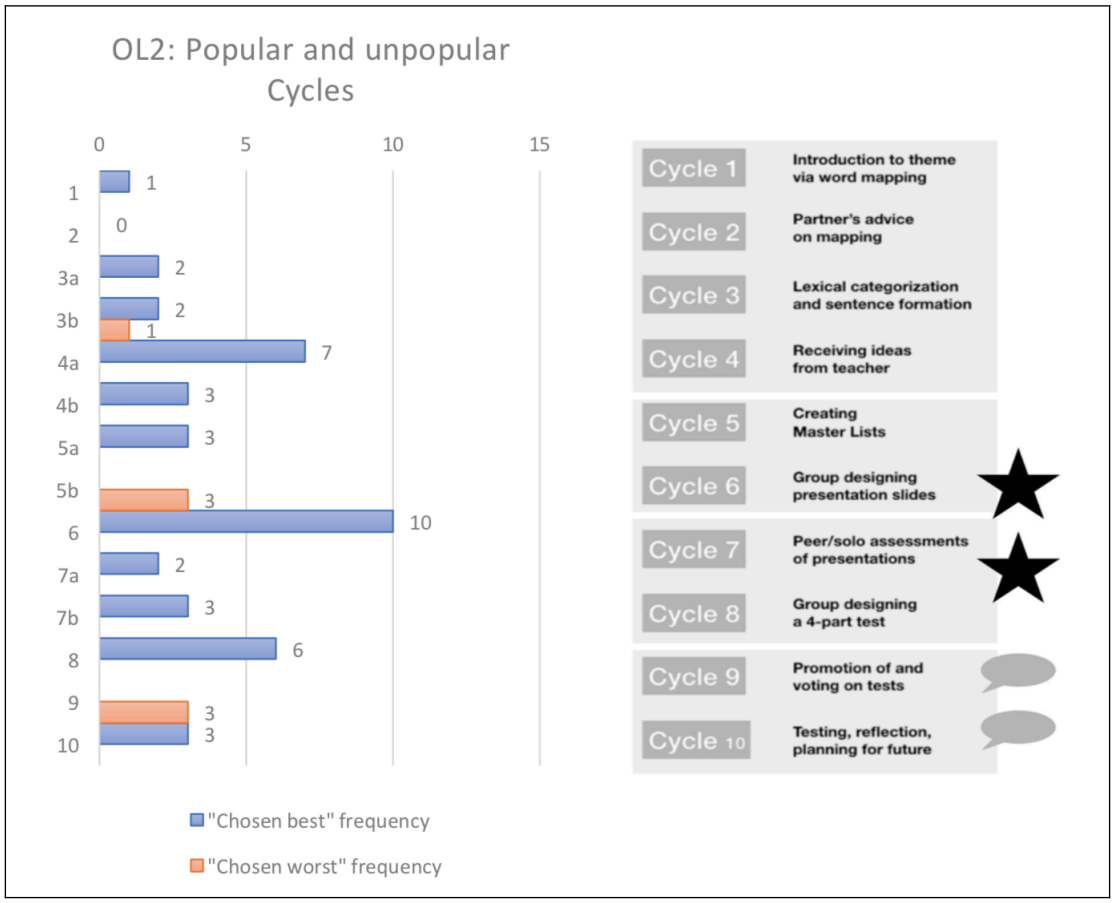


Figure 84: Graphic representation of OL2, popular vs. unpopular cycles

Next, is a side-by-side representation (Figure 84) of *OL1's* popular cycles (left), with the section explanations on the right side.

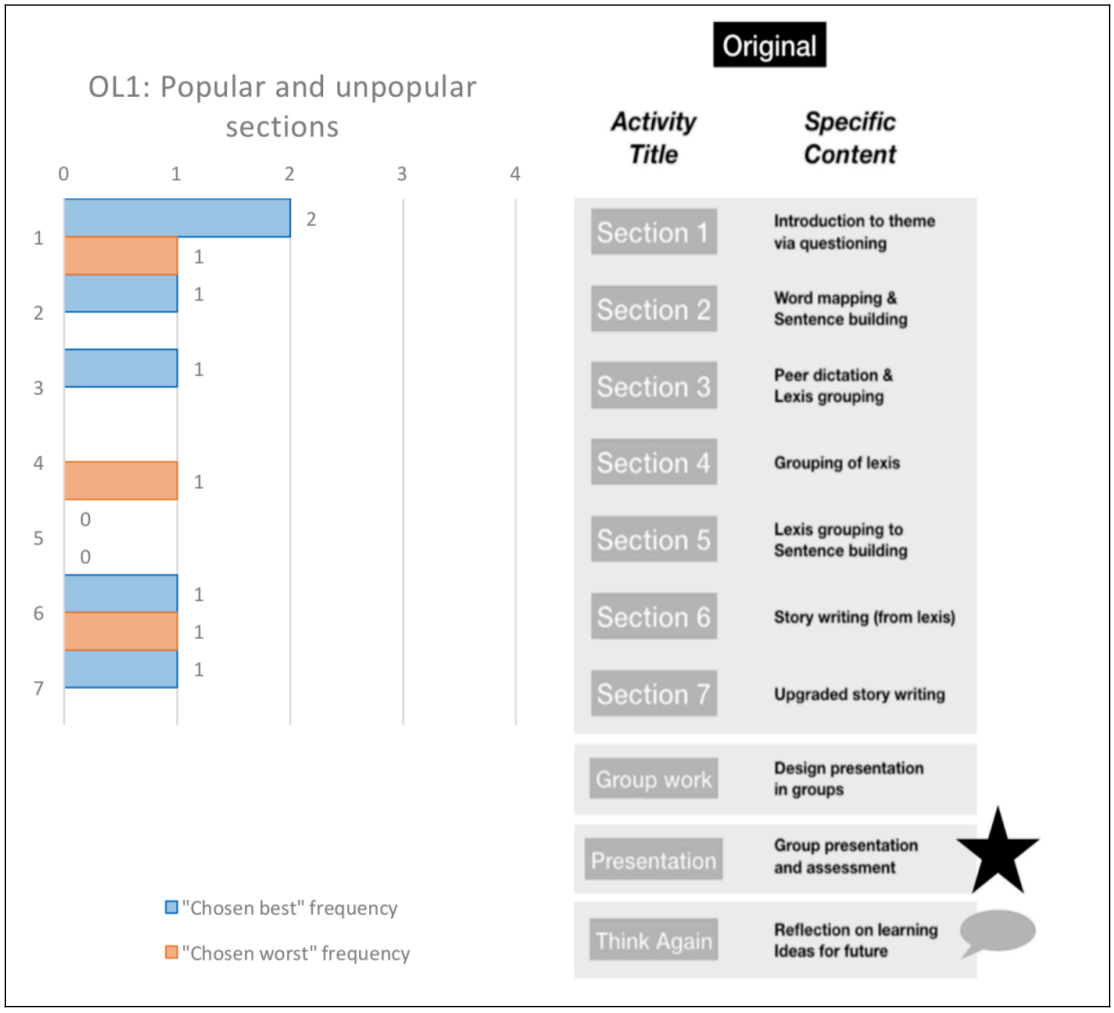


Figure 85: Graphic representation of OL1, popular vs. unpopular cycles

From the graphs, it is evident that for *OL2* (Figure 84) there are two climaxes (the starred area) that help create a natural bell curve of emotional flow across the duration of the module, which must be adding a visceral sense of cohesion to the module, most likely raising overall motivation—especially when compared to *OL1's* broken up flow (Figure 85). Again, this is likely the key to understanding why *OL2* generates so much more metacognition and elicits more demonstrations of deeper thinking than *OL1* does. This emotional bell curve was not an explicit design concept during the planning stages of *OL2*, but this attribute became apparent *and essential* during *OL2's* implementation in the classroom.

In contrast, *OL1* shows no such coherence of emotion—it is mostly flat and decidedly anti-climactic. What makes matters worse is having an indifferent Section

5 after an unpopular Section 4. As discussed in Chapter 5, the literature suggests that "horizontal cohesion" is imperative for good coursebook design. However, at the time of *OL1's* design, I had not yet uncovered the importance of "horizontal cohesion". Although it was not explicitly planned into *OL2*, the emotional curve climaxing seems to be a serendipitous feature of *OL2* that particularly distinguishes it from *OL1*.

Interestingly, Cycle 2 of *OL2* has consistently been considered "indifferent" by the students. It is a simple section that calls for pair work regarding the word-mapping activity that the students had just completed. Most coursebooks make a habit of adding this level of pair work (*"Let's review what you had just done. Discuss with a partner."*), and I do think such activities have merit, but within the standards of *OL2*, such pair work has been treated with indifference (too pedestrian?) by the *OL2* students.

The above point deserves more in-depth thought. Is "pair-work to review" too simplistic? Although none of the students had not been explicitly taught about the pedagogical theory that went into *OL2's* design, from their responses it seems that the sections designed to have more complex group interactions and peer-to-peer assessment (explicitly building up a high-support context for the students) are the sections that stood out for them. Those are the sections that they clearly regarded as being more useful and more helpful. This is an intriguing insight brought on by students' indication of their indifference to simple (now-pedestrian-to-them?) "pair-work for review." The students seem to require activities that are more in-depth in design—activities that are more cognitively engaging by design.

In sum, *OL2* creates and maintains a coherent and motivational arc that produces output akin to a high support context's output. And, it seems that many of the students grow to appreciate and expect that higher standard of engagement and growth. In stark contrast, *OL1* may have a critical break in the middle of each chapter that may be destroying the very cohesion that it is trying to build. As a consequence, students of *OL1* do not gain an expectation for that higher standard of engagement and growth that is apparent with in *OL2* usage.

8.10 Four years later: Data collection from the 2017-2018 school year

The primary studies for this thesis began in the first semester of 2014. Secondary studies were commissioned to further raise credibility of the 2014 data. The secondary studies are based on data collected during the 2017-2018 school year. Therefore, this section will review and compare the results of a qualitative study that was based on a collection of appraisal essays from 2014 and 2018. The rationale for this comparison is that if the trends from 2014 are similar to the trends found in 2018 then the data from 2014 (the primary study) is less likely to have been atypical; the data I will present in this section will show that the primary study participants' responses in 2014 were indeed similar to those from 2018.

8.10.1 Qualitative summary diagrams

The university has a compulsory questionnaire (end of each semester) which is mostly a multiple-choice questionnaire regarding students' feelings about their courses. The final section's data collection is conducted with a blank sheet of paper—free writing space for the students to "vent" or discuss anything they want to, regarding the course. This is an anonymous free-writing task. For my secondary study, I coded each student response from the 2014 term and the 2018 term with the following 1-5 rating protocol:

1 = Completely negative content
2 = More negative than positive
3 = Neutral (balanced mix)
4 = More positive than negative
5= Completely positive

The students were allowed to write in either language (Japanese or English), and the majority chose to write in Japanese, somewhat raising the validity and credibility of the responses. Being a bilingual speaker, I coded the responses myself. Table 50 shows the results from three of the main classes used in the primary 2014

studies (*OL2* users). The majority of the open-ended writing session resulted in a code score of 5 (N=42), meaning "Completely positive".

Table 50: 2014 data, University-administered questionnaire, open-ended section. Code on the top row.

Class designation	1	2	3	4	5
2014-3	1	0	1	2	11
2014-4	0	0	0	3	10
2014-5	0	0	6	4	21
Total	**1**	**0**	**7**	**9**	**42**

The results from the 2018 session were not very different from the 2014 results. I chose data from the same three courses (the "2017-2018 version" of each course). Regarding the specifics, Table 51 shows a slightly higher trend toward the upper range. Although there were a few 1s and 3s in the 2014 data, are no 1s or 3s in this collection. The vast majority of responses were coded a 5 (N=68). This trend across these classes is therefore very similar to the trends found in 2014.

Table 51: 2018 data, University-administered questionnaire, open-ended section. Codes on the top row.

Class designation	1	2	3	4	5
2018-3	0	0	0	1	26
2018-4	0	0	0	0	15
2018-5	0	0	0	5	27
Total	**0**	**0**	**0**	**6**	**68**

Although this data collection was not meant to be conclusive (the data was collected with a clear understanding of its limitations), it does highlight that: (a) there is a discernable trend, and that (b) the students from 2014 and 2018 responded similarly—regardless of the year, or class. So, even with its limitations, this secondary study effectively suggests that there was nothing particularly special or irregular about the 2014 students who participated in the primary study.

8.10.2 Relevant quotes for top themes

Quantifying qualitative data has utility—but in this section let us explore the actual quotes from the students. The main focus of this section is not to demonstrate commonalities between 2014 and 2018 students; that has already been demonstrated. Here I would like to focus on the differences and the composition of representative quotes from each group. One major difference between 2014 and 2018 is the amount of writing per group. The 2018 students tended to write more sentences and even added diagrams (illustrations and flowcharts) to their responses, describing the flow of the *OL2* learning process. I will use this section to review quotes from the participants that are representative of the top themes in the primary study, and to highlight students' insights—this exploration will provide vital insight into how *OL2* seems to affect the students.

> *"My presentation English got better because when I wanted to say something but it was not in my scripts, I was able to think that just say it out loud. It was the best thing. I got better in the class!!"* (Sem2, 301, 2018)

This quote vividly describes the moment that this student realized spontaneous L2 production. Moreover, she cheerfully described it as *the best thing.* If this had not happened in this student's past six years of English study in public schools, but suddenly happened to her this year, then it is likely that the cyclical design of *OL2* mentally prepared her with both the lexical capacity, and the psychological "can do attitude" necessary to make these spontaneous L2 utterances (ad-libbing) in public. It is naturally pleasing to notice that she went on to say it was "the best thing", fortifying the nuance that she realized the ability to ad-lib in English, and was happy to break free of the rote-memorization teaching that she had experienced until now.

> *"I can improve my creativity while discussing the presentation contents with my team members. I also really enjoyed listening to other team's presentations. I could learn a lot from them."* (Sem2, 302, 2018)

This quote highlights the metacognition that went into the realization of the birth of creativity due to group work. This is a complex realization that culminated from the embedded cyclical tasks in *OL2.* Moreover, this student discusses how s/he learned a lot, and enjoyed listening to, other group presentations. This hints at why *OL2* was good at raising metacognitive skills. From my own years of observations of other EFL classroom presentations (and indeed many professional presentations), the presentation sessions leave the majority of the sitting people quite bored; with nothing to do, even the most diligent people with the best intentions will eventually mentally drift from the content at some point and daydream. *OL2* was designed to combat this naturally occurring phenomenon.

With *OL2's* embedded peer-to-peer evaluation and self-evaluation activities that return cyclically with the expectation of improvement with each return, the students are compelled to pay much more attention than usual—on a variety of aspects of the presentations. Because they have so much to follow, and because they realize that paying attention to the content delivery will be advantageous for them with their own upcoming presentations, the sessions are naturally transformed into well-motivated learning and growing experiences (instead of daydream time).

The data suggests that *OL2* is successful in providing a relentless pedagogical flow that the teacher (and eventually the student) presumes and anticipates will give rise to rapid organic L2 maturation, inclusive of cognitive development, within the provided structure of the course.

> *"I was able to enjoy while gaining knowledge. I was able to learn how to speak while learning how to think"* (Sem2, 301, 2018)

This quote embodies the idea of cognitive growth *with* the L2. That is to say

that, unlike typical task-based projects, or content-based learning, for *OL2*, the primary function of the design is to raise cognitive skills as demonstrable in the L2. I feel this quote captures the essence quite well—and shows that this concept of *OL2* is being realized; it is not just a theoretical fantasy, but a practical realization of L2 maturity.

> *"In the second semester, not only did my presentation skills go up, but my teamwork, and my speaking skills also went up. This must be because of the "positive cycles" that you talk about. I realized that when I'm really having fun designing the presentation content, I do much better on presentation day, and even the content is better because of that. I now realize the importance of the cyclical relationship."* (Sem2, 311, 2018)

This quote shows great depth of understanding of the mechanics of the Cycles in *OL2*, and how they instigate complex skill development. Moreover, s/he is clearly connecting the enjoyment (flow?) of designing presentation content with better outcomes—and that this is all because of the "cyclical relationship". This is poignant insight from a second-semester student. It shows how students are capable of connecting the dots between raising motivation and raising skill levels—while appreciating the design of the pedagogy that got them there.

> *"As the course progresses I became more aware of the importance of logical flow, and I was able to raise the logical flow of my thoughts. I was then able to actually use my higher level logical thoughts in the presentations."*
> (Sem2, 312, 2018)

This quote connects well with the data to discussed in Chapter 7, and it also connects strongly with all of the top themes that emerged during the qualitative data collection. The student discusses an amazing progression from the realization of the importance of logical flow, to becoming able to think with higher level logic, and

then being able to embed that higher level thinking into English presentations. This vividly illustrates how *OL2* is creating a context that invites and nurtures higher level thinking while providing a creative avenue for its expression in the L2.

> *"(Because of this course) I gained a reason to proactively speak in English more than before. So, at my part-time job, a foreigner came and I was able to have simple but effective conversation with him."* (Sem2, 313, 2018)

The first part of the quote touches on how the course makes him/her become more proactive in English usage – and then was able to "rise to the occasion" when necessary. This ability to "rise to the occasion" is a primary (and recurring) theme regarding teaching outcomes of *OL2*. Although s/he was not in a context that looked like a high support context, s/he was able to not shy away, but to apply the emerging skills. This is significant in Japan because most students very openly shy away from speaking with foreigners.

> *"My level of vocabulary rose as I was preparing for the presentation. I came to like English more than before"* (Sem2, 314, 2018)

This is an interesting and important quote because it connects (the roundabout way of) studying lexical items in *OL2* to the realization of learning more vocabulary during the course. Moreover, the student seems to be enjoying this process "more than before". This is an important nuance because while the student is fixated in how s/he is learning more vocabulary in the quote, in the past six years of public school, they have been bombarded with explicit rote-memorization of (and the cramming of) vocabulary. So, it is an interesting point that this student is now not only learning new vocabulary, but s/he is enjoying it. Most interesting of all, this learning is *bottom-up* and implicit (meaning it is generated from the students on an internal level). This adds strength and credibility to the core design of *OL2,* as a

coursebook that facilitates this learning without mandating learning content through the reliance of the more typical direct/pragmatic *top-down* strategies that are globally accepted.

> *"As the term progressed, I went from a presentation novice, to somebody with realized the biggest gains in 'presentation skills'. What I mean is, as I do my presentation preparations, I think about talking speed, and flow. I think about how the listeners will understand what I'm saying and consider how to change the content so that it will be easier for them to understand. This sort of knowledge is something that I was able to bring to and use will in my other courses."* (Sem2, 316, 2018)

This quote shows deep metacognition with signs of empathy. This student is no longer thinking at the level of "what should I talk about?", but at the level of, "How will this come across? What adjustments should I make, so that it will be easier for *them* to understand?" This is a substantial step up in a cognitive complexity from the former level, and a sign of higher psychological maturity. In Maslow's terms, s/he is no longer worried about being hurt, but more at the level of, "What can I do for others?" On the complexity side, the student has the drive and the capability to adjust their language usage for the audience – and comment comes from a self-described presentation novice. This quote illustrates the complexity of growth being instigated at this rapid pace.

> *"I used to have a negative image (anxiety) of English language learning. But since this course is fun, my apprehension of English slowly dissipated."* (Sem2, 317, 2018)

I included this quote specifically because it sheds light on the functional mechanics behind the lowering of negative themes in the second term. The *OL2* classrooms follow the exact same flow of cycles for both terms, so why would the

negative themes lessen in the second term? One hypothesis that I discussed earlier was that perhaps the students realized their English is getting better, so they naturally had less bad things to say about the book. But here we see another very real affect of public school teaching. This student had anxiety regarding English language learning (a terrible consequence of public school English classes), but during the flow events that year, this student's anxiety levels lowered and s/he was able to enjoy the learning process. So, I may be able to add to the above hypothesis that first term students may still be coming to class with "bags of anxiety", which may very naturally influence them to see and raise more negative themes, but as the year progresses, anxiety lessens, more enjoyment sets in, the students are more relaxed, and they begin to notice the balance of bottom-up learning within the unwavering cyclical flow of events and that is why there are less negative things to say about the coursebook.

Did *OL2* help lower this student's anxiety? This does seem to be the case for this students, and therefore this is another important discovery from this study cyclical (no-surprises) pedagogy that embeds autonomy and encourages student creativity may help lower student anxieties that may have manifested over the years.

8.10.3 Student generated theoretical diagrams regarding OL2 pedagogical flow

There was a remarkable occurrence in Class 3 and Class 4 of 2018; there was a spontaneous trend of adding graphic flowcharts as illustrative explanations of student learning cycles based on the pedagogy of *OL2*. I am not sure how this phenomenon began. It seems to be an organic and natural product of effective and affective student communication and in-depth thinking regarding the course. I will present five of the student-generated flowcharts; they provide an invaluable look into the minds of the students – regarding the affect of *OL2* at the end of a full school year.

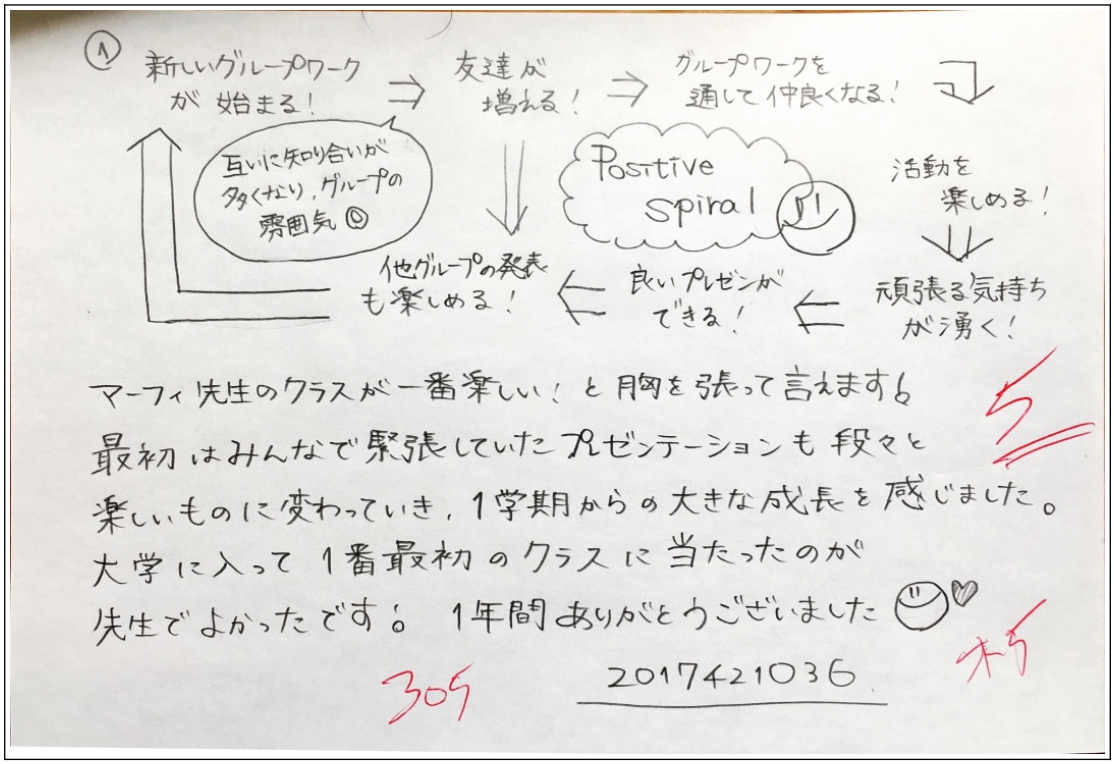


Figure 86: Student's theoretical diagram A

Diagram A (Figure 86) was drawn by a student who commented, "Mr. Murphy's class is the most fun – I can say this with pride." Her complex diagram has *two* cycles embedded. From the top left (clockwise): "New group work begins" "make new friends" "group work brings us closer" "enjoy the activity more"

"become more motivated" "able to make good presentation" "able to enjoy other groups' presentations" (return to "New group work begins"). The arrow down the center flows from "make new friends" to ""able to enjoy other groups' presentations".

In this way, the student has drawn up a diagram that shows how her motivation is raised through group work, making new friends, doing good presentations, and viewing others' presentations. Moreover, she quickly picked up on the cyclical design of *OL2*, and had two cycles concurrently producing affect.

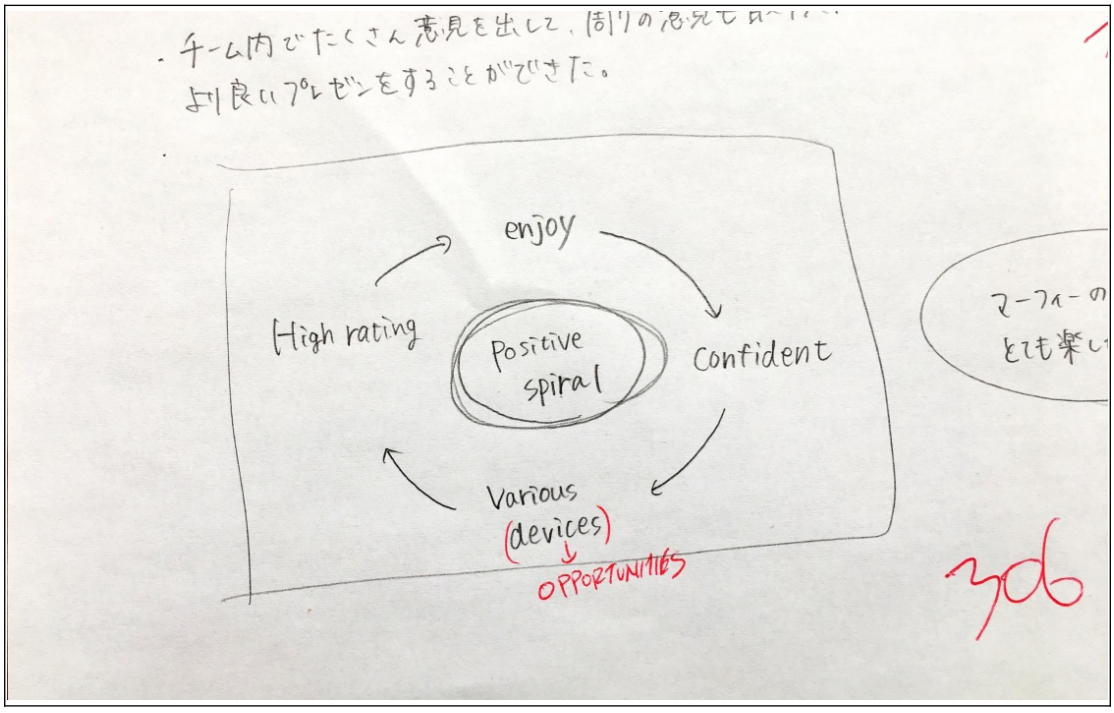


Figure 87: Student's theoretical diagram B

I was able to ask this student (she did not remain anonymous) what she meant by "various devices" and I learned that this was not about the computer devices in the classroom; she meant "various opportunities" created by the coursebook. Diagram B (Figure 87) shows that this student saw *OL2* as a book that creates a "positive spiral". This is a Japanese phrase synonymous with virtuous cycle. In her comments, she wrote about how well she was able to speak her mind in the group work. In her diagram, she points out the relationship between confidence, high

marks, and enjoyment via the various opportunities that were created by the flow of *OL2*. She also commented on how fun the class was for her.

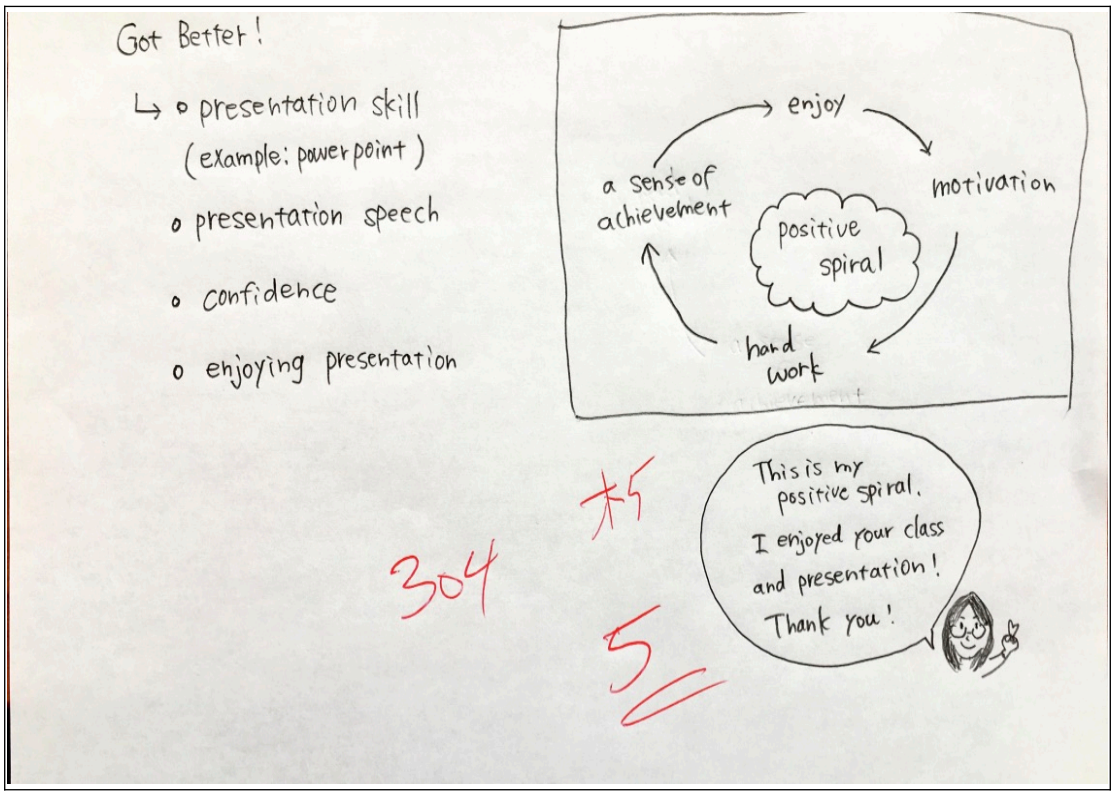


Figure 88: Student's theoretical diagram C

This student also used the term "positive spiral" to mean virtuous cycle. Diagram C (Figure 88) depicts this student's understanding of the relationship between motivation, hard work, sense of achievement, and enjoyment, in the *OL2* class. On the side, she also discussed the various skills that improved during this course. This student was not in any of the Focus Groups, and therefore I was surprised to see that her side bar comments were similar to the themes that emerged from the Focus Group discussions. It seems that regardless of the modality, similar realizations emerge from the *OL2* students. This adds another layer of trustworthiness to the multi-faceted approach taken by this thesis. Moreover, there are recurring themes even within these five diagrams. In the three diagrams thus far, the recurring themes are enjoyment, motivation, work, and good marks.

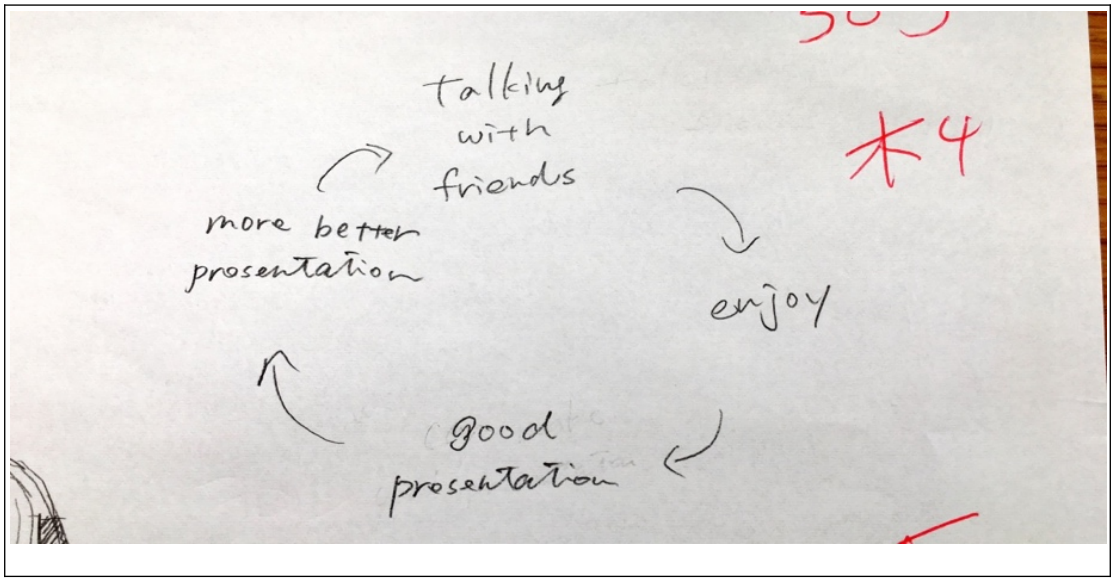


Figure 89: Student's theoretical diagram D

I included this fourth diagram (Figure 89) to strengthen the idea of the recurring themes—plus one new theme. In this diagram, the student illustrates the relationship between talking with friends, enjoyment, good presentations, and *improving* presentations. This final bit regarding improving presentations reflects the piloting I did with *OL2 Revised*, in their second term. In that revision, the students were given the opportunity to upgrade their presentations—and perform the upgraded version in front of the class. It seems this student picked up on the essence of the *OL2* revision, and made it one of the four key points in her illustration of the *OL2-R* classroom. The addition of this "upgrading task" is the most significant change that I made to the pedagogical flow for *OL2 R*. This will be discussed more detail in Chapter 9.

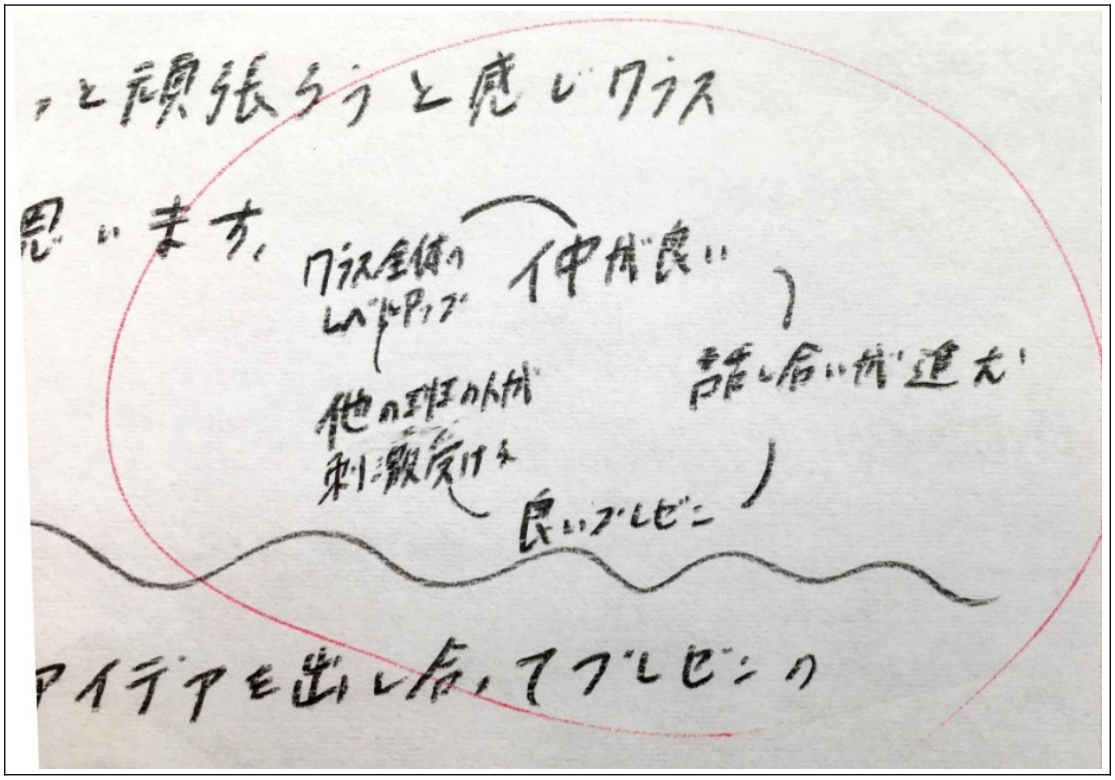


Figure 90: Student's theoretical diagram E

The fifth and final diagram (Figure 90) focuses on the concept of growth. The four points in the cycle are: (1) The whole class levels up, (2) we bond, (3) our discussions become more fruitful, and (4) we receive good stimulation from the other groups' performances. Again, there are recurring themes, but the nuance of this one is slightly different. This diagram happens to be drawn by a male student (the only male student of the five). The strongest nuances here are regarding the fruitfulness of discussion, and stimulation from other groups conveying the noticing of social learning. Moreover, this student (like the student who drew Diagram D) picked up on the "leveling up" element. Again, reinforcing the importance of the revision embedded into *OL2-R*.

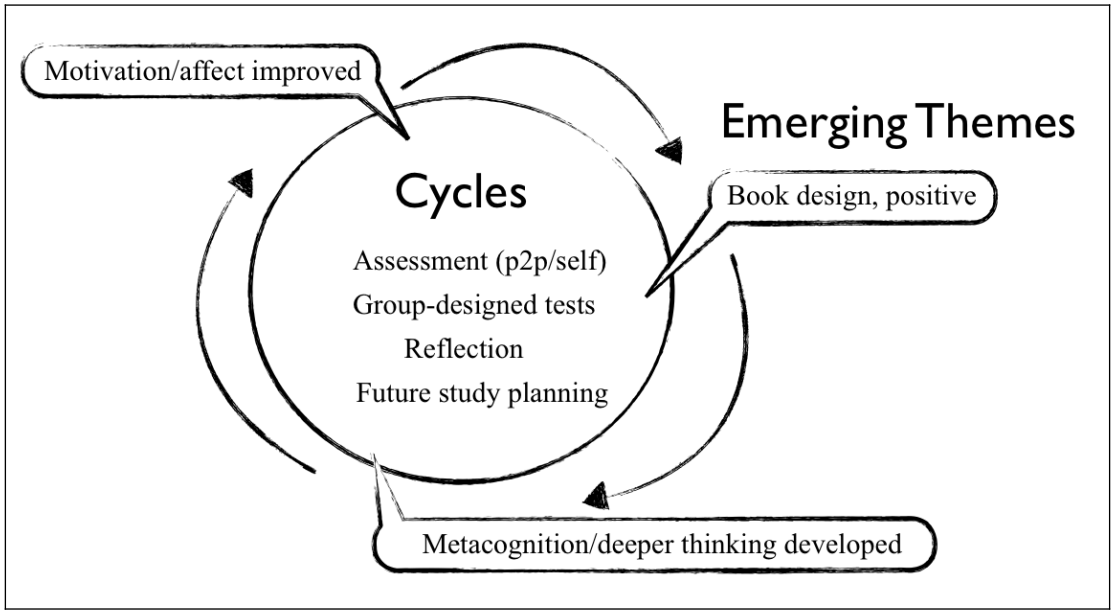


Figure 91: My diagram based on the student's diagrams

After thoroughly viewing the students' spontaneous work with their diagramed flowcharts, I combined the ideas and put the major themes into a single cohesive design (Figure 91). In a way, this is like a newer version of the COOPs protocol, but this time, instead of it being a designer's conception, it is a product from the eyes of the students. And, because it is from the users' perspective after a year of usage, it may be the most truthful depiction of *OL2*.

8.11 Five years later: The affect of teacher rapport on motivation and course evaluations

As this was an on-going action research project, each year of study produced a new set of data collection ideas that organically developed from previous years' data and conclusions. The data collection methods and ensuing deductions therefore evolved with each year. My final data collection sessions were concluded during the 2018-2019 school year. This section shall focus on data from that final research year. Until now, I presented the quantitative data and qualitative data in separate chapters—Chapters 7 and 8, respectively. However, because the deductions and conclusions regarding this final year of data collection is more easy to follow when the quantitative and the qualitative data sets are back-to-back, I have literally listed them back-to-back across the following sections (Sections 8.11.1 and 8.11.2).

As in the previous years, I conducted online surveys with the students taking *OL1*- and *OL2*-based courses, and the university conducted their annual course evaluations. My two groups of students for this final year were comprised of student groups that were essentially identical to the primary study of 2014. Just as in 2014, there was one *OL1* group of participants (N=22) and one group of five *OL2* class participants (N=111). However, for this final year, I decided to also study the potential affect teacher rapport has on student motivation and course evaluations – and how much my results from the *OL1* vs. *OL2* growth studies may have been affected by student-teacher rapport and student motivation (rather than the pedagogy—which had been my focus thus far). This idea that people may provide higher ratings based on rapport is colloquially known as "the halo effect." A famous example of this is the rise in Macintosh *computer* sales during the boom of the iPhone. Why would the computer sales be affected by phone sales? Business analysts attributed the rise in Macintosh sales to the "halo effect" – the affinity users had for their iPhones rubbed off on the Macintosh line of computers simply because they were made by the same company (Apple, Inc.), and the computers started to look more appealing to them. Was something similar happening in my classrooms? If so, how might that have affected my data over the years? This new line of questions had

the potential of dramatically changing major portions of the deductions and conclusions that were made thus far.

8.11.1 Quantitative data

To receive the best possible results, I created a high support-like context for the participants and received responses from them only after an elaborate set of deep-thinking tasks. Both groups were first provided with a 40-minute small-group discussion session regarding to what degree teacher rapport may affect student motivation and class evaluations. I made sure the students had deep and focused discussions on these topics. In the end, they were asked to write their conclusions down, and give a 0-4 rating on the potential of each of these two topics, along with their rationale for the 0-4 rating that they provided on that day. They were then given homework → they were asked to continue to think about these issues over the up-coming week and update their conclusions and ratings if their thoughts had changed in any way over the course of the week. In the subsequent class, the students had their most current version of their paperwork in-hand as they typed up a purpose-designed online survey.

On the survey, I asked the students to rate how much their personal rapport with the teacher affected: (a) their motivation to do well in the class, and (b) their class evaluation ratings on their annual university questionnaire. The responses were on a scale of zero to four (zero = no affect, 4 = very strong affect), with written rationales for each. By providing the participants with: (1) 40-minute group discussion time; (2) hand-written work with rationales; (3) one week to further their thinking—with written updates; and (4) an update-in-hand online survey session that summarizes their week of pondering, it would seem likely that the participants were able to produce fairly accurate and well-nuanced responses by the time it came to take that final online survey.

Regarding the survey results, first, let us review the raw data. Figure 88 and Figure 89 display the student reported affect of "Motivation" and the student reported affect on course evaluation, respectively, with totals in graphic form for

easier visibility of the trends across the groups and the affect scores that emerged. For the student reported magnitude of teacher rapport on their "Motivation" on a zero to 4 scale, the mean results per coursebook were: *OL1* (M = 2.82, SD = 1.26), *OL2* (M = 2.87, SD = 1.40). For the student reported magnitude of teacher rapport on their "Course Evaluation" on a zero to 4 scale, the mean results per coursebook were: *OL1* (M = 2.59, SD = 1.26), *OL2* (M = 2.57, SD = 1.32).

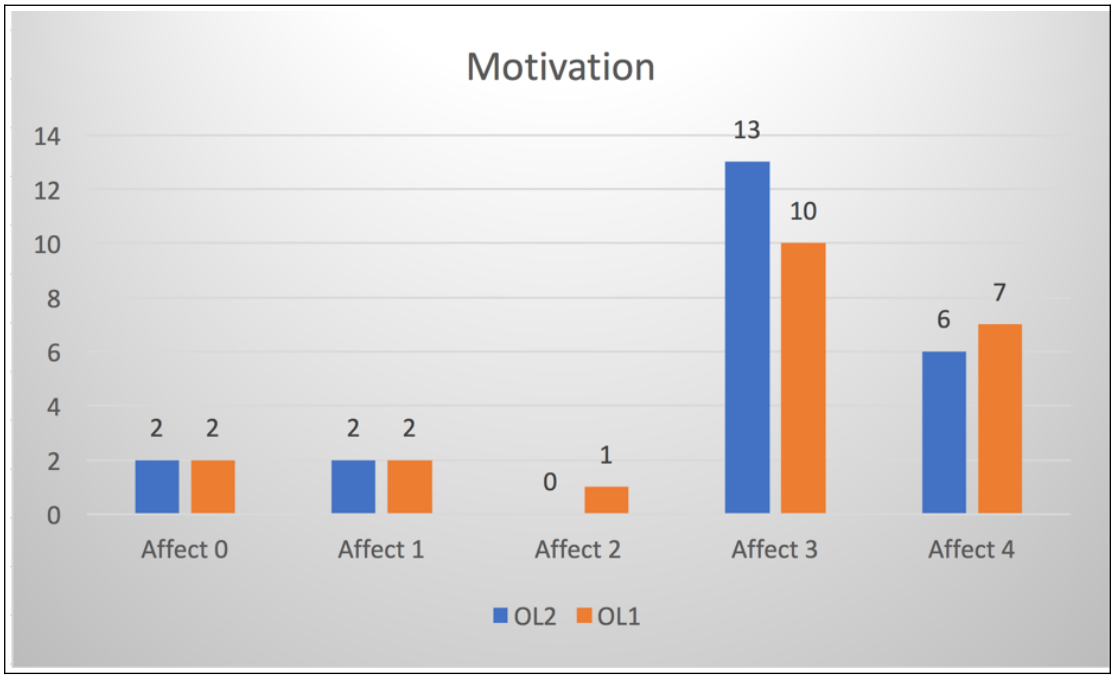


Figure 92: The reported affect of rapport on motivation, OL1 vs OL2

In raw data form (Figure 92), the reported affect of teacher rapport on student motivation is clear. The majority of the students, after a week of activities regarding this topic reported strong magnitudes of affect; the vast majority reported strong affect levels of 3 to 4. The trend is similar, but not identical when comparing the *OL2* data (blue bar) with *OL1* data (orange bar). With either book, rapport with the teacher strongly affects student motivation to do better in class. In Section 8.11.2, I will explore qualitative data (illustrative quotes from the participants) that shine more light on these findings.

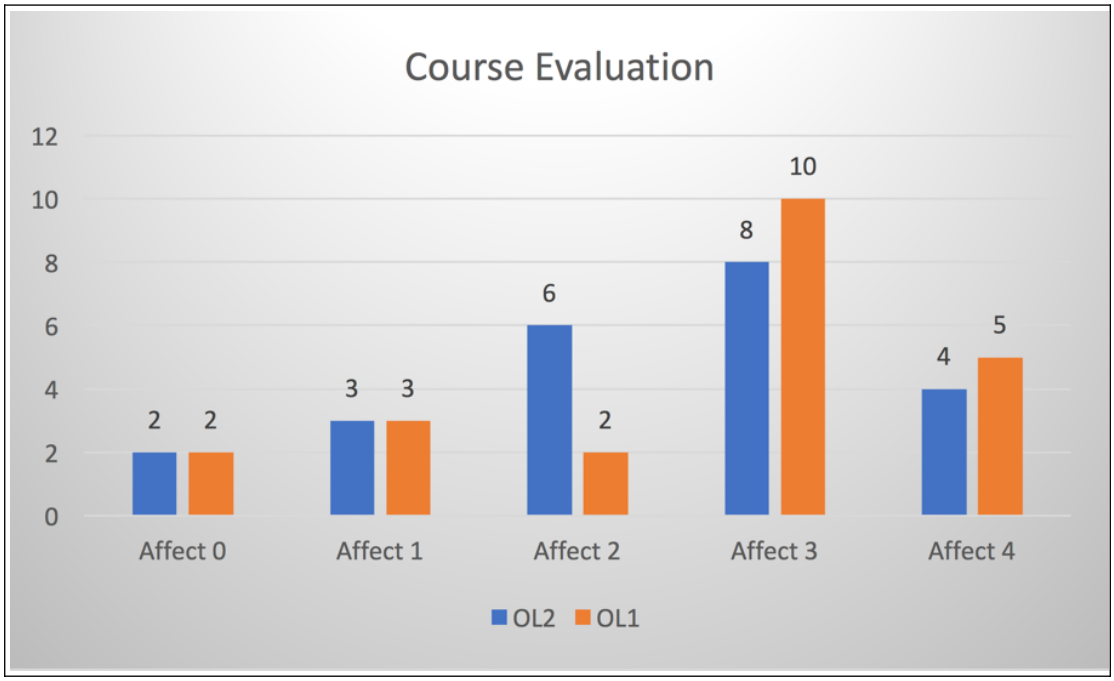


Figure 93: The reported affect of rapport on course evaluation, OL1 vs OL2

In raw data form (Figure 93), the reported affect of teacher rapport on students' course evaluation is less clear than the Motivation data, but it does unfold into an interesting story. The majority of the students reported affect levels between 2-4. The trends across the books are somewhat similar, but *OL1* shows a stronger pull toward the stronger affect levels of 3-4 than *OL2* does; *OL2* data shows a more even spread perhaps *OL2* students were *a bit* fairer? Nevertheless, it can be deduced that the majority of the students, regardless of the coursebook, reported that rapport with the teacher *does* affect course evaluation—often to a high degree. In Section 8.11.2, I will explore qualitative data (illustrative quotes from the participants) that shine more light on these findings.

The raw data above produced vital results—the results were an eye-opener for me as much as they were a confirmation of what I suspect most teachers already "know" (but do not talk about). Still, the raw data alone is not enough to paint a full picture; statistical analysis is necessary to fill in the gaps, and to substantiate these findings. I chose 2-tailed t-tests (SPSS). For the student reported magnitude of teacher rapport on their "Motivation" with a scale of zero to 4: *OL1* (M = 2.82, SD =

1.26), *OL2* (M = 2.87, SD = 1.40); t(131) = -0.173, p = 0.863. For the student reported magnitude of teacher rapport on their "Course Evaluation" with a scale of zero to 4: *OL1* (M = 2.59, SD = 1.26), *OL2* (M = 2.57, SD = 1.32); t(131) = -0.076, p = 0.939.

Regarding these two groups of data, the t-test analysis (2-tailed) showed p values of 0.863 and 0.939, respectively. Considering these t-test outcomes along with my previous data, it is clear that the differences between *OL1* and *OL2* for both "Motivation" and "Course Evaluation" are statistically non-significant. So, although both coursebooks stand on similar grounding regarding the affect of the "halo effect" on course evaluation and student motivation, only *OL2* spawned rapid L2 maturity; this is one of the most thought-provoking findings of the entire thesis.

8.11.2 Qualitative data

This section takes over from where the above section ends—by providing qualitative data to help better interpret the significance of the quantitative data. This section provides a collection of representative quotes that illustrate the meaning/ rationale of the zero to 4 scores as reported by the 2018-2019 participants (discussed above). I will first display and discuss the qualitative results regarding the affect of teacher-student rapport on student motivation, per Affect Factor (Table 52) After that, I will display and discuss the qualitative results regarding the affect of teacher-student rapport on their course evaluation, per Affect Factor (Table 53). For easy differentiation between *OL1* and *OL2* quotes, I have added a double asterisk (**) to the quotes from the *OL1* users.

Table 52: Illustrative quotes on rapport and motivation

Inquired Theme	Affect Factor	Illustrative example
Rapport's Affect on Motivation	4	"I think whether I like a teacher or not affects my performance strongly. Although there are some exceptions, but I can say like/dislike of a teacher affects my performance because if I liked that teacher my performance would be much better perhaps. " (*OL2*; Sem2, 001, 2019) "If I like the teacher, my performance is better. I will be interested in his/her any classes even if they are boring contents. On the contrary, if I do not like the teacher, my performance is worse. I can understand how much his /her class is very important, but my prejudice against him/her obstructs my concentration while the class. …I think my favorite or not teacher affects my performance in class clearly." (*OL2*; Sem2, 002, 2019) "Because I liked the teacher, I was motivated to go to every class, so I think there is a strong connection between rapport and motivation" (*OL2*; Sem2, 003, 2019) ** "There is a very strong connection between rapport and motivation. You cannot have motivation if you don't like the teacher." (*OL1*; Sem2, 004, 2019) ** "I can work on any difficult task, if I like the teacher" (*OL1*; Sem2, 005, 2019)
Rapport's Affect on Motivation	3	"I believe our willingness to do well in class changes (relates to) how much we like/dislike our teachers." (*OL2*; Sem2, 006, 2019) "I may be self-centered by saying this but I can't bring myself to listen in classes taught by teachers that I don't like, so there is a connection here. However, since this is an active style course and it is the students that need to do the work, the affect is less than other courses." (*OL2*; Sem2, 007, 2019) "If I don't like the teacher, not only do is it hard to learn from him, I don't even want to attend his class." (*OL2*; Sem2, 008, 2019) ** "You know, these things are really not only about teaching content. It's about people and relationships." (*OL1*; Sem2, 008, 2019) ** "Very helpful and kind teachers naturally motivate me to work harder and do better." (*OL1*; Sem2, 010, 2019)
Rapport's Affect on Motivation	2	"In active learning classes such as this one, most work is done between the students, so rapport with teacher does not affect motivation much. Motivation is more affected by who is in my group as we do the tasks." (*OL2*; Sem2, 011, 2019)

Rapport's Affect on Motivation	1	** "I am not a person whose motivation changes per teacher." (*OL1*; Sem2, 012, 2019) ** "(There is little affect because) regardless of if I like or dislike the teacher, I almost never talk to my teachers anyway." (*OL1*; Sem2, 013, 2019)
Rapport's Affect on Motivation	0	"I don't remember ever not going to a class because I didn't like the teacher. I was always motivated to come to this class. This had nothing to do with teacher rapport. Most importantly, I always looked forward to talking with my classmates. I never thought this course was a pain." (*OL2*; Sem2, 014, 2019) "Motivation is something that students need to bring on themselves. Motivation is not something that the teacher should have to go out of their way to awaken for the students." (*OL2*; Sem2, 015, 2019) ** "This class is about working with other students, so the teacher had no affect on my motivation." (*OL1*; Sem2, 016, 2019)

Table 52 lists representative quotes that explain students' rationale behind their zero to 4 Affect Factor marks regarding the connection between teacher-student rapport and motivation to do well in class. The quotes show us the criteria the students came up with as they worked on this task. What quickly became obvious was that there was no discernable difference in the the rationales between *OL1* and *OL2* users (although *OL2* users tended to produce longer responses). Here, I will provide a vivid example (quotes 003 and 004) of how similar the responses are; regardless of the coursebook, the students are providing rationales that are essentially identical (regarding the meaning of Affect Factor 4):

> *"Because I liked the teacher, I was motivated to go to every class, so I think there is a strong connection between rapport and motivation"* (*OL2*; Sem2, 003, 2019)

> ** *"There is a very strong connection between rapport and motivation. You cannot have motivation if you don't like the teacher."* (*OL1*; Sem2, 004, 2019)

It is refreshing to see these candid remarks about how strongly rapport with the teacher affects students' motivation. On the lower end of the spectrum, the few students who provided the zeros and the 1s displayed relatively deep, philosophically-oriented rationales. A prime example of this is quote 015:

> *"Motivation is something that students need to bring on themselves. Motivation is not something that the teacher should have to go out of their way to awaken for the students."* (*OL2*; Sem2, 015, 2019)

This quote calls to mind the age-old debate within *Language Teaching* and *Education* courses that questions → *Should the teacher be an entertainer? Must teachers seriously care what students think about them or the pedagogy?* We can deduce that the few students at the lower end of the spectrum may agree with the naysayers of such care. However, the vast majority of the students provided Affect Factors in the upper end of the spectrum. Combining that quantitative data with this qualitative data, we can rather confidently deduce that the majority of these EFL students did appreciate the air of entertainment, the student-teacher rapport, and the "fun" that was incorporated into these active learning classes they clearly stated that these qualities raised their motivation to do better in class.

The backbone of my classrooms using the COOPs protocol (what some of my students colloquially referred to as "active learning") —be it *OL1* or *OL2*—is autonomous group work leading to peer-assessed presentations of self-imposed and self-taught content. But, without that strong motivation to do well in class, this form of autonomous group work can rapidly degrade into an altogether unproductive waste of class time. What can be learned from all of these points? We can sum all of these parts together to form the conclusion that according to the majority of these EFL students' responses, positive teacher-student rapport helps provide the motivation to make good use of, and gain an appreciation for, the course-provided autonomy (be it *OL1* or *OL2*).

Table 53: Illustrative quotes on rapport and course evaluation

Inquired Theme	Affect Factor	Illustrative example
Rapport's Affect on Course Evaluation	4	"My evaluation of classes depends on the teacher. If it is my favorite teacher, I want to take the class again and again. And I will recommend the class to my friends. If I have a chance to answer a survey of the class, I will write what I want him/her to continue his/her teaching style. That is why, I think teachers affect our evaluation of the class." (*OL2*; Sem2, 017, 2019) "I think many of the students in this class liked the teacher and so there must be many high evaluations…I believe student-teacher rapport connects to how we evaluate the course." (*OL2*; Sem2, 018, 2019) ** "(There was a strong affect) because you always have upbeat and entertaining classes, students cheerfully did their group work, and that made an atmosphere where we could easily speak our minds." (*OL1*; Sem2, 019, 2019)
Rapport's Affect on Course Evaluation	3	"…Sometimes I can be affected by like/dislike of a teacher, so I checked 3 here. For example, even if the class is boring and not interesting, if I like a teacher, I can perform better and try to understand or learn from him/her. In the result, I can deepen my knowledge through the class…sometimes can be affected in evaluating the class." (Sem2, 020, 2019) "I do think that personal feelings for the teacher affect course evaluation. For example, if you do not like the teacher, you will not enjoy the class and that will lower the course evaluation." (*OL2*; Sem2, 021, 2019) "To be honest, I don't really understand the mechanics of how I end up evaluating the class. But I think I am going in the right direction." (*OL2*; Sem2, 022, 2019) ** "I love this class. The teacher's explanations are easy to understand and because of that I was motivated to do better presentations each time. Until now, public speaking was difficult for me, but I no longer feel that way." (*OL1*; Sem2, 023, 2019) ** "Because a good teacher's class is better than a bad teacher's class, I think" (*OL1*; Sem2, 024, 2019)
Rapport's Affect on Course Evaluation	2	"Because this is an active style class and students do most of the work, in such a class I don't think there is much of a relationship between rapport and evaluation." (*OL2*; Sem2, 025, 2019) "I believe I assessed fairly" (*OL2*; Sem2, 026, 2019) ** "I don't think there is much affect, but I admit if I like the teacher, my evaluation may be a bit affected." (*OL1*; Sem2, 027, 2019)

Rapport's Affect on Course Evaluation	1	"(There is little connection because) rather than personal feelings toward the teacher, what's important is how much progress I made during the course." (*OL2*; Sem2, 028, 2019) ** "I am capable of unbiased evaluation." (*OL1*; Sem2, 029, 2019)
Rapport's Affect on Course Evaluation	0	"There is no connection to evaluation." (*OL1*; Sem2, 030, 2019) "For example, I have never thought a class was difficult to understand just because I didn't like the teacher. I don't feel a connection with that and my progress. I feel I still have potential to progress." (*OL1*; Sem2, 031, 2019) ** "This is a class where we work with other students. Teacher rapport had no affect on course evaluation." (*OL1*; Sem2, 032, 2019)

Table 53 lists representative quotes that explain students' rationale behind their zero to 4 Affect Factor marks regarding the connection between teacher-student rapport and their course evaluations. The quotes show us the criteria the students came up with as they worked on this task. As with the data on the previous Table, there was no discernable difference in the quality of the rationales between *OL1* and *OL2* users.

It was fascinating to see how many students' rationales showed absolutely no hint of remorse regarding how strongly *like/dislike of a teacher* directly affects their course evaluations—as is represented well with this quote from an Affect Factor 3 response (021):

> *"...if you do not like the teacher, you will not enjoy the class and that will lower the course evaluation."* (*OL2*; Sem2, 021, 2019)

On the opposite end, the rationales for zeros and 1s discussed how unbiased they were capable of being for the sake of evaluation. Interestingly, and in agreement with above, there was no hint of the *virtues of being unbiased* for course evaluation. The responses were of a more pragmatic nature. Quote 030 is representative of this comparatively insipid quality:

> *"There is no connection to evaluation."* (*OL1*; Sem2, 030, 2019)

What can be deduced from this set of data? It is imperative to unravel the data with the quantitative results in mind. So, as with the prior table, the majority of these EFL students openly claimed that teacher-student rapport had a strong affect—in this case, it was on how they chose to evaluate their courses. I found the interpretation of this data and the reporting of it to be more difficult than any of the other data in this thesis. In most cases, the students straightforwardly responded that they provided higher ratings for teachers that they liked.

The few students that provided Affect Factors of zero to 1 dryly claimed no connection between the two. Consciously or unconsciously, it would seem that they felt that they were unbiased. However, they didn't take the extra step to discuss the significance/virtues of being unprejudiced. This may indicate that from the Japanese students' perspective it is not worth noting.

There was one mention of *fairness* "I believe I assessed fairly" (*OL2*; Sem2, 026, 2019), but that was for an Affect Factor of 2(!), which from a mature "Western" perspective, is a laughable combination; an Affect Factor other than zero may automatically be construed as an admission of guilt, or in the very least, partiality. Yet, in Japan an Affect Factor of 2 can be considered a fair assessment. This is not an atypical response in Japan, as can be deduced from nuances in the other quotes—and the quantitative data that plainly showed admission of strong personal bias. In this context, it can also be construed as an indication of a different connotation of the word fairness.

For this age group, fairness might mean an evaluation that is *genuine* regardless of subjectivity/objectivity. If that is the case, then this is likely an age-related (non-cultural affect)—not something specific to Japanese EFL students. If that is the case, then it is matter of lexical construal stemming from age-appropriate assumptions and little else.

It can be argued that because I had produced a high support context for this portion of the study, the students opened up more than they normally might have—and this is why they were so truthful with their favoritism. I do believe this to be the

case; the design of the context did allow the students to be truthful—the responses were likely more accurate and truthful than a low support context would produce. Therefore, the same high support context would have surely compelled the students to *concurrently* show remorse if they had a reason to be remorseful. This did not occur.

In this way, this portion of the study had uncovered a potential cultural divide. What is the best way to discuss these findings? This is problematic on many levels. Not only was favoritism rampant with Japanese students' course evaluations (which can be deemed almost scandalous from a "Western" perspective), but the majority of the students did not shy away from displaying their strong personal biases.

In short, integrity of this sort may not be a standing virtue for these Japanese students; for these students, it seems that this sort of integrity is neither feigned nor is it shameful to lack—this is why the students had little/nothing to be shy about. Yet, as as this thesis has a strong focus on cognitive development, it would be hypocritical if I did not mention that this level of response may instead be age appropriate. Below I will discuss the conclusions of both hypethetical explanations.

If we go down the *cultural affect* route, it can be hypothesized that because the Japanese culture does not prioritize the teaching of "fairness" in the Western sense of the term, the concept is not culturally ingrained. Moreover, because Japan is a neo-Confucian society, with strict adherence to *jouge-kankei* (hierarchal relationship of authority and control), when put in a position of power, the students naturally use this power in a hierarchal way—which can be at odds with the Western concept of fairness. In this way, this seems to be a two-pronged issue: (a) the Western concept of "fairness" is not a virtue in Japan, and (b) when given power within a hierarchy, the power is rather absolute and need not be "fair". And so, if we put A and B together, we can see how it would be unnatural for these students to show remorse over blatant favoritism. So, I am fairly confident with these observations. The derived conclusions are: (1) there was rampant favoritism with course evaluations, (2) there undoubtedly was a "halo effect", and (3) both *OL1* and *OL2* students clearly displayed these tendencies.

If we go down the *age-appropriate response* route (with no significant

cultural affect), the final conclusions are the same as above: (1) there was rampant favoritism with course evaluations, (2) there undoubtedly was a "halo effect", and (3) both OL1 and OL2 students clearly displayed these tendencies. And so, interestingly, regardless of which hypethetical explanation we choose, the ultimate conclusions remain unchanged. (However, although it has no impact on this current thesis, I am unsatisfied and a bit frustrated by having to leave this issue largely unresolved; in the future I would like to establish focus groups to help uncover the reasons behind these responses from the students, if only to satisfy my curiosity regarding students' understanding of "fairness" in this context.)

8.12 Conclusion

It seems that coursebooks have historically (and naturally) taken the most pragmatic approaches possible for the times—typically relying on rote memorization of facts. This was fine when schooling was only about preparing workers for manual labor during the industrial revolution (the reason for the flourishing of schools in the first place). I have discussed a more recent historical perspective of course and syllabus designs in Chapters 2 and 3—but, learning from that, I believe a more progressive design is being called for, especially in EFL contexts where students are *not* in a context that continuously provides them chances to listen to and use English for real purposes.

For the *ESL* learner in an ESL context, there is an on-going barrage of local language usage outside of the classroom that dynamically (consciously and unconsciously) helps create a usable and reliable linguistic system within the learner, based on real needs. This sort of dynamic learning is not only about learning to use the local language, per se, but it is more about learning to be a functional human being within that society—you become capable of fulfilling your real-world needs via appropriate language discourse.

It is therefore imperative for the person to develop higher-order skills that match the context they are going to live in—and our brains have evolved to do this for us in our native language from our native contexts; we just need to be in our

native context long enough for the brain to do its work.

The crux of the issue with EFL contexts stems from this biology; EFL contexts by definition do not inherently provide a context conducive of natural English language development. Moreover, the development of metacognitive skills is of course a staple of higher-order skill development.

In short, native English speakers living in English-speaking countries and ESL learners in English-speaking countries are living in contexts that naturally nurture, and later demand usage of, higher-order skills in English. *EFL* learners, on the other hand, such as Japanese students studying English in Japan, are seldom—if ever—provided with a context that requires them to, nor allows them to, foster deeper thinking in the English language. Complex thought has to be conceived in their native language, and then when called for, those complex thoughts are translated over to English—if they have acquired the appropriate translation skills to do so. If they do not have the high-level translations skills, then not only are these people not in possession of skills to have higher-order English thoughts, but they cannot even translate their deeper Japanese thoughts over to English when necessary.

As a response to this on-going issue within the Japanese EFL environment, *OL2* has been designed to have students enjoy the process of *developing* their own in-depth thoughts in English. This is somewhat comparable to teaching a person to fish in lieu of giving them a fish to eat. It is meant to provide a context that can compensate for the lack of being in an ESL or native-speaking context, instead of providing a quick fix on a micro level that does nothing for the macro level issues.

Regardless of the chosen coursebook's design, the development of metacognitive skills (based on personal needs) can surely be fit into almost any EFL programme—as long as those skills-to-be-nurtured are being used to encourage perceived-to-be-valuable cognition in English. The qualitative data discussed in this chapter can be summed into three major points:

(1) the *OL1* coursebook, which was at the time meant to be as learner-efficient as possible, has at least one major flaw – the potential "critical break" that occurs midway in every chapter. This break seems to have been overly disruptive, and, while good learning did occur, there was not enough of a forward-moving momentum that allowed for knowledge to compound as quickly and mature as efficiently as with *OL2*.

(2) In *OL2*, the realization of personal growth *with* the L2 (empowered by *autonomy* and students' ever-growing *feelings of ownership* of the course content) seems to be the core/archetypical theme that emerged from the students' data. This level of personal growth in English was accomplished by the upgrades that were incorporated into *OL2*. *OL2* requires teaching that cyclically fosters a real need for growth in the chosen domain of thought (the presentation theme) while at the same time having the students realize (on their own) the requirement of suitable lexical items (and/or grammar points) to accomplish this complex growth that is required to complete the task. This three-step process (create a visceral desire for L2 learning, further encourage students' desire to grow in the L2 in a conducive context, provide appropriate tools for deep and rapid growth in the L2) produces controlled—yet organic and rapid L2 maturity.

(3) There is a strong incidence of the "halo effect" in regards to student motivation and course evaluation. The majority of the participants discussed the manifestation of favoritism without a hint of guilt or shame—this potentially reveals a substantial cultural divide with "Western norms", or it could be an age-related response. Either way, even though there were no significant differences were found between the *OL1* and *OL2* groups in this area, only *OL2* instigated rapid L2 maturity.

There was a wide assortment of learning documented in this chapter. I have collected all of the major findings from this chapter below. As with any major study, although the data may be robust, there is always a margin of error, and there is always

room for improvement—from data collection methods to the interpretation of the data that has been collected. Future studies will surely call for reinterpretation of this data and invite revisions of the conclusions. With that caveat in place, here are the major findings from this chapter and their significance, as interpreted in 2019:

- Across the span of the year, students consistently discussed metacognition and deeper-thinking more than linguistic improvement – until the end of the year, when this suddenly flipped, and the students began to talk more about their linguistic improvements.

- Non-initiated first-year students (not in focus groups) were able to provide in-depth metacognitive insights seemingly by virtue of using *OL2* to study English. According to some students, this level of thinking is something they were not used to engaging in even in Japanese. This seems to indicate that the cyclical deep-thinking activities designed to be efficient language teaching approach also provided a pathway to deeper thinking in the students for domains that were not directly touched on by *OL2*. There was plenty of data regarding this in the five student diagrams (Section 8.10.3).

- Rather than proceeding with a typical lecture approach, or a simple provision of lexical item lists for use in a linguistic task, there is a huge payoff if teaching is designed to (artificially) manufacture a strong internal need for lexical items to complete a task that the students have already gained a personal attachment to. The lexical items are only provided to the students after they are "starved" for this help. Students' appreciation for the lexical items provided in this way is noticeably high, as is their motivation to use them.

- When students have been given time to design their presentations, but are still anxious and slightly frustrated with their progress—this is the optimal

timing for the teacher to provide a mini-presentation/talk that has been designed beforehand to alleviate their anxiety by showing tips for better presentations regarding the theme, inclusive of clever usage of the lexical items they are trying to embed into their own presentations. This seems to raise their aspirations toward better presentation skills, better linguistic skills, and it also seems to raise general rapport between the teacher and the students.

- The provision of well-timed teaching items is highly valuable for the entire course flow—this may seem obvious, but the tactics are not simplistic. If lexical items (or grammar points, etc.) are to be taught on a certain day, then the pedagogical cycle of events must include a purpose-designed "deprivation period" that produces a very real internal need within the students for that specific content. In short, the students must literally feel a *very strong need* for those lexical items—they must realize the need—and then the teacher can casually and introduce in a friendly way the lexical items, at which time the student will "jump on the opportunity", and be highly appreciative. This emotional intensity regarding the provision of the strongly desired content is what makes the teaching and learning more efficient than usual. A straightforward top-down lecture or the provision of a vocabulary sheet could never be as efficient as a cyclically designed pedagogical flow that induces the need for the knowledge beforehand. Several quotes from the student's support this concept.

- The frequency of the top five themes across the *OL2* year remained unchanged regardless of the modality. At first this was a surprise, but now it is an anticipated outcome.

- The themes regarding motivation and metacognition were consistently in the top frequencies, regardless of modality. It can be deduced that the

students realized *OL2* was challenging them to think deeply – at times in novel ways.

- *OL2* seems to have been able to sustain a cyclically compounding positive learning state of growth for a significant portion of the students.

- By the end of the year, the students' most frequently written about theme shifted to *how much their English had improved*, as metacognitive themes and motivation-related themes took slightly lower frequencies. However, all three theme genres were always high-frequency in any of the modalities. Upon consideration, this seems to be an ideal conclusion to the year of study with *OL2*.

- There seems to have been a critical break in the flow of *OL1*. My definition of a critical break is a break from which the built-up momentum of study cannot be fully recovered. It seems this is exactly what happens around the middle of the flow of every chapter in *OL1*. Although the concept of horizontal flow design in coursebooks is not new, the definition of critical breaks within chapter flows is new. Future coursebooks designs should keep this point in mind.

- Emotional climaxing occurred as a sort of a bell curve in the middle of every chapter of *OL2,* which had the opposite effect of a critical break within a chapter. Future coursebooks designs should keep this point in mind. The fact that *OL1* had a critical break around the time that *OL2* was having emotional climaxing is probably the most prominently significant difference between the two books—causing massive differences in their outcomes by the end of each term of usage.

- Typical ESL/EFL conversation coursebooks routinely "throw in" pair work

after a learning activity as a matter of course. Although this seems logical, according to my data, the pair work cycle was the cycle that was treated with the most indifference. This outcome can be interpreted in several ways. The major takeaway from this seems to be that the *OL* coursebook students (at some level) have a gut feeling that a task in the form of, "Ok, now talk about this…" is not a well-thought-out activity; the setup is decidedly pedestrian compared to what they have been acclimatized to from *OL* coursebook usage. The activity should, therefore, be much better integrated into the flow of the pedagogy—tuned for much more specific growth goals in mind. For *OL2,* I titled my pair work activity Cycle 2, but from the data, I now realize that this activity does not deserve the "Cycle" title; this is especially so when compared to the other cycles that are clearly more well-developed within the Module lineup. Cycle 2 should either be "demoted" to an *activity*, or, in order to keep the "Cycle" designation, it needs to be reorganized into a much more thematically connected undertaking with an explicit pedagogical complexity on par with the other cycles in the Module.

- The data shows that the sections and activities that were designed to have the most complex group interactions and peer-to-peer assessment (building and sustaining high support contexts) stood out as being the most useful and helpful, according to the students.

- The pedagogical flow of *OL2* seems to be able to nurture and sustain a cyclical and compounding flow of in-depth thinking and growth along with language learning. It seems that for most of the time, traditional "language learning", per se, is secondary; it seems that *personal growth **with** the L2* is the realized and recognized goal. This point is highly significant. Moreover, personal growth *with* the L2 is inherently tied to students' feelings of ownership.

- The data implies that students' *feelings of ownership* and their *personal growth **with** the L2* are naturally reciprocating qualities that are nurtured within the cycles of learning in *OL2*. This natural reciprocation is likely what is at the core of neuro-efficient learning and teaching; it entices the brain to follow a similar pattern of development that their native language took in their native language context. This is specifically why it can be regarded as neuro-efficient, especially when compared to pedagogy that is rote memory-based, then forgotten after the test. The implications of these points should be further developed in future studies. (The pre-frontal cortex can realize which language is being attempted at any given moment, but for the rest of the brain, language is language—but only if it is treated as a language by the brain. Too much language teaching methodology ends up asking to learn the L2 as a school subject, and not as a language akin to their L1. And so, future studies should focus on enticing EFL students' brains to carry on learning the L2 and using the L2 as it already does with the L1, and certainly not as a *school subject* that needs to be learned—and then soon forgotten).

- The "halo effect" was confirmed by data from the majority of the EFL students taking *OL1* and *OL2* courses. The majority of the Japanese EFL students reported that teacher-student rapport had a strong affect on both: (a) course evaluation and (b) on their motivation to do well in class.

- Although the presence and strong affect of the "halo effect" was confirmed in both groups by quantitative measures and qualitative measures, only *OL2* was able to produce rapid L2 maturity. I suspect that the reasons behind the significant difference in efficiency are complex and multifaceted; however, I believe the primary reasons are embedded in conclusions made two points above (regarding students' *feelings of ownership*, and *personal growth **with** the L2*).

9. Emerging tenets and implications

Whereas Chapter 7 and Chapter 8 were keenly focused on data involving the research of *OL1* and *OL2*, Chapter 9 is more of a general summation chapter. The main purposes of this general summation chapter are to: (1) review the pedagogical lessons that were learned from researching *OL1* and *OL2*, and display them in the form of tenets; then, (2) debut thoughts on the upcoming *OL3* design (to be titled "*Level Up!"*) whose modifications were wholly based on those tenets. In the way, this chapter will first provide theoretical ideas, and then it will display practical applications of those theoretical ideas, in the form of *OL3*.

9.1 Implications from the Literature review for the development of OL3

Material writing for language coursebooks has a relatively rich literature base within applied linguistics to rely on. However, much of the older literature is based on tradition, guesswork, and often a misguided understanding of how second language acquisition works. There has been, however, a change for the better in the past few decades. Research methods have improved. And, the science behind language acquisition and efficient classroom learning has become more prominent in the literature. The newer literature and progressive studies have spawned innovative applications that are likely affecting general classroom teaching practices—and leading to better materials writing (Kelly, 2012; Helgesen, 2014).

I have systematically combed through the literature reviews in this thesis (primarily Chapter 2 and Chapter 3) and found the most essential learning points regarding good teaching practices and good materials writing. I have boiled down that content into 19 guiding "classic" tenets for good teaching and good coursebook writing. I have listed them in the next section with concise notes next to each rule.

9.1.1 The 19 "classic" tenets based on the literature review

These 19 tenets are useful as "rules-to-write-by", for coursebook writers and "rules-to-teach-by", for teachers. I used the tenets as a practical analysis tool for this thesis. I used the tenet list as assessment criteria to evaluate and critique a full chapter of *OL2* (plus peripheral pages such as the "Think Again" page after the ending of the chapter). As such, this literature-driven, "classic" tenet list became a principal tool for this thesis.

This, however, was not my initial intent. My initial intent was to harvest the primary substance from the literature and to simply document the content as a part of my thesis. But, when the resulting list of tenets was closely examined, it became evident that the tenet list turned out to be a tailor-fit instrument—for this particular context. The list unexpectedly became a natural guide for coursebook and pedagogical assessment. With this literature-derived (formidable) instrument in hand, I was able to make a far-reaching analysis of *OL2*. All of the major "classic" violations that were uncovered during the evaluation of *OL2* were documented and have all been resolved in the designs for *OL3* (*"Level Up!")*. Those changes are highlighted in the second half of this chapter.

The 19 "classic" tenets

- ***Provide choices.*** Choices can automatically raise student motivation via a feeling of ownership toward chosen activities or chosen topics.

- ***Add emotional content.*** Boring information goes nowhere in the brain. Teachers much embrace the idea that emotional process building is at the core of all effective learning and teaching. Passion-based teaching leads to more engagement and confidence; this leads to increased diligence and increased willingness to study and learn more.

- ***Add "solvable mysteries".*** It is very rewarding for a student when they can solve a mystery that is just beyond their immediate understanding. Solvable mysteries should be the building blocks of coursebook design.

- ***Add potential "Aha moments".*** When students are able to reorganize their understanding and level up, they naturally feel good about it. These aha moments fuel the motivation for higher-level learning and understanding.

- ***Add empowerment activities.*** Prime examples of such activities are peer-to-peer assessment and peer-to-peer teaching. Empowerment activities make the learning more meaningful and more dynamic for the students.

- ***Add activities that are semi-controlled by the learner.*** Make sure the students realize that the teacher is allowing the students to have this control over their own learning.

- ***Focus on real-time issues and mistakes; ensure regular teaching content updates.*** Eliminate preemptive guesswork—the learning content should match the students, not the other way around. Conceptually similar to formative assessment, classroom teaching content should be based upon real-time outcomes and contextual variables—and not pre-emptive guesswork made before meeting the students.

- ***Add emphasis on social learning.*** However, do not remove all grammar exercises. (Moreover, grammar exercises can be adapted into social learning exercises so that social learning and grammar exercises are not mutually exclusive.)

- ***Ensure vertical coherence and horizontal coherence in the coursebook design.*** Not only is this more aesthetically pleasing, but it is also more logical for the students to follow.

- ***Make the goals clear.*** There are two sets of goals that should be made clear (but are often neglected). They are the course learning goals and each student's personal learning goals. Goal clarification of both sets can dramatically help raise focus and motivation.

- ***Add a strong sense of closure per unit of study.*** Do not allow a unit of study to fizzle away. Students must experience a real and meaningful sense of purpose for their studies; closure per unit of study with a climactic sense of mastery is ideal. This should be allowed to happen before moving on to the next unit of study.

- ***Add generative lexical concepts.*** To expedite L2 maturation, lexical items in the classroom should be treated as they have their own generative grammatical properties and unique nuances based on context. (Instead of simplistic rote-memorization of vocabulary.)

- ***Add "word maps" and "lexical item grouping" activities.*** Do this before sentence-level writing activities.

- ***Allow confidence to grow with "small group work" before "big group work".*** Allow students to build ample confidence with small group activities before large classroom-level activities such as presentations.

- ***Challenge students' intellect in the L2.*** For maturation of the students' L2s, have less focus on rote memory and more focus on critical thinking in the L2.

- ***Revisit content for meaningful and compounding growth—make sure it is not "just a review".*** Every time the content is revisited, the teacher should push the students' understanding a bit further than the previous visit. Compounded learning is learning maturation.

- ***Question: Is the learning well-arced?*** Content should not be simply "thrown" at students—expecting them to make sense of it. The learning points need to be logically and purposefully connected.

- ***Question: Will the learning tasks activate the teacher-student relationship?*** Autonomous learning in this thesis' context does not mean quiet study time—the students should be able to engage in fruitful co-creation of knowledge with the teacher.

- ***Question: Does the learning task have a real-world application?*** If the answer is no, what is the purpose of the task? For maturation of the L2 in an EFL context, no activity should be a wasted activity. A task that has no real-world application should be scrapped or re-written.

9.2 Implications from my own data (toward the development of *OL3*)

Concurrent with the development of the "classic" tenet list, I implemented a second set of "positive" guidelines, elicited directly from the student data of my studies. What emerged was a 20-point list that duly complements the "classic" list. In this way, I produced two tenet lists—one from the literature review, and one from my student-based research. The student-based tenets are primarily derived from prominent themes and explanations received from students discussing the efficacy of *OL2*.

Was this second list warranted? As good as that classic list was, it was based on literature that had been written before the conception of *OL1*. In contrast, I had carefully planned my primary and secondary studies to elicit a well-defined set of instructions for the improvement of *OL2*. In other words, the second list is a purpose-

designed tool made specifically for this assessment project. Perhaps it can be argued that this second list is more important for the purpose of this study—with its tight focus on the pertinent issues of the context. However, in practice, both lists were valuable and helpful in their own ways.

9.2.1 The 20 tenets based on EFL student data

In the section, I have listed 20 new consequential tenets that were uncovered from the student data. With this new list at hand, I assessed *OL2* a second time. The finalized results of the assessments from *both* lists are detailed in the next section. I have added notes to the tenets to clarify how they apply to this particular context.

The 20 "new" tenets

- ***Provide more "thinking" activities in the L2 than basic "linguistic" activities.*** To develop L2 maturity, metacognition and deeper-thinking in the L2 is more important than pure linguistic improvement.

- ***Provide a wide range of cyclically connected activities.*** *OL2*'s cyclically connected deep-thinking activities provided a pathway to deeper thinking in the students for domains that were not directly touched on by *OL2*. Prioritize wide-ranging (often serendipitous) growth for deep-rooted L2 maturation.

- ***Manufacture a strong need for lexis—then provide them like presents.*** Ideally, lexical items should only be revealed after the need for them has been manufactured. There is a huge payoff if teaching is designed to create a strong internal need for lexical items before they are revealed.

- ***Provide lectures (top-down teaching) only when students will strongly welcome them.*** The optimal timing for a teacher's lecture or presentation is when the students are anxious about a learning task. Conversely, lectures are not effective when students have yet to realize a strong desire for that content.

- ***Provide novelty; challenge students in novel ways that students will appreciate.*** Students realized *OL2* was challenging them to think deeply – at times in novel ways. This is why "metacognition" and "motivation" were common points of discussion in the focus groups and in the reflective writing. Novel challenges seem to be an effective component for classroom motivation and deeper task engagement.

- ***Design the syllabus to be able to sustain cyclical compounding learning tasks.*** OL2 seems to have been able to sustain a cyclically compounding learning state that expedited automaticity and the capability to adlib responses for a significant portion of the students. Innovative coursebooks should be designed to sustain a flow of learning momentum.

- ***Do not allow critical breaks to happen to students.*** (*OL1* likely had critical breaks; *OL2* did not.) Long vacations can easily become critical breaks for students, but so can badly designed coursebooks.

- ***Provide a "Performance-based learning climax" in each chapter of study.*** A performance-based learning climax (or, PuBLiC) occurred in the middle of every chapter in *OL2*, then continued to the end of the chapter. In other words, there was no sole climax at the end of the chapter. This flow maintained a high level of motivation.

- ***Do not simply throw in "pair work for discussion" after each learning activity.*** This sort of pair work is not efficient (specifically regarding Cycle 2 in *OL2*, as found in most EFL coursebooks as well). Group/pair work should be intricately designed to be maximally engaging. The data clearly shows that the chapter sections that were designed to have the most complex group interactions and peer-to-peer assessment (building and sustaining high support contexts) stood out as being the most useful and helpful, according to the students (Cycles 4a, 6, and 8 are good examples of this).

- ***Ensure students' realization that "Personal growth <u>with</u> the L2" is the ultimate goal.*** Personal growth with the L2 is inherently tied to feelings of ownership. (Personal growth in the L2 is maturation of the L2.) These are naturally reciprocating qualities that are nurtured within the cycles of learning built into OL2. This natural reciprocation is likely what is at the core of efficient learning and teaching.

- ***Disclose the pedagogical strategies to the students so that they are more self-aware of their learning and their learning process.*** In retrospect, it seems odd that many teachers don't disclose the pedagogical processes and theory behind the methodologies that are used in the course. If the students are "in on" the strategy, metacognition will deepen, and students will naturally come to appreciate the teaching more. This should help accelerate and deepen the learning.

- ***Regularity and unwavering cyclical designs lead to quicker automaticity in L2 skill building and L2 usage.*** Neural networks are literal structures in the brain; in the classroom, well-designed cyclical pedagogical tasks help build those literal neural pathways more quickly because regularly repeated usage of networks make their own response times quicker (due to myelination, etc.), as well as helping to expedite the creation of supportive networks.

- ***Be aware of "rapport with students" and how rapport affects student motivation, evaluation, and learning potential (halo effect).*** My studies clearly showed that the halo effect is real; the majority of the students studies reported strong affect factors regarding motivation and evaluation.

- ***Teach with "L2 Maturity" in mind, with a good understanding of students' L1 competence.*** Learn about each students' L1 capabilities -- so that you never mistakenly require a student to be able to accomplish an L2 task that that is too complex for them to complete in their L1. This mistake will lead to unnecessary stress for everyone involved.

- ***Assess and track L2 Maturation rates per student.*** Learn how to evaluate students on how quickly their are catching up to their L1 competencies, rather than have students compete with peers on arbitrary tasks set up by people who do not understand cognitive growth patterns. Striving for students' L2 to catch up to their L1 should be included in the criteria for almost any EFL syllabus.

- ***Provide "focus group" tasks in and/or outside of the classroom.*** Focus group discussions were found to be cathartic for the students because they were put in a comfortable context that allowed them to open up, and because they felt a real purpose. Real purpose and feelings of responsibility tap into innate feelings of altruism. These tasks make L2 learning a much deeper and meaningful (and perhaps even life-changing) experiences.

- ***Timing within cycles can be just as important as the content of cycles.*** Unwavering cycles are key to automaticity of skill building, but the studies showed that simple repetition of good ingredients alone was not sufficient. The complex dynamics of optimized timing should not be underestimated. The OL1 vs. OL2 studies made this abundantly clear.

- ***High support contexts nurture deeper thinking in students, but they are difficult to design.*** The studies indicate that coursebooks can be designed to effectively foster high support contexts in the classroom. That said, OL1 failed to accomplish this goal while *OL2* did accomplish this goal. It took years of research to realize the affective differences between the two books.

- ***Allow for critical self-review of past work.*** All to often students recieve marks on an assignment and then the course moves on. Little learned from this conclusion of events. Instead, this should be a learning moment. An essential component of an effective high support context is the allowance of students' critical self-review of their past work -- and allowing them to upgrade their work for a second review by their peers and the teacher. Evaluation should focus on how much dynamic improvement has been demonstrated by the student rather than pre-set goals that cannot take into account each student's idiosyncrasies.

- ***Establish explicit "learning from peer" tasks.*** The classroom is an artificial environment. In the real world, people learn language from interaction with other real people. Similarly, EFL students (who do not have ESL students' learning opportunities) should be given explicit learning time from peer demonstrations. Not only is this natural for the brain, but peer learning creates real bonding opportunities and a positive learning environment for the students.

When combined, the two tenet lists serve as a catalogue of standards for progressive EFL coursebook writing and efficient pedagogical design. As both of these lists were used to revise *OL2*, the entire evaluation process became one large aspect of my research and provided many of the deeper implications of this thesis.

Another substantial aspect was the group of specific problem areas in *OL2* that needed to be rectified; finding those negative aspects of *OL2* were also an integral part of my study. So, in this way, there emerged a solid list of "positive standards" that needed to be upheld, as well as a list of undesirable areas/mistakes in the *OL2* design (that need to be excluded from *OL3*). The next section details those undesirable areas in the form of *proposed modifications*.

9.3 Proposed modifications for OL2

From this work with the tenet lists, there emerged 35 modification proposals for the transformation from *OL2* to *OL3*. Three tiers of modifications emerged from the group of 35 proposals. The main modification was so substantial that it merits its own tier. This first tier is discussed in Section 9.3.1. The second tier of modifications entails the deletion of pages and/or changes to pages that substantially affected the

flow of events (Section 9.3.2). The third tier is the collection of the remaining modifications.

9.3.1 Main modification tier

The most significant proposed modification that was made for the optimal development of *OL3* (*"Level Up!")* is highlighted in orange in the flowchart below (Figure 94). This highlighted modification requires four pages of additions to the chapter design. The modification idea is simple-sounding, but the advantages of the implementation are likely to be substantial.

How does this modification work? After the group presentations and the peer-to-peer assessments are completed (Figure 94), the presentation groups get back together and discuss how they should *"level up"* their presentation. This presentation upgrade is based on the peer-to-peer assessments at hand (in their books), and their own impressions and thoughts from all of the presentations that they had just witnessed that day. According to the data, this modification brings an evocative change to the air and flow of the course—and therefore to the outcomes. I decided to name this coursebook version after this primary modification—hence, *"Level Up!"* will likely be the publishing title for *OL3* when it goes to print.

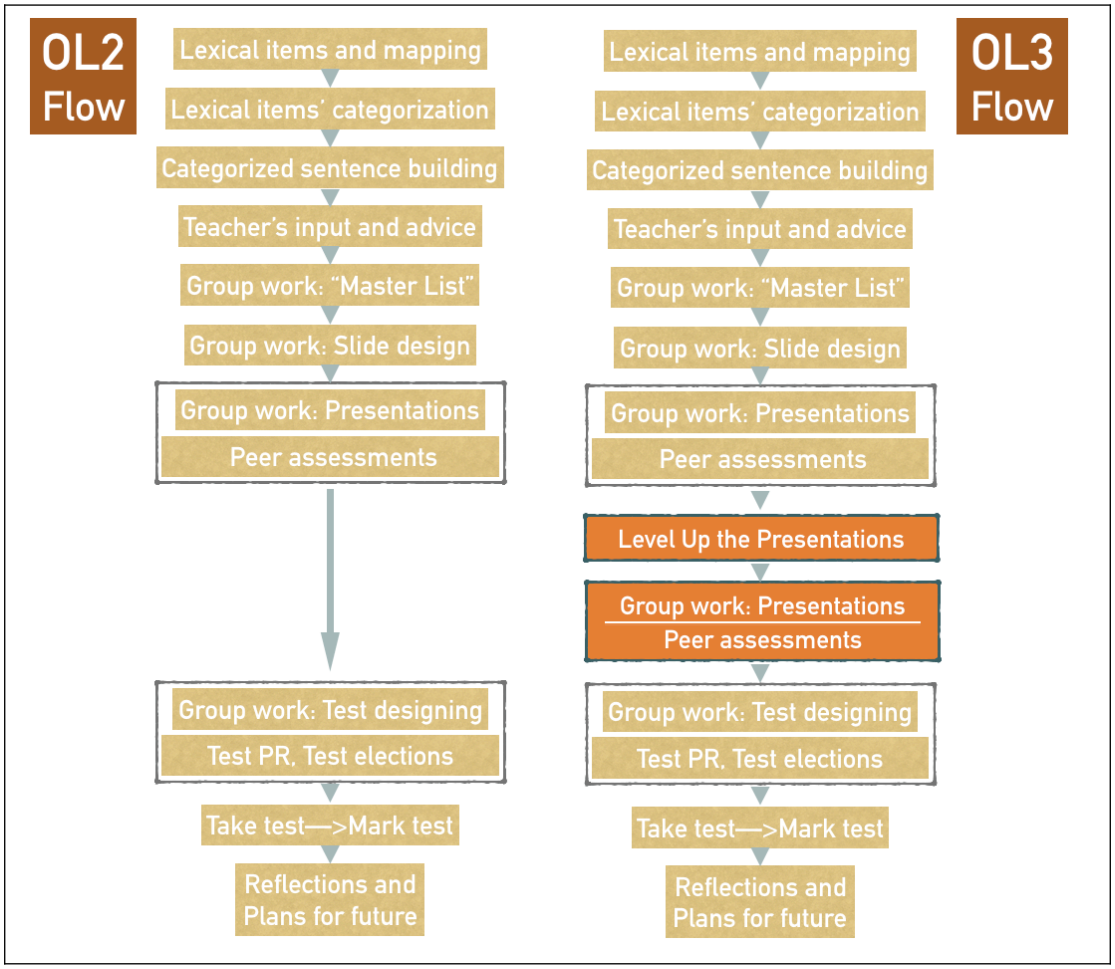


Figure 94: Comparing two flowcharts, OL2 vs OL3 ("Level up!")

Having the students notice what they *could have done better* immediately after they watch their peers perform, seems to trigger an intimate and very real feeling of "self-responsibility" toward what they have been doing with this presentation project. It seems to engage something primal inside of them which is not unlike playing a video game where they "die" but immediately resurrect to jump into action again. If they are even slightly "addicted" to the game, then they are ready to have a second go at it with a resilient motivation to perform better this time. This motivation to do better at the game, and do better in the classroom should naturally provide an adrenaline rush in most cases. The flow of events in this *leveling up action* in a video game and in the coursebook is highly relatable to the university students because they are accustomed to "falling into" this mode.

However, what I propose here is not a simplistic gamification of the content.

From the academic perspective, this modification literally doubles the presentation time per topic. This is a substantial change to the student performance time per semester. Due to this modification, instead of having three presentations on three themes per semester (15 weeks), the addition of this *leveling up* to the syllabus requires students to have FOUR presentations on two themes, per semester (15 weeks). By the end of the year, each student will have completed (and peer-assessed) eight presentations instead of six presentations—albeit, with a total of four themes instead of six themes.

These numbers highlight a trade-off; doubling presentation time necessitates one less theme per semester, so only two themes can be explored per semester. But, having four presentations *leveled up* to create a total of eight presentations per year turned out to be a much better strategy than six presentations that do not implement *level up* sessions.

In short, *OL2's* strategy was wide but shallower (no upgrades). On the other hand, the proposal will have *OL3* users go deeper with the content (consistent upgrades) with a more focused approach, albeit with less topics. Why is this important? The latter pattern seems more appropriate for L2 maturation purposes in the EFL context.

Although *OL2* has proven to be a strong performer as an EFL coursebook, the key element that was missing from *OL2* was the *leveling up* of the student presentations. As writers know, with any form of writing, the first draft is seldom their best possible achievement. Similarly, although it is still relatively uncommon in EFL contexts, student should be given the chance to *level up* their presentations, and mature from the ensuing insights and accomplishments that could not have been achieved in any other way.

The practical experience the students gain from the *leveling up* task, and the confidence that they gain from being given time to level up will likely compound into a culmination of senses including pride, motivation, peer-to-peer bonding, and positive teacher-student rapport. This occurs across two semesters. I hypothesize that if tested, the Lectical assessment growth scores for an *OL3* (*"Level Up!"*) test group

would show a significantly higher rate of L2 maturation than that of the *OL2* test groups. This experiment will therefore be one of my first post-thesis experiments.

Let us examine the justifications for this leveling up modification, with the two tenet lists as a guide (Table 54). From the tenets, the quality and quantity of the theoretical justifications that underlie this modification are impressive. There is a total of twelve tenets from the two lists that support this modification. Six are from the first list and six are from the second list.

Table 54: Justifications for the major modification—the leveling up task

Justifications: Corresponding tenets from the first list (Literature review):

- ❑ *Provide choices.*
- ❑ *Add potential "Aha moments".*
- ❑ *Add empowerment activities.*
- ❑ *Focus on real-time issues and mistakes; ensure regular teaching content updates.*
- ❑ *Allow confidence to grow with "small group work" before "big group work.*
- ❑ *Revisit content for meaningful and compounding growth—not "just a review".*

Justifications: Corresponding tenets from the second list (My own data):

- ❑ *Provide more "thinking" activities in the L2 than basic "linguistic" activities.*
- ❑ *Provide a wide range of cyclical activities.*
- ❑ *Design the syllabus to be able to sustain cyclical compounding learning tasks.*
- ❑ *Do not allow critical breaks to happen.*
- ❑ *Do not simply throw in "pair work for discussion" after each learning activity.*
- ❑ *Ensure students realize that "Personal growth with the L2" is the ultimate goal.*

As discussed earlier, this first tier modification for *OL3* (*"Level Up!"*) will surely have the largest impact out of all of the modifications; it radically changes the flow that was set by *OL2*, but by doing so, it adds substantial L2 maturation potential

for the students.

To test my hypothesis, I piloted this leveling up idea in the second semester of the 2017-2018 year. The participants quickly and noticeably become more excited about their presentations the second time around, making the event more climactic than it ever was during the standard flow. With my 10+ years of teaching "TEFL with presentations" at the university level, I can state with no inhibitions that this modified semester (pilot test) produced upgraded performances that were collectively the most engaging and interesting presentations I had ever seen from university EFL students.

9.3.2 Other noteworthy modifications: the second tier.

There are several noteworthy modifications that distinguish *OL3* from *OL2*, and there are two types of these modifications. *Type 1 modifications* are changes that were made to directly comply with complaints from the students. They are meant to fix faults in the design that were uncovered in the discovery process with the students, in the primary and secondary studies. *Type 2 modifications* are changes that were made to comply with issues that were discovered *after* the studies were completed. These modifications answer the questions, "Based on my new knowledge, what could have been done better?" And, "What is missing here?"

Type 1 modifications:

The focus groups and the writing studies (Chapter 8) uncovered that there were three Cycles in *OL2* that were unpopular with the students. They were Cycles 3b, 5b, and 9. Of those, Cycle 5b was clearly the most unpopular Cycle.

- Unpopular Cycles: 3b, 5b, 9
- Deleted and replaced Cycles: 5b
- Cycles modified to comply: 3b, 9

Cycle 3b: [Emotional valence activity missing] → This task (Figure 95) is important in the flow of the Module, but students likely do not see the relevance yet. Rather than "lecture to them" why this page is important at this stage, I have applied the emotional valence tenet here, to give the students a more personal attachment to this task. This task now (on the right) requires the students to add emotional valence to each of the sentences they come up with, per lexical item. In this way, students will have personal justifications, based on emotions, per sentence. This may appear to be "gimmicky" in getting the students to like this page—but it is good pedagogy nevertheless. The students will not only feel more attachment to this task and their sentences, but they will likely develop better recall of the sentences.

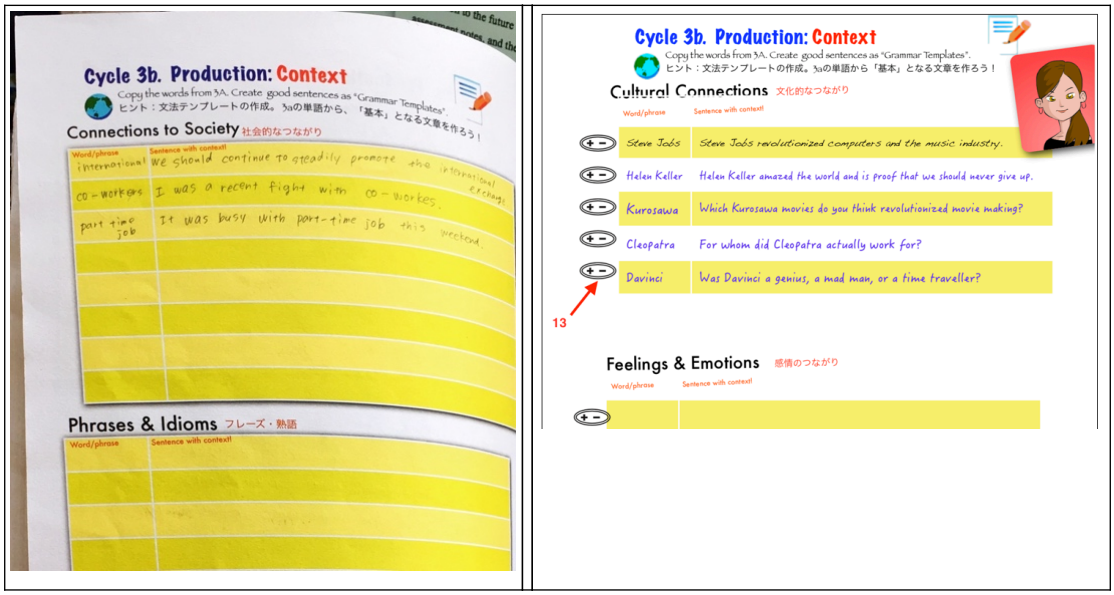


Figure 95: Modification set A (right side is the modified version)

Cycle 5b: deleted from coursebook

Cycle 11: [correction of bad background color choice; horizontal coherence, connecting to the future] → (1) Page color changed to white (Figure 96, right side) for improved legibility in the student-marked chart. (2) By now, after working in deep thinking modes, this page seems to be a bit too simplistic for them *and* it did not provide engaging horizontal coherence, nor a strong connection to the future for them. So, for this modification, I ask students to view the assessment notes, and then write down *"Ideas for my next presentation."* This transforms the task to a more personalized challenge and it creates forward momentum with the otherwise static task.

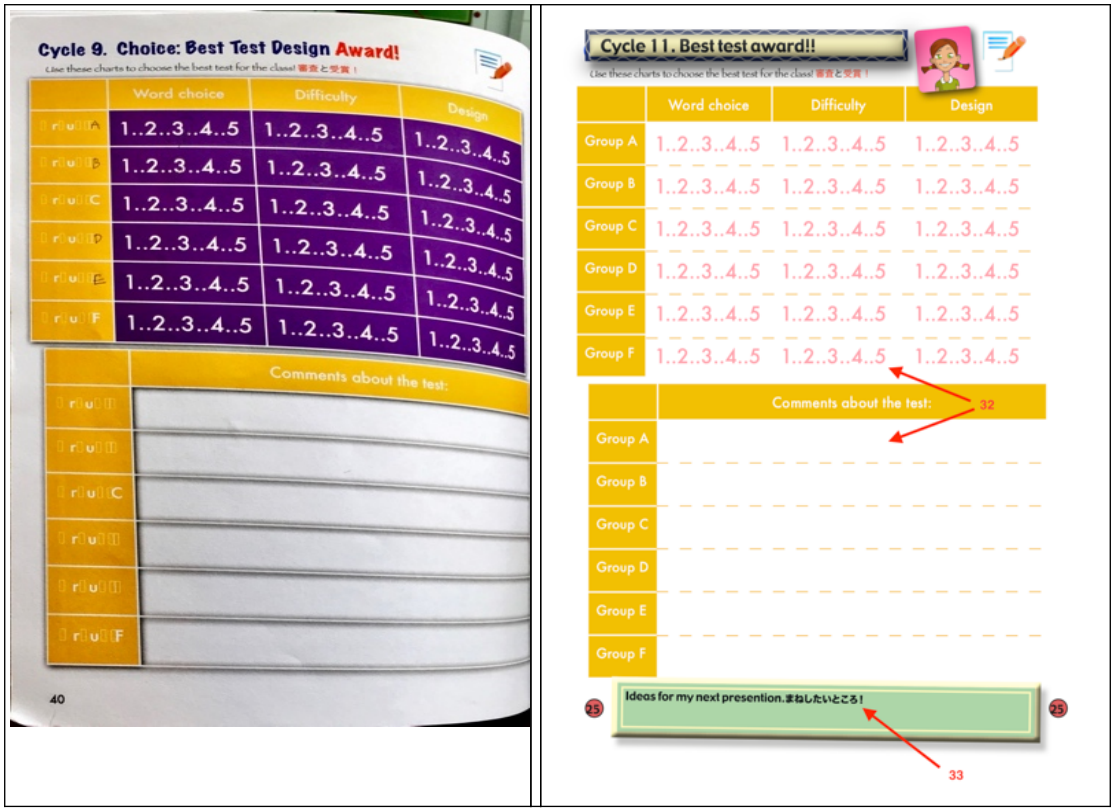


Figure 96: Modification set B (right side is the modified version)

Type 2 Modifications:

With the tenet lists as guidelines, several Cycles stood out as candidates for improvement. Below is a list of the tenets and the Cycles that were modified in light of those specific tenets:

Cycle 2: [Do not *simply* throw in "pair work for discussion" after each learning activity] → The original design in *OL2* had a simple paired discussion at the end of this page (Figure 97). This is exactly what the tenets call out as bad design. For the revision, questions were added here to make it a reflexive task. Moreover, it also asked the students to then choose which skills they need to improve. It therefore, starts them off (on the very first day) on a trajectory for improvement, instead of the limping start of *OL2* (that was without discussion guidance nor an explicit connection to the future).

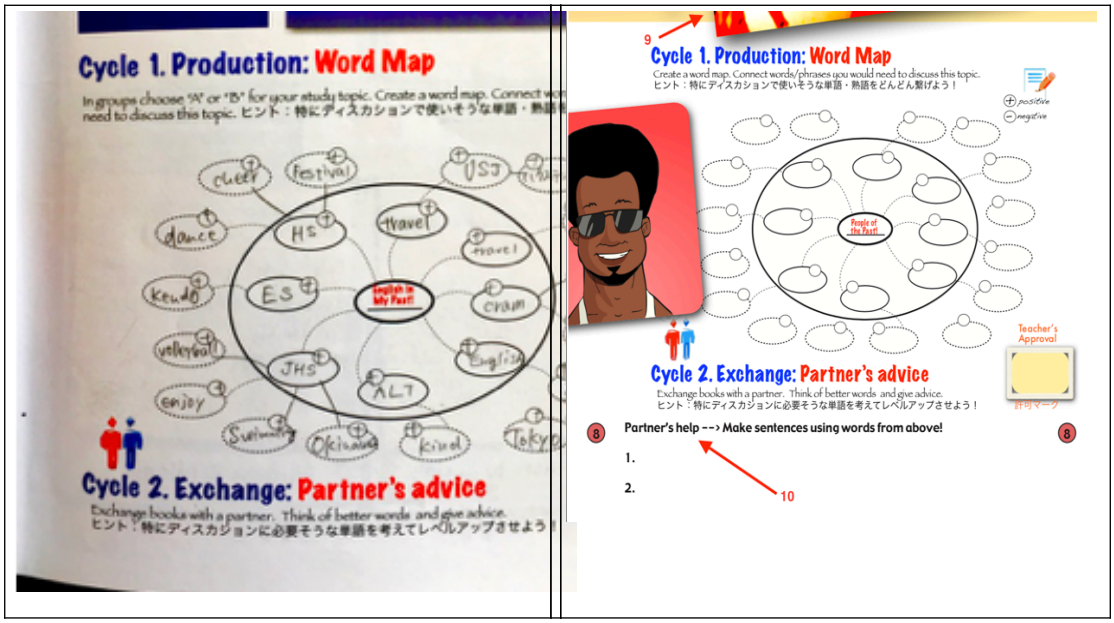


Figure 97: Modification set C (right side is the modified version)

Cycle 4b: [Do not *simply* throw in "pair work for discussion" after each learning activity] → Being a new Cycle, there is nothing to compare this one to (Cycle 98). Building upon previous "Partner's help" tasks, here the partner is being asked to not only review the student's work, but to choose the best lexical items for

the student for the upcoming presentation, and consider how they might best be used. It is an excellent example of how to add pair work after an exercise that is not a simple talk-to-your-partner task.

There are three "index card" note-taking areas on the page. The first card asks the partner to look at the previous work and choose the best words for the upcoming presentation. The next card asks the partner to think and write three sentences for the student, using the words from the first card. The third card goes so far as to ask the partner to think of innovative ideas for the student's upcoming presentation. This is a radical departure from a "get advice from your partner" task because it is highly structured (three tiers) and designed to be motivational. These tiered tasks nurture empathy and create a connection to the future with horizontal coherence, and fosters innovative/creative ideas for the student's upcoming presentation.

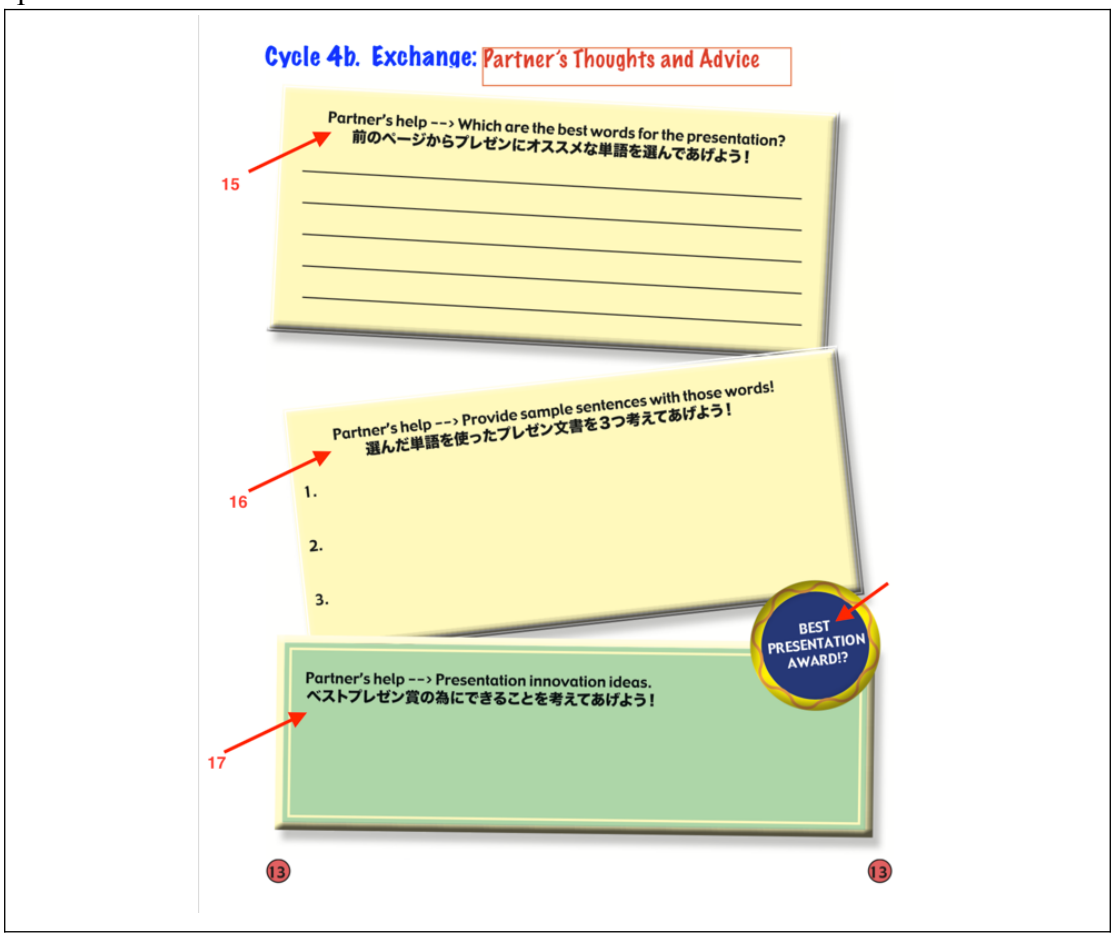


Figure 98: Cycle 4b modification

"Think Again" Page: [Do not *simply* throw in "pair work for discussion" after each learning activity; add social learning; add a clear ending to the chapter] → In *OL2*, this was a simple "quiet reflection" task (Figure 99). Although such tasks are commonplace in coursebooks, according to the students, it was not optimally engaging. Some students skipped this page altogether.

With the modification, this "quiet" page as now been re-tasked as a deep-thinking interview task. It has become a social learning event that clearly marks closure of the chapter—while looking toward the future too. By now the students expect a certain level of "excellence" in partners' responses—and so they are likely to mutually provide more in-depth answers than they were accustomed to providing in EFL courses prior to this one.

This may not seem like a noteworthy change, but this modification is far from being a simplistic change in regards to L2 maturation. With this modification, theoretically, this page now accentuates feelings of empathy, helps in metacognitive development, nurtures feelings of ownership, and naturally asks students to be more autonomous with their L2 learning—all of these components help intensify L2 maturation.

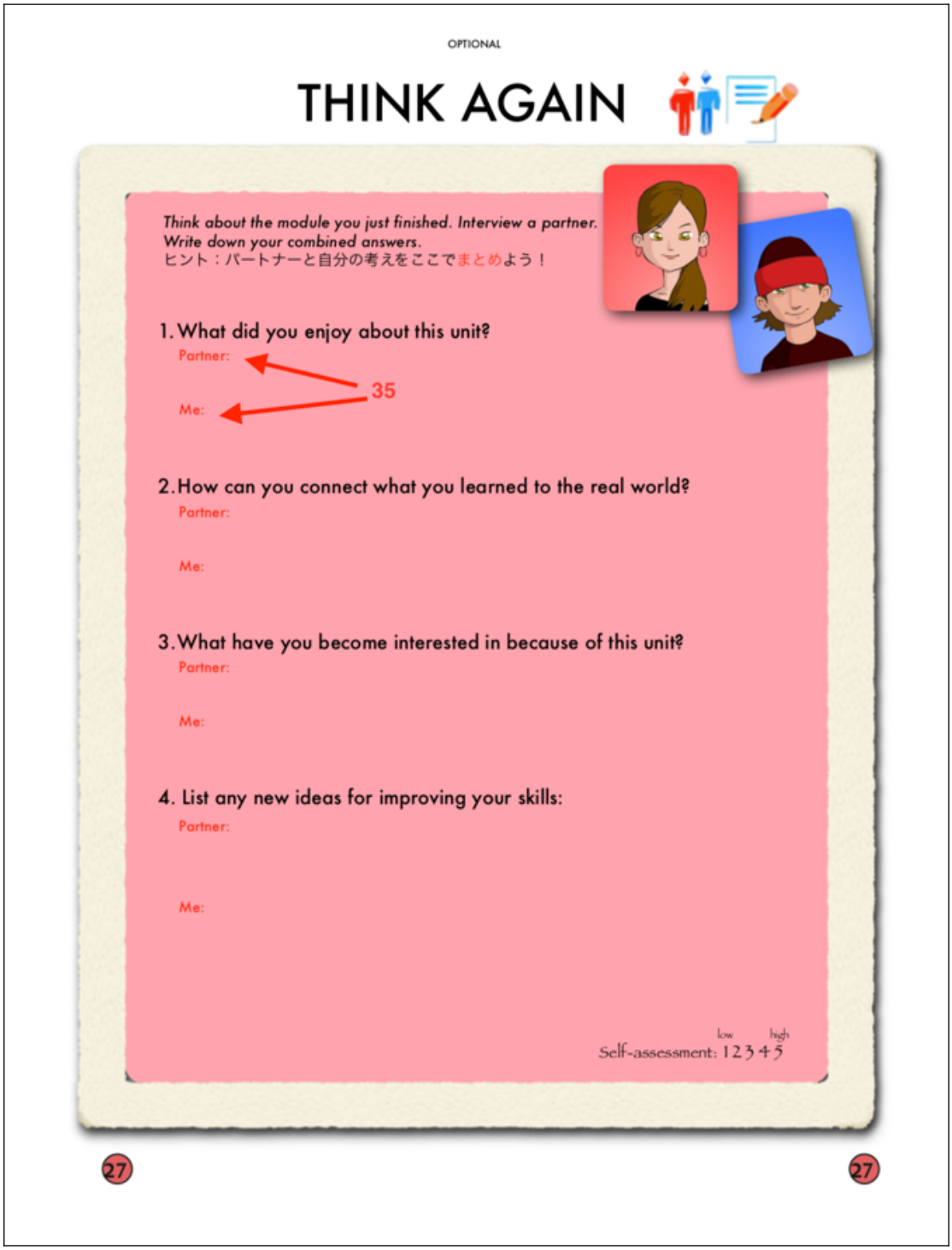


Figure 99: “Think again” modification

9.3.3 Other minor modifications: the third tier

The tier of modifications for *OL3* can be called minor modifications. However, collectively, they have a strong impact on the overall design of *OL3*. Although this newest coursebook design has yet to be tested in the way *OL2* has, with the data uncovered by this thesis, I hypothesize that the third-tier modifications will meaningfully strengthen the overall pedagogical affect of *OL3*. I have gradually gained confidence in this notion because collectively, these well-principled modifications (based on the data from this thesis) carry the essence of the newest "best approach" that has emerged from this research project.

Similar to the other noteworthy modifications that were highlighted in the previous section (9.3.2), these third-tier modifications were motivated by two areas: (1) explicit feedback from students requesting changes [Type 1], or (2) my own concerns regarding areas in *OL2* where the positive tenets' list could be better reflected [Type 2]. As they are minor modifications, instead of listing them here, I have listed them in the Appendix. However, for a visio-chronological perspective, the next section graphically represents the details all *OL3* modifications in a page-by-page graphical format.

9.4 Conclusion

In materials design, the "proof of the pudding" is in the outcome of the designed materials. Therefore, this was the underlying assumption of Chapter 9. This is why, after presenting the established theoretical tenets, comes a detailed summary of the practical ideas that emerged (revolving around the idea of modifying *OL2* to *OL3*). In this section, it would not make sense to reiterate all of the detailed points. Instead, what follows is a graphic summary of the 35 modifications made for *OL3*, with their primary justifications below them. The graphic layout spans Figure 100 to Figure 104. All of the 35 modifications are being presented in chronological order (following the exact flow of the new coursebook, page-by-page).

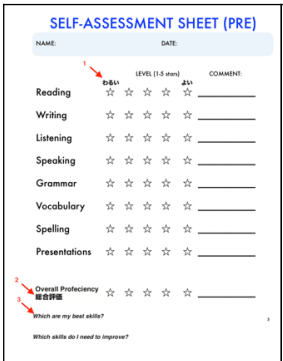
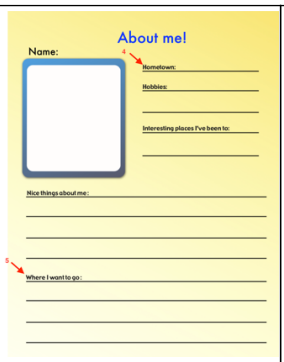
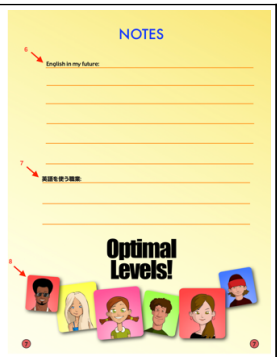
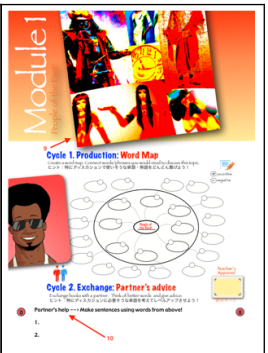
			
• Clarify goals • Provide more thinking activities • Personal growth in the L2 as goal	• Clarify goals • Aha moments • Solvable mysteries	• Horizontal coherence • Aha moments • Solvable mysteries	• No "simple pair work" tagged on • Social learning • Empowerment • Novelty

Figure 100: Modifications for Self-assessment to Cycle 2

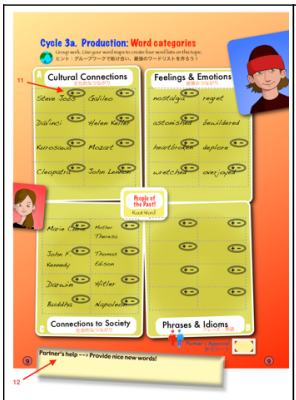
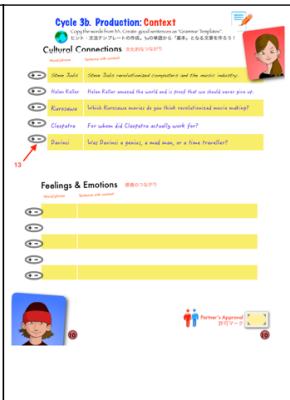
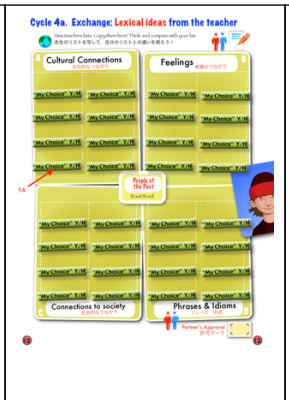
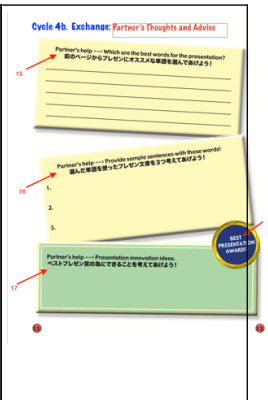
			
• Emotional valance • Social learning • Challenge intellect in L2	• Emotional valence • Add more thinking activities • Personal growth in the L2 as goal	• Empowerment • Aha moments • Solvable mysteries • Teacher-student rapport building	• No “simple pair work” tagged on • Social learning • Empowerment • Novelty

Figure 101: Modifications for Cycle 3 to Cycle 4b

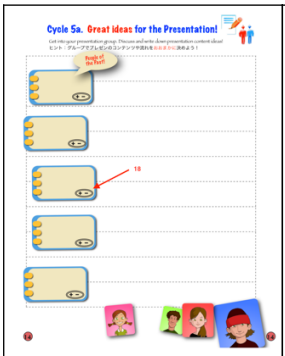
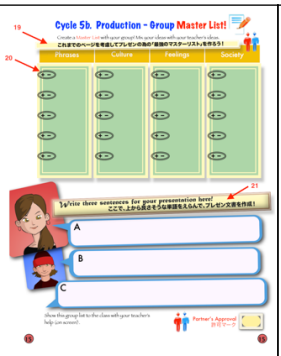
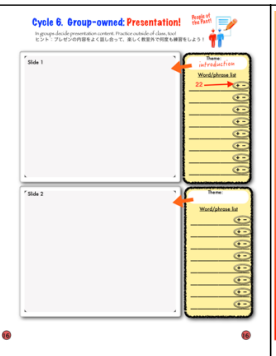
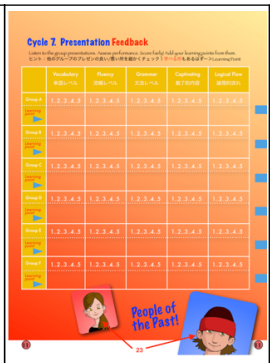
			
• Emotional valence • Horizontal coherence • Clarify goals	• Emotional valence • Horizontal coherence • Aha moments	• Horizontal coherence • Aha moments • Solvable mysteries • Social learning (group negotiation)	• Add emotional content (aesthetically pleasing design)

Figure 102: Modifications for Cycle 5a to Cycle 7

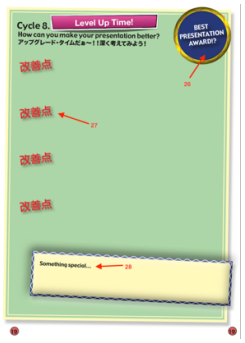
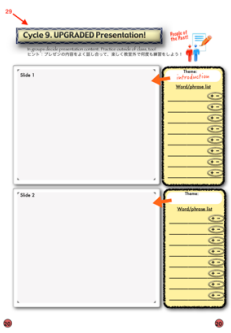
			
• Clarify goals • Provide more thinking activities • Teacher-student rapport building • Empowering	• Clarify goals • Aha moments • Solvable mysteries • More thinking in the L2 • Personal growth in L2	• Horizontal coherence • Personal growth in the L2 as goal	• Clarify goals • More thinking in the L2

Figure 103: Modifications for Cycle 7b to Cycle 10

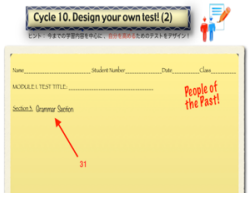
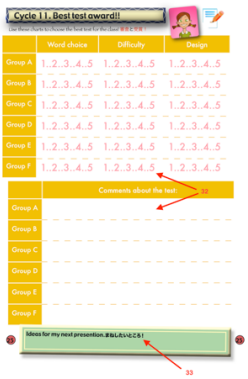
			
• Clarify goals • Provide more thinking activities in the L2	• [fix the color] • Clarify goals • Aha moments • Solvable mysteries • Horizontal coherence	• Horizontal coherence • Clarify the goals • Personal growth in the L2 as goal	• No “simple pair work” tagged on • Social learning • Empowerment • Novelty • Personal growth in the L2 as goal

Figure 104: Modifications for Cycle 10 to “Think again”

This concludes the reporting of the theoretical and practical learning points from this research project that have made their way into *OL3*. The next chapter is the Conclusion (Chapter 10). In it, I will depart from the realm of EFL coursebooks to summarize the broader theoretical issues that emerged from this exercise, and the pedagogical implications from them.

10. Conclusion: The implications that go beyond textbook design and assessment

Unlike the previous chapters that focused on coursebook designs and their evaluation processes, in this final chapter I will focus on the broader themes regarding *good teaching* that emerged from these studies; the lessons that I have learned via the research in this thesis have implications that transcend coursebook writing.

Chapter 9 serves as a showcase of practical applications of my findings. Therefore, in this chapter I will focus on the macro-level lessons learned, their results, and ensuing propositions. In Section 10.1, I will discuss my conclusions regarding *L2 Maturation*. This is a term that grew out of my discussion of how I used Lectical Analysis in a unique way to measure L2 growth in EFL contexts. That discussion makes a good segue to Section 10.2, in which I discuss metacognitive development—in this case, how the learners benefit from *thinking about thinking* in the L2. Then, in Section 10.3, I shall discuss "good teaching" from the prospect of having the proper balance between top-down teaching and bottom-up teaching. Next, I will continue on with the "good teaching" theme with discussions of the COOPs protocol, *"Teaching for the DATC"*, and the overlap the tenet lists have with neuroELT maxims. I will discuss the limitations of my research in 10.6. Future directions will be in Section 10.7. In that future directions section I will put special emphasis on the motivational dimension of coursebook materials with the application of concepts from research in directed motivational currents. The final words will be in Section 10.8

10.1 L2 Maturation: Lectical assessment

When coursebook efficacy was test with Lectical Analysis methods (*OL1* vs. *OL2*), the *OL2* group showed an 8x faster rate of demonstrated growth over the *OL1* group on the Lectical scale. ANCOVA in SPSS was used to adjust for potential differences between the two groups, and yet the data was clear – the *OL2* group demonstrated a much faster maturation rate compared to *OL1* (in a 15-week single semester of study).

The results were staggering when they first arrived from Lectica—my first instinct was that they were too good to be true. And, I was worried that when published, it may seem as though I had altered and/or misrepresented the results. However, upon consultation with the co-founder of Lectical Analysis (Theo Dawson), it was confirmed that *OL2* did indeed produce data showing maturation that was 8x faster than *OL1*. But, as the *OL1* group speed mean was 1/2 that of the mean growth for native English speakers in the same age group in the United States, it can be said that *OL1* did not produce "bad" results—it is just that the results did not show as much pedagogical efficiency or learning depth as *OL2* did. So, in terms of L2 Maturation, we can say that *OL2* nurtures L2 Maturity more efficiently than *OL1* does.

From the data, it seems that *OL2* has achieved its most important goal. The *OL* series was, from the very beginning, a design that was meant to teach more than just English – it was meant to nurture what I have come to call *L2 Maturity*. I will define L2 Maturity in the next section. First, I will define L2 Maturity in the form of a mathematical equation, and then as a recognizable and discussable *construct* to be used in Applied Linguistics discourse.

10.1.1 L2 Maturity as a mathematical equation

L2 Maturity – as an equation – is made possible by tracking cognitive complexity scores (such as from Lectica). In this study, I used Lectical Analysis to produce each students' L2 complexity scores. *L2 Maturity* ($\boldsymbol{L2^{M}}$) can be presented as a percentage that is calculated by dividing the student's L2 complexity score ($\boldsymbol{L2^{lec}}$) by the same student's L1 complexity score ($\boldsymbol{L1^{lec}}$) if available:

$$\boldsymbol{L2^{M} = \frac{L2^{lec}}{L1^{lec}}}$$

For this doctoral study, although there were Lectical Analysis scores for each students' $\boldsymbol{L2^{lec}}$, the students did not also take the Lectical test in their L1. Therefore, there was no provision for each student's $\boldsymbol{L1^{lec}}$ score in this study. As a functional workaround, a *norm* complexity score ($\boldsymbol{L1^{norm}}$) for similar university students in their L1s can be used as a substitute denominator:

$$\boldsymbol{L2^{M} = \frac{L2^{lec}}{L1^{norm}}}$$

*An example with real numbers where $\boldsymbol{L2^{lec} = 7.5}$, and $\boldsymbol{L2^{norm} = 10.3}$

$$\rightarrow \quad \boldsymbol{L2^{M} = \frac{7.5}{10.3} = 0.72}$$

The result can be expressed as follows: $\boldsymbol{L2^{M} = .72}$, or we can say the student demonstrated ***72% Maturity of L2***. This is an easy to understand evaluation of the student's L2; it is a *maturity* percentage based on their L1 production. If the student's L1 and L2 are equivalent, the Maturity percentage would be 100%. If at some point

the student's L2 overtakes the L1 (such as is often the case when a family moves to a new country), the L2 Maturity percentage may achieve an evaluation over 100%. In this way, the L2 Maturity percentage is straightforwardly logical, easily calculable, and easily understandable. (However, this diverges from Lectica's literature, as they do not use *Maturity/Maturation* in their literature.)

L2 Maturity comparisons across groups (such as with *OL1* vs *OL2*) are possible as long as the $\boldsymbol{L1}^{norm}$ remains an equivalent denominator for both groups; without an equivalent denominator in such cases, group comparisons would produce less reliable results. Therefore, although using the $\boldsymbol{L1}^{norm}$ is *not* ideal for individual student evaluations (because the student's actual L1 score is not being used as the denominator), the method of establishing a common denominator works well for *group* comparisons when a single $\boldsymbol{L1}^{norm}$ is established as the common denominator in all of the pertaining group-based data calculations.

Similarly*, L2 Maturation Rates* can be calculated by rate of growth between two or more Lectical tests, in comparison with an established growth of an $\boldsymbol{L1}^{norm}$. Longitudinally, $\boldsymbol{L2}^{lec}$ scores should be expected to change dynamically across tests.

It is also important to note that if an individual's $\boldsymbol{L1}^{lec}$ happens to grow atypically over any given year, perhaps due to intensive studying of L1 that year (or some other natural or contextual affect), the gap between their $\boldsymbol{L1}^{lec}$ and $\boldsymbol{L2}^{lec}$ can widen at an atypical rate, dramatically lowering/raising the L2 Maturity percentage via no real fault of the L2 learning context. The same goes for the the $\boldsymbol{L2}^{lec}$ *mean* if it is a group study.

Most importantly, it can also be said that this is a strong argument for the collection and assessment of $\boldsymbol{L1}^{lec}$ even in EFL classrooms. This is because the L1 data can help the L2 teacher better understand each student's *current* linguistic growth and their potential to express deeper thoughts in either language. All in all, an EFL teacher who is ready to capitalize on strong performance in their students is in an ideal position to trigger dramatic L2 maturation in their EFL classrooms. This is a new realization that branches from the usage of the L2 Maturity construct.

10.1.2 L2 Maturity as an Applied Linguistics construct (non-mathematical approach)

From an Applied Linguistics perspective, L2 Maturity can be used as an "umbrella term" that points to a complex paradigm of evaluations that include areas of development that may not typically appear in traditional L2 evaluations. Even so, it answers a single simplistic question → "How *good* is the person's L2?"

For the layperson, this is a common question that commonly receives an instant response (often without any hard data to go by) → *"His Japanese is pretty fluent for only being in Japan for 6 months."* Or, *"Not very good for being in Japan for six years."*

Why can the layperson produce quick responses to such a complex question? What are they basing this evaluation on—and why do they come to these conclusions so quickly and confidently? These are thought-provoking questions.

In contrast, for the SLA and Bilingualism researcher, this sort of questioning represents a complex set of issues. I have hypothesized that key factors in this paradigm are displays of *socio-cultural understanding*, *non-linguistic aptitudes* (body language), and displays of *critical thinking in the L2* (Murphy, 2008). The crux of the hidden issue here (literally and figuratively) is that although the above three areas are not typically assessed with standardized tests such as entrance exams, they *do* exist as psychological constructs (that are nurtured in context concurrently with linguistic growth, so they end up belonging to the same/similar neuro-networks in the brain) in all students with an L1 and an L2—albeit, in context-influenced proportions. It is weirdly ironic that the layperson can summarize a person's L2 Maturity in a split second, yet for a researcher it may take weeks or months of data collection (and debates on the criteria) to come up with a similar answer. And more to the point, it is ill-fated that current standardized tests do not typically to take these hidden yet socio-culturally and psycholinguistically necessary-to-succeed-in-life constructs into account.

As discussed in Dynamic Skill Theory, new higher-level learning for students around university age occurs with the connection of diverse systems of knowledge—

causing a radical reconstruction of knowledge and aptitude (discussed as *cognitive reorganization*). So, when areas as diverse as socio-cultural understanding, body language, and critical thinking eventually work together as a unified web to raise a student's L2 Maturity on a holistic level of complexity, the emerging reorganization of knowledge produces neuro-networks capable of demonstrating new skills and new insights to problems that are exponentially more complex than those from networks that were established by simple rote memorization of L2 forms (as can be deducted from Fischer, 2008).

When students have been in a context that nurtures networks capable of demonstrating exponentially deeper levels of thought (over rote memorization-based networks), those same students may also display a stronger "intent to use the language" –instead of deferring to the L1 whenever possible. It can also be said that this "intent to use" the language is also a catalyst for further learning, making the entire process a virtuous cycle of learning that ensures further natural growth of the networks—which induces even further learning potential.

The cycles of learning discussed throughout this thesis have been put in place to specifically help foster and maintain these virtuous cycles of learning. For further clarification of *L2 Maturity* as an Applied Linguistics construct, below, I have listed examples of low L2 Maturity and high L2 Maturity. I have also considered how the layperson quickly surmises L2 Maturity—and therefore I have listed three key factors that can be used to accurately reveal L2 Maturity, albeit informally. This is a quick form of appraisal method because: (a) hard data is not used, and (b) because aspects such as length and speed are not factored in. That said, the establishment of the factor list has made my own understanding of L2 Maturity more tangible, and somewhat ironically, this quick approach to maturity appraisal has made my evaluations of students' aptitude more precise. Because it is a quick yet useful form of appraisal, it can be used easily by any EFL teacher while teaching class. It is beneficial for teachers to be able to make practical appraisals such as these, without having to set up elaborate equipment and/or spend extra hours with students just for testing.

Examples of low L2 Maturity in university students:

- ❑ After 6 years of English study in public schools in Japan, as an 18-year old exchange student in the USA, the student still cannot order a hamburger in English.
- ❑ When asked where the train station is, in English, the university student flees from the situation instead of providing help.

Examples of high L2 Maturity in university students:

- ❑ When asked a political view in the L2, although it may be outside of their current linguistic L2 capability (but it is within their L1 capability), the student pushes on with a robust dialogue laden with justified view-points, even though the L2 usage is not at a native level of precision.

Hypothetical 3-factor list for quick L2 Maturity appraisal:

(a) the user is capable of demonstrating complex thoughts in the L2 at a level that is reasonably close to their L1's level

(b) the user displays a willingness to use the L2 in context

(c) the user displays a willingness to improve their L2 competence

Although the 3-factor list (above) has can be useful when quick appraisal is called for, there is no substitute for accurately calculated and appropriately deduced student growth documentation and evaluation. The next section will put the focus back on more scientifically accurate evaluation procedures, and the insights gained from work in that area.

10.1.3 Why is this important (the new insights)

Returning to the mathematically-based evaluation of L2 growth as discussed in 10.1.1, what does Lectical Assessment have to offer that can provide *new* insight into language development? Lectical Assessment data (as used in 10.1.1) allows researchers to put a finger on L2 Maturity and gauge it scientifically. At the same time, it helps provide a quick but most accurate answer to the question, "How good is George's Japanese?" → *"He has reached 70% maturity, compared to English."* This is a straightforward, accurate, and practical reply.

Via analysis of the Lectical data, I was able to demonstrate that it is possible to gauge the growth of cognitive complexity as demonstrable in the L2—and compare that growth speed with the norms found in L1 contexts. This is a novel approach as it does not look at right or wrong answers from the students. It needs not look at grammatical rules, nor vocabulary lists to come to a conclusion. And yet, it is an exceptional tool for the evaluation of the, and the understanding of, the complexity of language development; and it is an exceptional tool for gauging the rate of language development because: (a) not only does it provide objective scientific data, but because (b) the results are easy understand and the Lectical Assessment tests easy to implement for any teacher--making is eminently practical on a several levels.

10.1.4 L2 Maturity in detail

From a historical perspective, the fact that L2 maturation has become a prominent area in this thesis is not coincidental. *OL1* was the centerpiece of my MA dissertation (2009). I consciously adopted an integrated a tri-elemental model of linguistics, emotion, and socio-cultural aspects into the pedagogical design (Figure 101). I gave it the pedagogical slogan, "Teach for the DATC", as a way to propagate what I felt to be an important model for teachers to learn and to teach by.

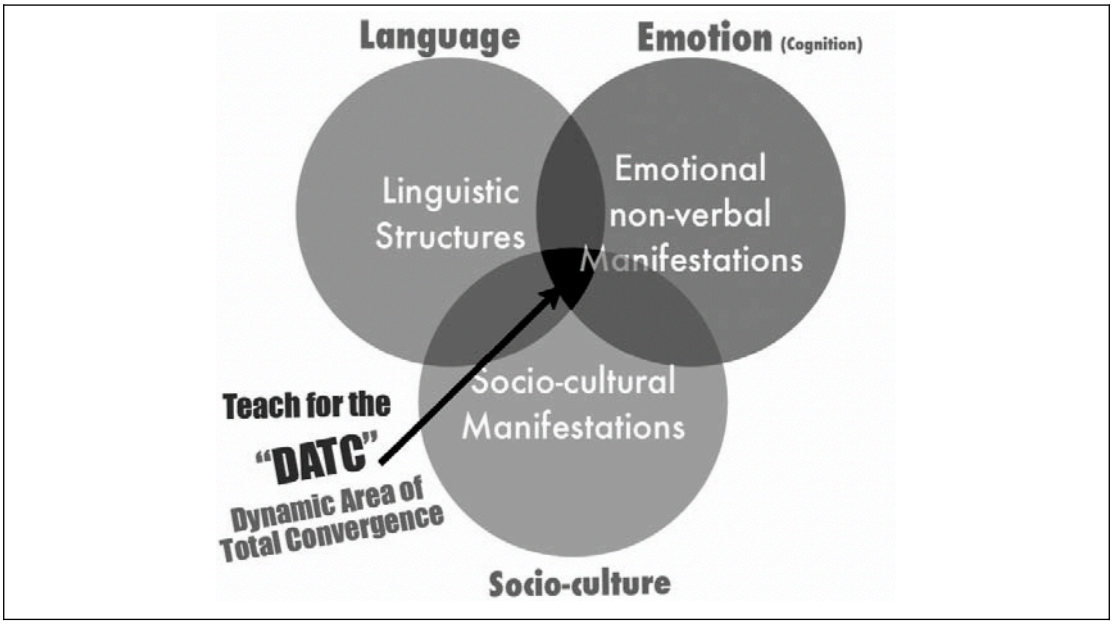


Figure 105: Tri-elemental model from the MA thesis (Murphy, 2008).

Looking back, I believe that the underlying intention to put a focus on these three areas is sound—and at the core, it forms a large part of what emerged as L2 Maturity assessment. Because of this, the integration of these three areas is still being explicitly embedded within the tasks of *OL3*. I shall return to a discussion of the *Teaching for the DATC* construct later in this chapter.

Refocusing on maturity, we can say that when maturity/maturation is discussed in SLA, it is typically either regarding a peak of acquisition, or acquisition in the context of elderly people. Here I use *L2 Maturation* (and *L2 Maturity*) to primarily denote the demonstrable cognitive growth levels within the learner's L2 in comparison to his/her demonstrable L1 levels. Moreover, because the norms of students' growth rates per year in the L1 have already been established by Lectica, results from EFL students using the same test can be used in pre- and post-tests to assess *L2 Maturation Rates* (in comparison with established L1 growth rates). Lectical assessment has allowed me to scientifically measure this gap in complexity of output (gap between L1 and L2), and the speed at which this occurs. Where there is a difference in speed, it is natural to ponder the causes of the differences. When two coursebooks are pitted against each other, the difference in L2 maturation rates between the two coursebook groups can be used to deduce which coursebook

provided a more efficient learning context for the students. This is exactly what was done for this thesis—I compared *OL1* and *OL2* and the Lectical data allowed me to deduce which book had the more efficient pedagogical design.

Is the concept of *L2 maturity* (and the gauging of *L2 maturation rates*) important outside the territory of coursebook evaluation? I would to argue that understanding L2 maturity (and the stages of cognitive development) is just as important for the fields of SLA and Bilingualism as it is for coursebook evaluation. If teachers and SLA/Bilingualism researchers realize the potential of gauging L2 complexity growth rates against L1 complexity growth rates, how could it not affect their approach to teaching and their understanding of language acquisition? Below I have listed changes that can occur if students' L1 performances are used as benchmarks against each students' L2 output and longitudinal growth in the L2:

- **Level-appropriate teaching.** Foremost, L2 teachers will have to stop the teaching of complex language that is already beyond students' capabilities in their L1s. Put in this way, it sounds silly that teachers might be doing this—but if there is no L1 benchmark to gauge against, then the L2 teacher would never know "how hard is too hard" per student, so this can easily happen—and go on, unnoticed. If this change is called for, it has the potential of rectifying this problem of having incorrect level choices in the L2 classroom. It calls direct attention to the current L1 level of each student as well. Knowing the current L1 level will certainly be helpful for teachers in other subjects. In sum, to ensure level-appropriate teaching in either language, the demonstrable complexity of both the L1 and the L2 should be gauged for growth each semester. This change will "work magic" because it will strong-arm teachers into choosing level-appropriate tasks per student, and it will compel teachers to group students into level-appropriate clusters. This will also remove teacher guesswork regarding student levels, making level-appropriate teaching more easily attainable. Level-appropriate teaching will naturally become normalized.

- **Functional L2 interaction over rote memorization.** Lectical Analysis looks at the *complexity* of each student response to challenging questions that require deep thought. This is innovative, and I do believe this is how primary assessment of L2 should be handled. If Lectical Analysis growth marks are provided to the teacher (and/or the teacher learns how to do this assessment), the teachers will naturally begin to focus on the *growth* of functional L2 systems within the students—in lieu of a focus on rote memorization and test scores of grammar points. Teachers will begin to notice, if they had not already noticed, that rote memorization alone typically produces dysfunctional interaction. I am not advocating the eradication of rote memorization tasks altogether. However, I believe they should be incorporated into the classroom as supplementary tasks; they work better in that capacity. If teachers realize that rote memorization alone typically produces dysfunctional interaction, the development of a *functional* L2 system within each student would naturally become the focal point of the L2 teaching. So, this change has the potential of rectifying major problems that are deeply rooted ELT issues in Japan (and in other countries)—where "correct" usage is weighed more heavily than fluent interaction.

- **More motivational classrooms, with a consistent string of *solvable mysteries*.** If the Lectical scores of both the L1 and the L2 are known to the teacher, they will more easily be able to challenge each student with a level-appropriate string of *solvable mysteries* that are just beyond the current potential of each student (but are achievable with effort from the student). As discussed earlier in this thesis, solvable mysteries are tasks that are just right for each student—not too easy, but also not too difficult for them. Embedding a string of solvable mysteries across the semester can help create an ideal learning context for students—because of the

naturally high levels of engagement and motivation it brings to the classroom. Similar to how addictive video games are to children, there is a naturally occurring adrenaline rush that comes from solving classroom challenges that are "just right"—helping to keep the student learning efficient. Moreover, the students will likely begin to appreciate their progress and notice their fuller engagement with their studies. This phenomenon will be discussed in the next section.

- **Looking beyond the language.** For rapid L2 Maturation purposes, it seems likely that systematic teaching that focuses on the development of students' needy areas for their development as real people who can live life in the L2. Teachers who take the concept of L2 Maturation seriously should design learning contexts that foster this attitude toward L2 learning and teaching. One of the goals of the teacher must then be to affect the attitude of the student as well – ultimately, it is the student that must see L2 as something more than just a school subject. Ideally, EFL students must gain the feeling that English is not only a school subject, but that represents a culture and a different way of life. For this attitude change to happen in a student, a combination of processes must culminate: (a) The student needs to feel autonomy via bottom-up learning tasks, and realize growth from this autonomous context. (b) They need to feel personal ownership of the language and the learning context. And, (c) to optimize any student's rate of maturation, it is imperative that the student also develops level-appropriate metacognitive skills. These metacognitive processes should be inclusive of self-knowledge of their strengths, self-knowledge of their weaknesses, and also the development of a clear vision of attainable goals in their L2 Maturation journey. As such, all three of these points (a, b, and c) are interwoven matters.

10.2 Metacognitive development

The hard data that was used for Lectical Analysis and the insights from that form a natural segue to this next section regarding the growth of metacognition. Above was a discussion about the growth of the L2 within students—this section is about students' perceptions of their growth. This is not something that was measured by Lectical Analysis, but it was a common theme for the Focus Groups, and a common theme among the written responses from the students. The most striking data from Chapter 7 in this regard was the huge difference between the results of *OL1* and *OL2*. The *OL1* group fixated more on the more tangible and concrete aspects of language learning when asked about their coursebook, while the *OL2* group consistently discussed the *hows* and *whys* of their L2 growth. If the coursebooks are the primary reason for this discrepancy between the demonstrated metacognitive skills of these two groups, then is seems that only *OL2* enticed the students to be more self-aware of their L2 growth, with more complex abstractions.

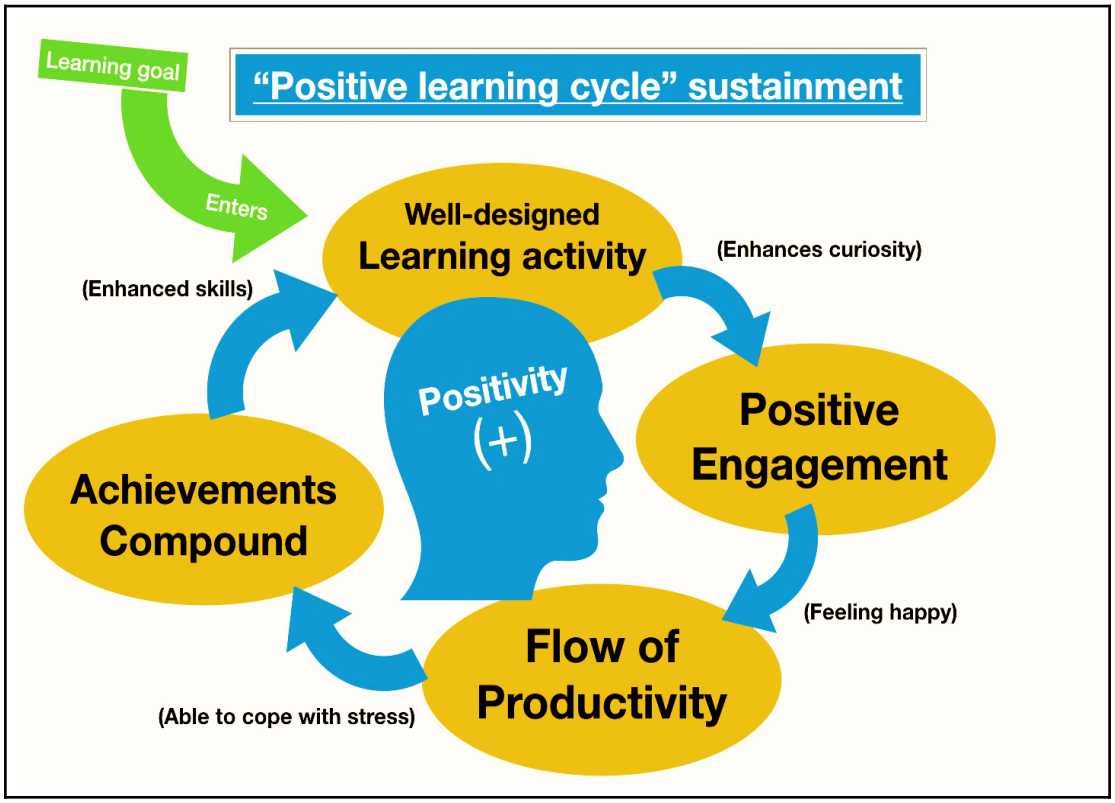


Figure 106: Illustration of students' understanding of their learning cycles

Learning from responses of the *OL2* students regarding their metacognition, I designed the "Positive learning cycle" sustainment diagram (Figure 102). In the diagram, the bracketed words describe how the students said that they felt during each particular phase of the cycle. The large orange circles are what the students perceived as "milestones" in the flow of events.

The milestones that the students perceive are appreciably different from my two protocols for the *OL* series; they do not closely follow the *OL* flowcharts in Chapter 9 either. This version of events from the students' perspective illustrate how different the students' perception of their learning was from my perception of events as the course designer. Their presentations and their designing of tests are smaller issues within the grander scale of their perception. Their metacognitive discussions revolved around how engaged they were with the course--rather having a clear focus on the individual tasks. This is an eye-opener. But, in retrospect, I can see why their deeper thoughts were regarding personal engagement—it is only natural for them to be ego-centric when asked to think deeply about the course.

EFL learners in Japan seem to seldom be in a context that nurtures deep thinking *in* the English language. English is hardly ever heard or used outside of the English classroom, so where and when would they be in such a context? Teachers in Japan should not assume that this happens regularly. And, considering that student thinking in the English classroom is primarily done in Japanese, and then just translated over to English when necessary, hardly any class time is spent on the development of *thinking-in-English* skills.

Is this a problem? If the student will only ever have to translate over like this until they pass an exam, and then they can literally forget about English after that, then there is nothing "wrong" with this. It is the most pragmatic solution for the student. However, if communicative competence and near-native fluency is a part of the goal, the sooner the student can learn to think in English and have metacognition in English, the better. In Japan, this is a tall order. Luckily, my studies have found a substantial variance in metacognition between the two coursebook groups that I examined. Therefore, with this data, it is possible to deduce methods that will

promote metacognition in the English classroom—even in the context of Japanese EFL. The next further develops these notions and discusses implementation ideas.

The relationship between metacognition and rapid L2 maturation

My studies showed that there was significantly stronger metacognitive development in the *OL2* classes (compared to the *OL1* classes). I taught both classes without discrimination, and I so am confident in claiming that the design of coursebooks *can* affect metacognition—some promote it, while some may even hamper it. Even with the admission of the "halo effect" and the magnitude of the affects of student-teacher rapport, the flow of events that are designed into a coursebook seem to have a substantial impact on the development of metacognition.

By this point in the thesis it is quite clear that this phenomenon is related to why the *OL2* group showed more rapid growth in L2 complexity than the *OL1* group did. There is an intriguing relationship between quicker growth in demonstrable L2 complexity and stronger metacognition in the students. This link is the key in helping to deduce what can be done to design better English courses in the EFL context.

When asked to discuss the course they took in open-ended questions, how did the output differ between *OL1* students and *OL2* students? The *OL2* students discussed the *hows* and *whys* their English learning improved, while *OL1* students focused more on *what they did* in the classroom. According to cognitive development researchers such as Fischer and Piaget, these *how* and *why* type responses represent profoundly more complex thinking (requiring more complex networks) than *what I did* responses. So, when the students were given open-ended questions about the course and the *OL1* students were more focused on the *what I did*, while *OL2* students were more interested in discussion of *why* I learned this way, a serious difference in maturity was discernable; frankly, the *what I did* level of response is utterly childish in comparison. This concretely supports the earlier discussed hypotheses regarding how *OL2* was able to achieve *cognitive reconfiguration* within the 15-week semester, while *OL1* did not.

Vocabulary test responses and grammar drill outcomes cannot easily

determine *L2 Maturity* (as it is defined in this thesis). Moreover, from what I have gathered thus far, a strong reliance only on rote-based language teaching and teaching with grammar translation methods seems to impede L2 Maturation (as it relies heavily on the cognitive prowess of the L1 to enable a framework memorization understanding of the covered material).

What sort of teaching, then, might assist in the development of efficient L2 Maturation? It may be regarded as a waste of time to some teachers, but I do think that the study of *other* subjects via English has excellent potential for assisting in efficient L2 Maturation. This is likely because the process "expects" the students to use the English to achieve the goal—English usage is never the goal in such classes; it is only the means to achieve the goal. But, the ensuing flow of events advances the students' command of English *as* the primary goals are being achieved. This is identical to the way children learn their native language—it happens as a byproduct of tasks meant for achieving other goals.

The move from simplistic grammar translation and "rote-teaching of English" to "learning English via the study of xxx" has become increasingly popular across the past two decades. Approaches such as CLIL (Content and Language Integrated Learning) and CBLT (Content-based Language Teaching) now have their own place in the Applied Linguistics literature and in conferences around the world. So, the idea of learning English via the study of something else is no longer a novel approach

The big question is, will CLIL, CBLT, and the like *really* promote efficient L2 Maturation? And if so, how efficient will they be? The results hinge on their implementation. There are two ways of going about teaching with these methodologies: (A) the pre-emptive teaching of the assumed-to-be-necessary language (*before* it becomes necessary for the students), or (B) the provision of the problem first (making the students think first, and become anxious), *then* providing linguistic help in achieving the solutions, as necessary. This dichotomy represents the top-down vs. bottom-up issue that I will be discussing in detail in the next section (10.3), but here I would like to only focus on the metacognitive aspects regarding

this dichotomy.

From the work that I did in this thesis, I hypothesize that the *Type A Process* (providing the answers before the problem) lacks much of the cognitive processing that is necessary to promote metacognitive thinking in the students regarding their learning process. This means that, after the learning, the students can probably tell you what they did, but they wouldn't *think to tell you* "why" they came to those conclusions or why they used the language in the way that they did. The students were provided with hints and answers beforehand, so why should their minds develop metacognition in those areas? Moreover, the dopamine rush when solving problems will likely be comparatively small because the answer is already there – the challenge is not very interesting for them.

In contrast, I hypothesize that the *Type B Process* (problem comes first, appropriate help comes after anxiety starts) primes students for satisfying "aha moments" with a proper dose of dopamine. As a byproduct, this also creates an inclination to discuss how and why they came to that solution—students often want to tell you how they got there, to receive that second dopamine rush, much like when people discuss and re-live sports the day after. There is an remarkable inherent biological function about the *Type B Process* that not only makes students more aware of the conclusion that they come up with, but also makes them crave "closure" by explaining the process to somebody (and being biologically rewarded again).

My revelation here is that not only should teachers aim for more *Type B Processes* in the EFL classroom, but teachers should also realize the importance of giving students the time and place to discuss their metacognitive discoveries (share their *"triumphs!"*). Not only would it be a lost opportunity for rapport building, but it would be a bad—almost cruel—pedagogical design in a lesson if the teacher did not provide ample space for this *metacognitive growth, venting*, and *closure of the task*. So, to answer the question regarding the efficacy of CBLT/CLIL and the like, it would seem that a focus on *Type B Processing*, with the addition of ample *metacognitive* tasks (as discussed above) will most likely lead to efficient L2 Maturation.

10.3 Good teaching: Importance of top-down vs bottom-up ratios in the classroom

This section will discuss an issue that spans much further than EFL—it touches the heart of educational design. In my study, the *OL2* group was the group that displayed stronger and more elaborate metacognition regarding their learning and their understanding of how *OL2* worked pedagogically for them. *OL2* students even spontaneously began to design flow charts to further demonstrate their growing understanding of the efficacy of *OL2*. Several students also discussed (directly and indirectly), the benefits and differences between top-down and bottom-up teaching styles.

I use the terms top-down and bottom-up differently from some of the other authors in *Education* literature. At times, this causes confusion at conferences and workshops. However, throughout this thesis (and in all my prior publications and presentations), I have used top-down teaching to mean authoritative, lecture-styled, "conveying" of information to students. In top-down teaching contexts, the students are passive learners. On the other hand, I have used bottom-up to mean student-centered, passion-following, interactive information "discovery" by the student. In bottom-up teaching contexts, the students are active learners, while the teacher take the role of a facilitator.

In the literature review I provided solid critiques regarding the six syllabus types that I discussed. Among them was TBLT (Task-based Learning and Teaching). I was not critical about having tasks in the classroom – I strongly agreed with that part, but I pointed out that for teachers who are not confident in assessing free discussion among students, they may not want to choose this approach for their own classroom—and this is very logical. Why would you choose an approach that potentially highlights your own weakest area? For teachers who are not confident with on-the-spot judgment and learning facilitation, the ultimate approach might be the "listen to what I have to say – no questions" approach. This is the extreme version of the top-down method (a stern lecture with no Q&A), and in this extreme form it may be questionable as a form of teaching—the students may be able to learn

more by sitting at home, using a computer, or just reading a book. So, with this very real issue firmly in place, the question is—what is the best balance between top-down and bottom-up teaching?

If given a casual thumbing, the *OL1* and *OL2* coursebooks might be called coursebooks of "bottom-up" design; they do look more like a "workbook" than a "textbook," literally and figuratively. There are no dense reading passages anywhere in the books. It is therefore up to the students to follow their passions and discover what they need to know to accomplish the provided tasks. However, even though both books look mostly empty and generally follow the same principles, *OL2* produced significantly better results. Relatively small changes within the top-down facilitation in the *OL2* design is what must have caused this significant difference. This sort of difference is akin to the way interest compounds in the bank—a small difference in the interest today compounds into a substantial difference accrued down the road. From this notion and the results of my main study, we can deduce that not only is the ratio between top-down and bottom-up teaching a substantial choice to be made by the teacher, but also that even a small change within top-down guidance can make statistically significant differences by the end of the semester.

Because both of my tenet lists (Chapter 9) were borne from the literature review and studies regarding the outcomes of student data, the lists would likely be a good guide for teachers in the planning stages of their courses. However, the tenets cannot offer guidance on the exact ratio between top-down and bottom-up teaching. My current general rule of thumb here is: more top-down if *accuracy* is the goal, and more bottom-up if *fluency* is the goal, but these ratios must not stay rigid. They need to change dynamically to meet the-then current needs of the students.

10.4 Good teaching: Revisiting COOPs and "Teaching for the DATC"

The unconventional aspects of *OL2* and *OL3's* coursebook design are part and parcel of the approach. It provides opportunities and purpose-designed learning contexts for the EFL students to automatize English usage so that it becomes more natural for the students to actually use English—without having them "studying" and rote-memorizing phrases the way they are accustomed to. Each Module in *OL2* and *OL3* a cyclic set of well-thought out compensation strategies that make up for the fact that the students are in an uphill battle within a mostly unaccommodating EFL context. This issue is permanently linked to the top-down/bottom up teaching ratio matter discussed above.

Although I used the protocol to design *OL2* and *OL3*, the *Cyclical Optimized Ownership Protocol* (COOPs) does not outline specific task designs within *OL2/ OL3*; that was accomplished in the charts in Chapter 9. What COOPs does is lay out founding principles of effective EFL coursebook design. For easy reference, I have reposted the protocol here (Figure 107). All of the points that were discussed thus far in this chapter (10.1 to 10.3) culminate in the central sphere in the COOPs model. That sphere represents the amount of displayed student ownership of the L2. The border of the hard boundaries around the sphere represent the boundaries of that student's ownership of the L2. However, there is a larger sphere in the model as well.

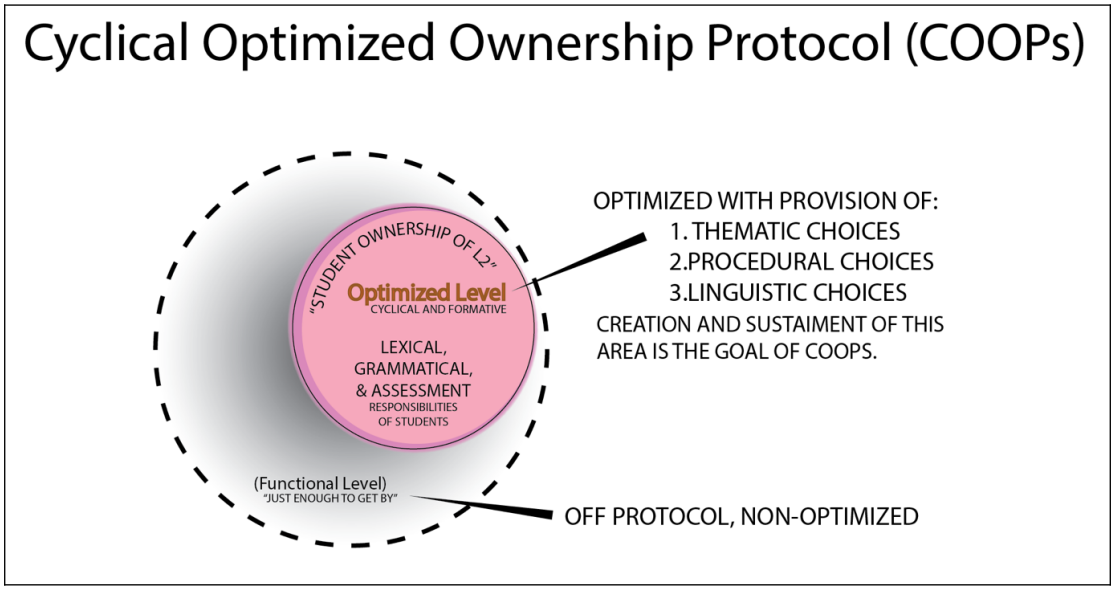


Figure 107: Revisiting the COOPs protocol.

It is imperative for students to actually feel ownership of their L2 learning – without that attitude toward L2 learning, teaching will not be effective nor *affective* for them. As can been seen in the model, there are two levels of ownership possible: the functional level and the optimized level. The larger sphere is the un-optimized, just-enough-to-get-by level. This level is tenuous and unreliable; therefore it is modeled with a dotted line instead of a solid boundary.

My data from this thesis suggests that with *OL2*/*OL3* as a tool, it is possible to create an optimized level of L2 learning within students (that smaller sphere in the COOPs model). There are three primary factors that help form and sustain this smaller, optimized L2 sphere.

The three factors are *student responsibilities.* For sustainability of the inner sphere, the students must carry carry out: lexical responsibilities, grammatical responsibilities, and assessment responsibilities. What support is necessary from the outside (from the teaching context) to help sustain this optimized L2 ownership? Well-designed autonomy-enhancing support is necessary to maintain this optimized level of ownership within students. Students will be able to sustain a high level of L2 ownership if they are allowed to follow their passions within a framework that binds them, but gives them freedom within that framework to make choices regarding the

whats and *hows* of the learning, plus the freedom to "play" with lexical items, to be able to find their own voice, instead of simplistically mimicking what is written in the text.

The first word in the COOPs acronym is Cyclical. *OL2* and *OL3* are deliberately built on systematic *cycles of development*. The idea of cyclical teaching/ learning was discussed in Chapter 2 as a design for teaching/learning that is especially useful as an optimized compensation strategy for the exposure-deprived EFL context. It may not always be easy to implement, but the idea of cyclical teaching is at the core of COOPs. Discussion with the focus groups revealed that the students not only liked the implementation of the cycles, but that they noticed their *motivation*, *metacognition*, and *appreciation of the coursebook design* improved per round. Those three revelations were always at the top of the list of their discussion topics each month.

As discussed in the above sections, there is an interesting and observable growth-related phenomenon with the *OL2* classes → it has to do with the correlation between the frequently recurring appearances of themes (*metacognition*, *motivation*, *appreciation of the coursebook design*) and their realization of L2 Maturation. Although correlation does not necessarily equate to causation, the correlation among these themes and outcomes does not appear to be accidental; I hypothesize that this convoluted flow of events is specifically what is creating an optimized context for L2 Maturation within that uphill battle within the EFL context.

In sum, the EFL classroom must not only rote teach the language, but it must also aspire to be a practical compensation strategy – it must be a *stand-in* for all the external relationships that would have occurred to the students had they been in an ESL context instead (such as Japanese exchange students on a multi-year exchange program with a school in the USA).

"Teaching for the DATC" was conceived of during the early design stages of *OL1* (Murphy, 2009) and it continues to be fundamental in the design of *OL2/OL3*. This is because after all this data has been compiled, it is still evident that EFL teaching should incorporate emotional constructs and socio-cultural constructs along

with linguistic constructs.

For this convergence to be the most beneficial for the student (for it to positively affect their L2 Maturation), the total convergence area of the three constructs should be large—as large as possible. How can this be accomplished?

The three constructs must be nurtured in a way that allows them to grow together organically and dynamically—the required type of dynamic and organic growth cannot occur within the students by simplistic memorization of facts. This argument is similar to the point made in 10.1, "Look beyond the language"; L2 Maturation occurs most efficiently when extra-linguistic factors become organically tied to, and help compound into, L2 Maturation.

Rather than leave the convergence up to chance, the *Teach for the DATC* slogan calls for explicit teaching designs that zero-in on the dynamic meshing of the three constructs. Considering that *OL2/OL3* has been designed with the *Teach for the DATC* protocol in mind, and considering the success of *OL2*, it is plausible that *Teaching for the DATC* is not only a slogan, but a valid protocol for a wider range of EFL contexts.

10.5 Good teaching: Connection to the neuroELT Maxims

I found it reassuring that, although the tenet lists were directly formulated from the emerging data of this current study, the lists showed uncanny similarities with previous work that I have been involved with. I lead a team that compiled a list of 50 *neuroELT Maxims*. The maxims were compiled based on neuroeducational principles spanning several years of research with colleagues and interns. Seven of the tenets turned out to express virtually identical messages (Table 55). This strong compatibility across these lists that were developed separately strengthens the position of both lists.

Table 55: The seven neuroELT maxims that are fully compatible with the tenets from this thesis

NeuroELT maxim	Explanation
"Choices" fuel learner motivation	Autonomous choice-making commonly activates neural networks that provide feelings of satisfaction. Thus, acting with autonomy is typically a naturally pleasurable experience.
"Real-time feedback" is at the core of cognitive development	The human love of the prediction-outcome dichotomy stems from the rewarding rush of dopamine received when we *perceive* our solutions to be correct. Thus, real-time feedback satisfies and even nurtures our craving for the dopamine rushes.
"Aha moments" enhance neural networks	Pleasurable dopamine rushes also result from successful problem-solving. The neuronal networks involved become more *efficient* through myelination; the most often used networks therefore become enhanced for superior processing (speed and quality actually become enhanced).
"Happy students" learn better	Studies show that happier individuals perform better academically, with enduring motivation. *Positive Psychology* has found that engagement in *perceived* meaningful activities may sustain happiness more effectively than the simplistic pursuit of pleasure.
"Performances of Understanding" are essential for good assessment	PoU activities provide learners with the opportunity to make their learning and understanding visible by having them demonstrate their *Understanding* (real word mastery) of the concepts. PoU-based pedagogies are aimed at synthesizing student learning and advancing it all to the next level in the form of a real-world applicable performance. *This is often lacking in traditional top-down, rote-memorization only teaching.*
"Assess" in three ways	There are 3 synergistic ways to assess performance (referred to as the *Trinity of Assessment*): 1. Self-assessment 2. Peer to peer assessment 3. Teacher-student assessment It is the culmination of, and synergy of, *all three* that is ideal for nurturing students' metacognitive prowess, problem-solving skills, and overall academic growth.
"Encourage" mistakes; celebrate mistakes	People can and do learn from mistakes. However, in contexts where people are *afraid* of making mistakes—they are in an extreme disadvantage. The negative context may be detrimental to natural learning processes. An environment where mistakes are treated as special teaching/learning moments that are *essential to learning* may lead to more neuro-efficient cognitive development.

10.6 Limitations

There are several limitations to this study. Firstly, my research design did not follow what might be called a standard experimental design—that is, a design that establishes and compares a control group and an experimental group. The primary reason for this was because my PhD project necessitated dynamic changes within the *what* and *how* aspects of this action research; my project called for continuous development of a coursebook series rather than taking a finished product and comparing it to a more established counterpart. Secondly, accessibility to students and time factors caused numerous constraints that needed to be worked around.

To overcome these issues, I: (a) organized a replication study of my primary study (the year after my primary study, with different first-year students); (b) used questionnaire data from other classes being taught by other teachers, and compared them with my own results to ensure validity, and to see how well *OL2* did against other textbooks, in the eyes of the students; (c) the *OL1* group was treated as a real control group, enabling scientific comparison with the *OL2* group's results—this allowed for meaningful comparisons that provided intriguing data.

I believe that the cumulative evidence from the comparisons helped confirm that the project was at least moving in the right direction. That said, nothing can take place of a properly laid out experimental study; this is something that I must find a way to carry out in the future.

Another limitation of the study was that across all six years of data collection, for any given year, *OL1* was constantly represented by only one course per year (for the primary study, n=30). The study would have been more robust had I had more *OL1* coursed to work with per year. This was a significant limitation that was wholly dependent on the university teaching schedule; the course ratio was set by the university at 6:1 (*OL2*:*OL1*) across all five years of study. A different ratio may have produced different results.

The next limitation of this study is regarding Lectical Analysis. For financial reasons, the initial plan had always been to study the growth of students across just one semester with Lectical Analysis. With hindsight, since the L2 Maturity growth

was vastly different between *OL1* and *OL2* (roughly 8x), I now would have preferred to have followed the growth of those students longitudinally into their second semester—perhaps even though the entire second year, to complete a 2-year study (four 15-week semesters). I am curious to see if the L2 Maturation effect of *OL2* was sustainable longitudinally into the second year, and beyond. Therefore, in spite of the remarkable data that Lectical Analysis supplied for that first semester, the single semester data of growth studies via Lectical Assessment is an unfortunate and significant limitation of this study. Since my curiosity for this is strong, studies based on Lectical Assessment of *OL3* classes, longitudinally, are on the list for future research projects.

Another limitation regards teacher-student rapport. A good portion of the final year of research was dedicated to uncovering how a "halo effect" might have affected student motivation and course evaluation – and how *that* may have affected the overall results of my study. Two potential limitations emerge from here: (1) the study of the "halo effect" occurred in the final year (2018-2019). Although it has been established several times in this thesis that any given year across all the years of study provided relatively similar student responses (making the data transferable), ideally, the "halo effect" study should have been designed into the primary study in 2014 and/or the secondary study in 2015. More importantly, (2) the majority of the studied Japanese EFL students unabashedly confirmed favoritism. The t-tests then showed that the choice of *OL1* or *OL2* had no relevance on this aspect—the majority of both groups reported that their *favoritism* had a significant affect on their own motivation to do better, and how they evaluated the course. I must therefore concede that my rapport with my own students affected the results, albeit rather steadily across the years.

Although my yearly results may have been relatively consistent, it can easily be hypothesized that a less experienced teacher, or a teacher who is less/more popular with his/her students might produce different results. It can also be hypothesized that a radically different EFL context may have provided different results. Of course, this argument can be made about *any* coursebook, which is why

new coursebooks are typically piloted by a wide range of teachers—in a variety of contexts before they get published. Such an epic study was not possible at the time of this project, but a more varied study may be possible in the future with *OL3*.

The final limitation stems from the longitudinal questionnaire studies. Because the questionnaire-based statistics required participant consistency (each partner had to have participated in every questionnaire session held), I lost roughly 25% of my potential data. This is because roughly 25% of the various questionnaire participants missed one or more of their questionnaire sessions and/or they input their data incorrectly in one or more of their sessions, invalidating their responses for the purpose of this study. Although student attrition is not uncommon for longitudinal studies, I would have preferred a lower attrition rate. The results may have been significantly different had there been less student attrition.

10.7 Future research directions

OL2 and *OL3* are both products of this thesis, yet *OL3* has yet to be studied in the classroom. Therefore, it is only natural to have the study of *OL3* on the top of my list for future research. As discussed in the section above, one of the limitations of my primary study is the length of the Lectical Analysis portion of the study. So, the next research project shall replicate my primary study, but it shall follow the students L2 maturation rates across at least two semesters of *OL3* usage. Currently, it is not possible to receive the students' L1 Lectical scores because the Lectical coders do not code Japanese responses, but in the future, it may be possible to:

- implement human or computer translation [Japanese-English]
- train new Lectical coders who are fluent in Japanese, to code the L1 results
- train bilingual Lectical coders, to code both the L1 and the L2 results

I would also be interested in researching the efficacy of the tenet lists with EFL courses that do not use the *OL* series. The connections that can be made with the neuroELT maxims should also be explored in a way that fortifies the content and

helps to further validate both bodies of work.

10.7.1 DMCs: Directed motivation currents

With this new-found understanding of the students' flow of engagement and motivation, I could see definite periods where the students fell into coursebook-induced motivational currents which were reminiscent of *directed motivational currents* as discussed in the SLA literature (Dörnyei et al., 2014). I was tempted in several placed to include a more prominent motivational element in my thesis research—as I am also strongly committed to DMC research (Murphy et al., 2017) —but in the end, my supervisor and I decided that it would divide the attention away from the core material design discussion. However, as part of the possible future research, here I will briefly summarize how I see DMCs are related to my thesis.

DMC theory explains the *whys* and *hows* of the occurrences of prolonged directed motivation toward tasks. It also touches on how they can be artificially created – this where DMC theory connects to my thesis. A DMC requires a project framework, typically with these elements:

- there is a clear goal
- there is a viable pathway to that goal
- the person desires to take that path
- progress is visible with progress checks and feedback on a regular basis

These four points fully overlap with the tenets used to design *OL3.* Clear goals have been set. There is a viable pathway to the goal that is semi-controlled by the students. And, progress is visible – as was made clear with my study of the *OL2* students. As I was observing student reactions, I noticed a new element that seemed to be generating DMCs which has not been mentioned before in the DMC literature. This is the recurring (cyclical) tasks and routines aspect, which is an integral part of the design of *OL2/OL3.*

However, as was made clear with the comparison of *OL1* and *OL2*, the mere

presence of these elements cannot guarantee the successful launch of a DMC. But a good way to help launch and sustain a DMC is to facilitate those recurring tasks/ routines. This is an exciting new avenue for DMC research.

Because of the repetitive nature of the cycles, the *OL* students gained increasing familiarity with the pedagogical process, but unlike many *top-down* coursebooks, students' familiarity with the pedagogy and the relentless repetitiveness of the course did not result in boredom. In contrast, the familiarity seems to have liberated them –engaging them in a way that enticed the students to contribute their own ideas *bottom-up*. In other words, the more solid their familiarity was with the process of the pedagogical framework, the more free they felt to follow their passions and capitalize on the provided opportunities. In short, although it may sound ironic, the boundaries of the cyclical framework became a catalyst for the students to explore and express their ideas more freely than usual. This led to intensive project work. All-in-all, this flow of events approximated the criteria for DMCs. I believe there is a strong future in examining the motivational implications of the *OL* coursebook series, from the DMC perspective.

There is a final observation that fully connects DMC research with one of the more prominent tenets (*"Do not allow critical breaks to happen for any student"*). The concept of critical breaks has already been explored in light of DMC theory in Murphy et al, 2017. It speaks directly of the potential of premature closure of DMCs. Naturally, teachers would not want any critical break to occur mid-chapter—nor at any point during the whole semester. And so, as the provision of a consistent flow of challenging-yet-achievable *solvable mysteries* (a chain of proximal sub-goals) may be the key to the sustainment DMCs, this tenet is the key to sustaining motivation in the *OL2/OL3* series.

In this way, the connections with DMC theory and the findings in my thesis are marked. Now that DMC research is becoming more prominent in Applied Linguistics literature, and as my own thesis comes to a close, a stronger pursuit of research that compliments both bodies of study seems to be a natural path for future research.

10.8 Final words

From my initial meeting with my supervisor, it became evident that there was little in the-then current literature to assist in what I was setting out to do. On one hand, this was to be expected because I was setting out to do a form of action research and assess my own coursebook and my own pedagogical methodology as I was developing it—so there couldn't have been a preexisting template In this way, this has been a liberating experience. On the other hand, it has also been a daunting experience because I had to create my own research protocols and procedures while in the process of improving the product. It is (somewhat comically) fitting that the cyclical designs in my coursebooks and the on-going *growth-through-upgrading* approach played out its own part in the design of *this* thesis; my thesis ended up imitating the gist of my coursebooks' pedagogical experience.

What have I accomplished on this project? I attempted a systematic data-driven revision of two coursebook designs, and I believe that the upcoming *OL3* will be a better book than *OL1* on all counts.

This project has been a quite an adventure. Including the "getting my feet wet phase" during my MA thesis, this entire EFL coursebook project has taken me on a path of maturation that spans over a decade of my life. I am grateful for the opportunity to pursue my own academic interests in this way. I feel I have made the right choice in pursuing a topic that happens to be as practical as it is theoretical.

It was a joy to realize that I conducted qualitative research that was such a cathartic experience for the focus group participants that they requested more meetings! On the quantitative side, I was ecstatic about the rate of L2 maturity displayed by the *OL2* group—8x growth speed was far beyond what I had imagined. This has given me the confidence to move forward with the publication and implementation of *OL3* in the coming years.

My outlook has dramatically changed along with the depth of the various hypotheses that have grown into the tenet lists. Both lists should prove to be invaluable across the next several years of my career. Moreover, as I am only

pursuing this topic to help teachers become better teachers, it has been my deepest joy and honor to have been invited to discuss and teach the learning from this project to peers around the globe, while I was conducting the research. This inherent virtuous cycles of teaching while learning, and learning while teaching had a strong impact on the efficacy of my own studies. I consider myself privileged and I hope that my continued research into better pedagogical methodologies has a real chance in making a difference in the world.

References

Abd-El-Khalick, F. (2005). Developing deeper understanding of nature of science: the impact of a philosophy of science course on preservice science teachers' views and instructional planning. *International Journal of Science Education*, 27 (1): pp. 15–42

Allright R. L. (1984). 'Why don't learners learn what learners teach? The interaction hypothesis.' I.R.A.L.

Apple, M. and Jungck S. (1990). "You don't have to be a teacher to teach this unit". Teaching technology and gender in the classroom'. *American Educational Research Journal*, 27/2: 227- 51.

Ausubel, D. P. (1977). The facilitation of meaningful verbal learning in the classroom, *Educational Psychologist*, 12:2, 162-178.

Bailey, K., Madden, C. and Krashen, S. (1974). 'Is there a "natural sequence" in adult second language learning?' *Language Learning,* 24: 235-243.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122-147.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bardovi-Harlig, K. And Hartford, B. S. (1997). *Beyond Methods: Components of Language Teacher Education* (1st ed.). McGraw-Hill Humanities/Social Sciences/ Languages.

Bell, J. and R. Gower (1998). 'Writing course materials for the world: a great compromise' in Tomlinson, B. (ed) *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, 116-129.

Berk, L. (2001). *Awakening children's minds*. New York, NY: Oxford University Press.

Berko-Gleason, J. and Murphy, R.S. (2011). 'The Wug Test and Major Developments Since Then: The Jean Berko Gleason Interview'. *ETJ Journal* Vol. 9: Issue 1 http:/ /ltprofessionals.com/en/resources/etjjournal.html

Breen, M. (1984). 'Process syllabuses for the language classroom' in Brumfit, C.J.

General English Syllabus Design. Oxford: Pergamon

Brown, H.D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy.* New York: Longman.

Brumfit, C. J. (1981). 'The notional syllabus revisited' *Applied Linguistics*, Vol 2/1, 83-89.

Brumfit, C. J. (1984). *General English Syllabus Design.* Oxford: Pergamon

Bruner, J. (1960). *The process of education*. Cambridge, MA.

Buchweitz, A., Mason, R. A., Hasegawa, M., and Just, M. A. (2009). Japanese and English sentence reading comprehension and writing systems: An fMRI study of first and second language effects on brain activation. *Bilingualism*. 28, 141-151.

Carter, R. A. and McCarthy, M. J. (eds) (1988). *Vocabulary and Language Teaching*. London: Longman.

Cheng, C. L. (1999). Constructing self-representations through social comparison in peer relations: The development of Taiwanese grade-school children. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 59(9-B), 5132.

Childs. A. (2018). Summary, unit 2 week 8. [Class lecture slide]. Retrieved from https://weblearn.ox.ac.uk/

Cook, V. (2001). *Second Learning and Language Teaching*. London: Arnold.

Cox, D. (2005). 'Can we predict language items for open tasks?' in Edwards, C. and J. Wallis (eds) *Teachers Exploring Tasks: in English Language Teaching.* Cambridge: New York: Palgrave Macmillan, 171-186.

Csikszentmihalyi, M. (2008). *Flow: The Psychology of Optimal Experience* (1st ed.). Harper Perennial Modern Classics.

Cunningsworth, A. (1995). *Choosing your Coursebook*. London: Longman.

Damon, W & Lerner, R. M. (Eds.), *Theoretical models of human development. Handbook of child psychology* (6th ed., Vol. 1, pp. 313-399). New York: Wiley.

Dawson, T. (2018). Personal communication.

Dawson, T. (2019). Personal communication.

Dawson, T. and Stein, Z. (2011). *Virtuous Cycles of Learning.* Downloaded March 18,

2018 https://s3.amazonaws.com/lecticalive/assets/pdf/2012_0929_virtuous_cycles.pdf

Donnelly, J. (2005). Reforming science in the school curriculum: a critical analysis, *Oxford Review of Education,* 31 (2) pp. 293-309

Dörnyei, Z. and Ushioda, E. (2009). *Motivation, Language Identity and the L2 Self.* Multilingual Matters.

Dörnyei, Z. and Ushioda, E. (2011). *Teaching and Researching: Motivation* (2nd Edition). Pearson ESL.

Dörnyei, Z. (2001). *Motivational Strategies in the Language Classroom.* Cambridge University Press.

Dörnyei, Z. (2005). *Psychology of the Language Learner: Individual Differences In Second Language Acquisition* (New Ed.). Lawrence Erlbaum Associates.

Dörnyei, Z. (2008). *The Psychology of Second Language Acquisition*. Oxford University Press, USA.

Dörnyei, Z. and Murphy, R. S. (2010). Where does psychology and second language acquisition research connect? An interview with Zoltán Dörnyei. *The Language Teacher*, 34(2), 19-23.

Dörnyei, Z., Muir, C., and Ibrahim, Z. (2014). 'Directed Motivational Current': Energising language learning through creating intense motivational pathways. In Lasagabaster, D., Doiz, A., and Sierra, J. (Eds.), *Motivation and foreign language learning: From theory to practice* (pp. 9-29). Amsterdam: John Benjamins.

Dörnyei, Z. and Ushioda, E. (2009). *Motivation, Language Identity and the L2 Self.* Multilingual Matters.

Dörnyei, Z. and Ushioda, E. (2011). *Teaching and Researching: Motivation* (2nd Edition). Pearson ESL.

Dörnyei, Z. (2001). *Motivational Strategies in the Language Classroom.* Cambridge University Press.

Duran, G. and Ramaut, G. (2006). ‘Tasks for absolute beginners and beyond: developing and sequencing tasks at basic proficiency levels’ in Van den Branden (ed) *Task-Based Language Education: from theory to practice.* Cambridge: Cambridge University Press, 47-75.

Edwards, C. and J. Willis (2005). *Teachers Exploring Tasks: in English Language Teaching.* Cambridge: New York: Palgrave Macmillan.

Ellis, R. (1984). *Classroom Second Language Development*. Oxford: Pergamon.

Ellis, R. (1998). 'The evaluation of communicative tasks' in Tomlinson, B. (ed) *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, 217-38.

Ellis, R. (2002). 'Methodological options in grammar teaching materials' in Hinkel, E. and Fotos, S. (eds) *New Perspectives on Grammar Teaching in Second Language Classrooms*. New Jersey: Lawrence Erlbaum Associates, 155-179.

Ellis, R. (2003). *Task-Based Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. (2012). *Language Teaching Research and Language Pedagogy* (1st ed.). Wiley- Blackwell.

Erduran S. and Dagher, Z.R. (2014). *Reconceptualizing the Nature of Science for Science Education*. Springer Academic Publishers, New York

Evans, R. (2001). *The Human Side of School Change: Reform, Resistance, and the Real-Life Problems of Innovation* 1st Edition. Jossey-Bass

Fahnestock, J. (2005). Rhetoric of Science: Enriching the Discipline, *Technical Communication Quarterly*, 14:3, 277-286

Finocchiaro, M. and C. Brumfit (1983). *The Functional-Notional Approach*. Oxford: Oxford University Press.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87, 477-531.

Fischer, K. W. and Bidell, T. R. (2006). Dynamic development of action, thought, and emotion. In W. Damon & R. M. Lerner (Eds.), *Theoretical models of human development. Handbook of child psychology* (6th ed., Vol.1, pp. 313-399). New York: Wiley.

Fischer, K. W. and Yan, Z. (2002). The development of dynamic skill theory. In R. Lickliter & D. Lewkowicz (Eds.), *Conceptions of development: Lessons from the laboratory*. Hove, U.K.: Psychology Press.

Fischer, K.W. (2008). Dynamic cycles of cognitive and brain development: Measuring growth in mind, brain and education. In A. M. Battro, K. W. Fischer, and P Lena (Eds.), *The educated brain.* 127-150. Cambridge, U.K.: Cambridge University Press

FLANG (2007). http://www.flang.keio.ac.jp/webfile/AWC/AWC2007/UG1_RU.pdf

Froese, T., Iizuka, H., and Ikegami, T. (2014). Embodied social interaction constitutes social cognition in pairs of humans: A minimalist virtual reality experiment. *Scientific Reports* 4, Article number: 3672

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks.* New York, NY, US: Psychology Press.

Graves, K. (2000). *Designing Language Courses: A Guide for Teachers* (1st ed.). Heinle & Heinle.

Hardiman, M. (2012). *The Brain Targeted Teaching Model.* California: Corwin.

Harwood, N. (Ed.). (2010). *English Language Teaching Materials: Theory and Practice.* Cambridge University Press.

Helgesen, M. (2012). Personal communication. Unpublished.

Holliday, A. (1994). *Appropriate Methodology and Social Context*. Cambridge: Cambridge University Press.

Holloway, J. H., (2000). *Preparing Teachers for Differentiated Instruction*. Educational Leadership, 58(1).

Hong, J. C, Hwang, M.Y., Lu, C. H., Cheng, C. L, Lee, Y.C, & Lin, C. L. (2009). Playfulness- based design in educational games: A perspective on an evolutionary contest game. *Interactive Learning Environments*, 17(1), 15-35.

Hornby, A.S. (1954). *Guide to Patterns and Usage in English*. Oxford: Oxford University Press.

Howard-Jones, P., Demetriou, S., Bogacz, R., Yoo, J. H., Leonards, U. (2011). 'Toward a Science of Learning Games'. *Mind, Brain, and Education.* 5(1), 33-41.

Hutchinson, T. and E. Torres (1994). 'The textbook as agent of change'. *ELT Journal* 48/4: 315-28.

Hutchinson, T. and Waters, A. (1987). *English for Specific Purposes.* Cambridge: Cambridge University Press.

Immordino-Yang, M.H. (2013). Studying the Effects of Culture by Integrating Neuroscientific with Ethnographic Approaches. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, 24(1), 42-46

Immordino-Yang, M.H. and Faeth, M. (2010). In Sousa, D. A. (Ed.), *Mind, Brain, and Education: Neuroscience Implications for the Classroom* (1st ed.). Indiana: Solution

Tree.

Immordino-Yang, M.H. and Damasio, A.R. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Jossey-Bass Reader on the Brain and Learning* (pp.183-198). San Francisco, CA: Jossey-Bass.

Immordino-Yang, M.H., McColl, A., Damasio, H., and Damasio, A. (2009). Neural correlates of admiration and compassion. *PNAS.* 106(19) 8021-8026.

Jenkins, E. (2007). School science: a questionable construct? *Journal of Curriculum Studies,* 39 (3) pp. 265-282

Johnston, H. (2012). *The Spiral Curriculum. Research into Practice*. EPI

Kelly, C. (2012). Personal communication. Unpublished.

Kelly, C. and Murphy, R. S. (2017). *What neuroscience tells us about language learning*, Plenary Session, FAB11: KOTESOL, http://fab-efl.com/page8/index.html

Kagan, D.M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62(2) pp. 129-169

Kennedy, B. (1994). *The development of self-understanding in adolescence*. Dissertation

Kimball, M.E. (1968). Understanding the nature of science: A comparison of scientists and science teachers. *Journal of Research in Science Teaching*, 5(2), pp.110–120.

Krashen, S. and T.D. Terrell (1983). *The Natural Approach*. Hayward, CA: Alemany.

Kuhn, T.S. (1996). *The structure of scientific revolutions* (3rd ed.) Chicago: The University of Chicago Press

Larsen-Freeman, D. and Anderson, M. (2011). *Techniques & Principles in Language Teaching* (3rd ed.). Oxford: Oxford University Press.

Lederman, N.G., (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), pp.331–359.

Lederman, N.G., (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), pp.497–521.

Ledoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional*

Life. New York: Simon & Schuster.

LeDoux, J. (2002). *Synaptic Self: How Our Brains Become Who We Are.* New York: Viking Adult.

Lee, E. and Luft, J. A. (2008). Experienced Secondary Science Teachers' Representation of Pedagogical Content Knowledge. *International Journal of Science Education*, 30(10), 1343–1363. https://doi.org/10.1080/09500690802187058

Liebenberg, L. (2017). *What are 21st century skills?*

Lightbown, P. and Spada, N. (1999). *How Languages are Learned.* Oxford: Oxford University Press.

Loftus, E. F. (1997). Creating False Memories. *Scientific American.* Vol 277 #3:70-75

Long, M. (1985). 'Input and second language acquisition theory' in Gass, S. and C. Madden (eds) *Input in Second Language Acquisition*. Cambridge, MA: Newbury House Publishers, 377-393.

Long, M. and G. Crookes (1991). 'Three approaches to task-based syllabus design'. *TESOL Quarterly* 26: 27-55.

McCarthy, M. J., McCarten, J., & Sandiford, H. (2005). *Touchstone Level 1 Workbook.* Cambridge University Press.

McDonough, J. and C. Shaw (2003). *Materials and Methods in ELT.* Second edition. Oxford: Blackwell.

McGrath, I. (2002). *Materials Evaluation and Design for Language Teaching.* Edinburgh: Edinburgh University Press.

Medina, J. (2009). *Brain Rules: 12 Principles for Surviving and Thriving at Work, Home, and School* (Reprint.). Seattle: Pear Press.

Meltzer, L. (2007). *Executive Function Education.* New York: Guilford Press

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*. 16(5):1936-1947.

Murphy, R. S. (2009). *Models for EFL theory and methodology derived from an SiR-based pilot study on Japanese cognitive development*, Published MA Dissertation, University of Birmingham.

Murphy, R. S. (2012). *Creating Champion Students with Neuro-based Pedagogy*.

http://www.youtube.com/watch?v=-SQYnlz5NwY&feature=g-upl

Murphy, R. S., Stubbings, C, & Uemura, T. (2017). NeuroELT perspectives: Directed Motivational Currents--Theoretical, Experimental, and Pedagogical Research--, University of Kitakyushu, *Kiban Center Kiyo Journal*, Vol 29

Murphy, R. S. (2015). *Why you should know about Dynamic Skill Theory*. Featured Presentation, KOTESOL

Murphy, R. S. (2016). *neuroELT Maxim: Decide Top-down vs Bottom-up teaching ratios*, https://www.youtube.com/watch?v=9WtQb4gs8KI

Murphy, R. S. (2017). 3D:CG (Constant Gradient) Evaluation System: New Rubrics for the Evaluation of Teaching and Professional Development Programs. *Kiban Center Kiyo Journal*, Vol 28, University of Kitakyushu

Niggs, J. T. (2009). *What Causes ADHD? Understanding what Goes Wrong and Why*. New York: Guilford Press.

Nott, M. and Wellington, J. (1998). *Science & Education* 7: 579.

Nunan, D. (1987). 'Communicative language teaching: making it work'. *English Language Teaching Journal* 41/2:136-145

Nunan, D. (1988). *Syllabus Design*. Oxford: Oxford University Press.

Nunan, D. (1989). *Designing Tasks for the Communicative Classroom.* Cambridge: Cambridge University Press.

Nunan, D. (1993). 'Task-based syllabus design: selecting, grading and sequencing tasks in Gass, S.M. and G. Crookes (eds) *Tasks in a Pedagogical Context*. Clevedon: Multilingual Matters, 58-59.

Nunan, D. (1999). *Second Language Teaching and Learning.* Boston, MA: Heinle and Heinle.

O'Neill, R. (1977). 'The limits of functional/notional syllabuses or My guinea pig died with its legs crossed'. In *English for Specific Purposes*, ed. Susan Holden. Mansfield, U.K: Modern English Publications.

O'Dwyer, F., Noriko, N., Collett, P., Sullivan, K., and Smith, A. (2011). 'Framework & Language Portfolio SIG Forum: Looking forward'. In A. Stewart (Ed.), *JALT2010 Conference Proceedings*. Tokyo: JALT.

Pajares, M.F. (1992). Teachers' beliefs and educational research: Cleaning up a messy

construct. *Review of Educational Research* Vol. 62, No. 3 pp. 307-332

Palmer, H.E. [1922] (1968). *The Principles of Language Study*. Oxford: Oxford University Press.

Pica, T., R. Kanagy and J. Falodun (1993). 'Choosing and using communication tasks for second language instruction' in Crookes, G. and S. Gass (eds.) *Tasks and Language Learning, Integrating Theory and Practice,* 9-34.

Pink, D. H. (2009). *Drive: The Surprising Truth About What Motivates Us* (Reprint.). Riverhead Trade.

Posner, D. (2004). What's Wrong with Teaching to the Test? *Phi Delta Kappan*, Sage Publishing. Vol 85, Issue 10, pp 749-751

Prabhu, N.S. (1987). *Second Language Pedagogy.* Oxford: Oxford University Press.

Pustejovsky, J. (1998). *The Generative Lexicon.* A Bradford Book.

Prabhu, N.S. (1987). *Second Language Pedagogy.* Oxford: Oxford University Press.

Pu, Y., Liu, H.-L., Spinks, J. A., Mahankali, S., Xiong, J., Feng, C.-M., et al. (2001, June). Cerebral hemodynamic response in Chinese (first) and English (second) language processing revealed by event-related functional MRI. *Magnetic Resonance Imaging,* 19, 643-647.

Pustejovsky, J., Bouillon, P., Isahara, H., Kanzaki, K., and Lee, C. (Eds.). (2012). *Advances in Generative Lexicon Theory* (2013rd ed.). Springer.

Ralabate, P.K., and Nelson, L. (2017). *Culturally Responsive Design for English Learners, The UDL Approach,* CAST

Richards, J. (1998). *Beyond Training*. Cambridge: Cambridge University Press.

Richards, J. and Rodgers, T. (2001). *Approaches and Methods In Language Teaching*. Cambridge: Cambridge University

Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in Language Teaching: An Anthology of Current Practice.* Cambridge University Press.

Richards, J.C. (2001). *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.

Robinson, P. (2001). (ed) *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.

Robinson, P. (2011a). *Task-Based Language Learning* (1st ed.). Wiley-Blackwell.

Robinson, P. (2011b). *Second Language Task Complexity: Researching the Cognition Hypothesis of language learning and performance.* John Benjamins Publishing Company.

Rosen, C. (2008). The Myth of Multitasking. *The New Atlantis*, 20,105-110.

Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary Teachers' Mathematics Subject Knowledge: the Knowledge Quartet and the Case of Naomi. *Journal of Mathematics Teacher Education,* 8(3), 255–281. https://doi.org/10.1007/s10857-005-0853-5

Rubdy, R. (2003). 'Selection of materials' in Tomlinson, B. (ed) *Developing Materials for Language Teaching.* London: Continuum, 37-57.

Samuda, V. and Bygate, M. (2008). *Tasks in Second Language Learning.* (C. N. Candlin & D. R. Hall, Eds.) (First Edition.). Palgrave Macmillan.

Scaife, J. (2008). Science Learning, *Science Teaching:* Chapter 4. New York, Routledge

Schacter, D. (1997). *Memory distortion.* Cambridge, USA: Harvard University Press

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*. 67(8):603-13

Schacter, D., Addis, D., and Buckner, R., (2007). Remembering the past to imagine the future: the prospective brain. Nature Reviews Neuroscience, 8, 657-661

Schmitt, J.B., Debbelt, C.A., & Schneider, F.M. (2017). 'Too much information? Predictors of information overload in the context of online news exposure'. *Information, Communication & Society*

Schmitt, N., and McCarthy, M. (1997). *Vocabulary: Description, Acquisition and Pedagogy.* Cambridge University Press.

Schwartz, B. (2004). *The Paradox of Choice*. New York: Harper Perennial.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency* (Cognitive Science and Second Language Acquisition Series). Taylor and Francis. Kindle Edition.

Seung, S. (2013). *Connectome: How the Brain's Wiring Makes Us Who We Are*. Boston: Houghton Mifflin Harcourt Publishing.

Shaw, S. (2016). *The memory Illusion*. London, UK: Penguin Random House

Shizgal, P., Arvanitogiannis, A. (2003) Gambling on dopamine. *Science* 299:1856-1858.

Shulman, L. S. (1986). *Knowledge Growth in Teaching,* 15(2), 4–14. Retrieved from http://www.jstor.org/stable/1175860

Shulman, L. S. (2004). *The wisdom of practice: Essays on teaching, learning, and learning to teach.* S. Wilson (Ed.) San Francisco: Jossey-Bass, Inc.

Sinclair, J. M. and Renouf, A. (1988). 'A lexical syllabus for language learning' in Carter and McCarthy (1988).

Skehan, P. (1996). 'Second language acquisition research and task-based instruction' in Willis, J. and Willis, D. (eds) *Challenge and Change in Language Teaching.* Oxford: Heinemann, 17- 30.

Skehan, P. (1998). *A Cognitive Approach to Language Learning.* Oxford: Oxford University Press.

Sousa, D. A. (2011). *How the Brain Learns* (Fourth Edition.). Thousand Oaks, CA: Corwin Press.

Sousa, D. A., & Tomlinson, C. A. (2010). *Differentiation and the Brain: How Neuroscience Supports the Learner-Friendly Classroom* (1st ed.). Indiana: Solution Tree.

Sousa, D.A. (2016). *How the Brain Learns*. Corwin. New York

Stern, H.H. (1992). *Issues and Options in Language Teaching*. Oxford: Oxford University Press.

Sylwester, R. (2002). *A biological brain in a cultural classroom* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Tagg, C. and Woodward, E. (2011). *Syllabus and Materials*. University of Birmingham: CELS.

Thornbury, S. (2011) In Simpson, J. (Ed.). *The Routledge Handbook of Applied Linguistics* (1st ed.). Routledge.

Tokuhama-Espinosa, T. (2010). *Mind, brain, and education science., a comprehensive guide to the new brain-based teaching.* NY, USA: Norton

Tokuhama-Espinosa, T. (2010). *The New Science of Teaching and Learning: Using*

the Best of Mind, Brain, and Education Science in the Classroom. New York: Teachers College Press.

Tokuhama-Espinosa, T. (2014). *Making classrooms better, 50 practical applications of mind, brain, and education science.* NY, USA: Norton

Tokuhama-Espinosa, T. (2018). *What neuroscience tells us about how to teach foreign languages*. (downloaded, July 2018) http://traceytokuhama.com

Tomlinson, B. (1998). *Materials Development in Language Teaching.* Cambridge: Cambridge University Press.

Tomlinson, B. (2003). *Developing Materials for Language Teaching*. Continuum.

Tomlinson, B. (Ed.). (2008). *English Language Learning Materials: A Critical Review*. Continuum.

Tomlinson, B. (Ed.). (2010). *English Language Learning Materials: A Critical Review* (1st ed.). Continuum.

Tomlinson, B. (Ed.). (2011). *Materials Development in Language Teaching* (2nd ed.). Cambridge University Press

Tomlinson, C. A. (1995). Differentiating instruction for advanced learners in the mixed-ability middle school classroom. *ERIC Digest* E536.

Tsai, C. (2002). Nested epistemologies: Science teachers' beliefs of teaching, learning and science. *International Journal of Science Education,* 24(8), pp.771–783.

Valaki, C. E., Maestu, R, Simos, P. G., Zhang, W., Fernandez, A., Amo, C. M., et al. (2004). Cortical organization for receptive language functions in Chinese, English, and Spanish: A cross- linguistic MEG study. *Neuropsychologia,* 42(1), 967-979.

Van Avarmaet, P. and S. Gysen (2006). 'From needs to tasks: language learning needs in a task-based approach' in Van den Branden, K. (ed) *Task-Based Language Education: from theory to practice.* Cambridge: Cambridge University Press, 17-46.

Van den Branden, K. (2006). T*ask-Based Language Education: from theory to practice.*

Vygotsky, L. S. (1978). M*ind in Society: The Development of Higher Psychological Processes* (14th ed.). Harvard University Press

Watcyn-Jones, P. (1981). *Pair Work.* Harmondsworth: Penguin.

Watson, F. G. (1983). SCIENCE EDUCATION: A DISCIPLINE? *National*

Association for Research in Science Teaching, John Wiley & Sons, VOL. 20, NO. 3, PP. 263-264

West, M. (1953). *A General Service List of English Words*. London: Longman.

Wilkins, D.A. (1976). *Notional Syllabuses*. Oxford: Oxford University Press.

White, R. (1988). *The ELT Curriculum*. Oxford: Blackwell.

Wexler, B.E. (2006). *Brain and Culture*. Cambridge: The MIT Press.

Widdowson, H. G. (1979). *Explorations in Applied Linguistics*. Oxford: Oxford University Press.

Widdowson, H. G. (1988). 'Grammar, and nonsense, and learning' in Rutherford, W. and M. Sharwood Smith (eds.) *Grammar and Second Language Teaching: a Book of Readings*. New York: Longman, 146-155.

Widdowson, H. G. (1990). *Aspects of Language Teaching*. Oxford: Oxford University Press.

Willingham, D. T. (2009). *Why Don't Students Like School: A Cognitive Scientist Answers Questions About How the Mind Works and What It Means for the Classroom* (1st ed.). Jossey- Bass.

Willis, D. (1990). *The Lexical Syllabus.* Oxford: Macmillan Heinemann.

Willis, D. (1996). 'Accuracy, fluency and conformity' in Willis, J. and Willis, D. (eds) *Challenge and Change in Language Teaching.* Oxford: Heinemann, 44- 51.

Willis, J. (1990). First Lessons: Real English for Real Beginners. Collins CoBUILD.

Willis, J. (1996b). 'A flexible framework for task-based learning' in Willis, J. and Willis, D. (1996a) (eds) *Challenge and Change in Language Teaching.* Oxford: Heinemann, 52-62.

Willis, J. (1998). 'Concordances in the classroom without a computer: assembling and exploiting concordances of common words' in Tomlinson, B. (ed) *Materials Development in Language Teaching*. Cambridge: Cambridge University Press.

Willis, J. and Willis, D. (1996a). (eds) *Challenge and Change in Language Teaching.* Oxford: Heinemann.

Willis, J. and Willis, D. (1996b). 'Consciousness-raising activities in the language classroom

Willis, J. and Willis, D. (2007). *Doing Task-Based Teaching*. Oxford: Oxford University Press.

Van den Branden, K. (2006). *Task-Based Language Education: from theory to practice.* Cambridge: Cambridge University Press.

Yan, Z., and Fischer, K. (2002). Always under construction, Dynamic variations in adult cognitive microdevelopment. *Human Development*. 45:141-160

Young, G. (2011). Development and causality. *Neo-Piagetian perspectives*, New York, Springer

Appendix

Appendix A. Tier 3 modification list

List of modification numbers ranging from 4 – 34.

Modification 1 → Instead of the simpler notebook styled "blank lines" of *OL2,* I have now added writing prompts.

Modification 2 → This modification has the students plan for the future, making this page not only a reflection page, and aha moment page, but also a page that connects to the future, making this page now a forward-planning page as well.

Modification 3 → In Japanese I have specifically asked them to imagine "Work environments that use English". I proposed that this specific prompt/question should be in their L1 to alleviate the chance that it might be misinterpreted.

Modification 4 → I have added very friendly looking characters to the design.

Modification 5 → Engaging graphics. The graphics and placements have been modernized.

Modification 6 → (Type 1: This was an ignored Cycle) Cycle 2 was much too simplistic. It did not significantly engage the student's partner's mind. So, here I have added a new task that requires the partner to write two sentences using the student's words from above.

Modification 7 → Added emotional valence per lexical item.

Modification 8 → (Type 1: This was an ignored Cycle) Social learning and *significant* partner help is key for this proposed change. Here, the student's partner is required to of new lexical items for the sake of the student—not for themselves. This is a solvable mystery/aha moment activity that also helps to bond the students.

Modification 9 → (Type 1: Cognitive overload.) Initially, this task had the students have this page open while the students listen to the teacher discuss potential usage of various new

lexical items. The students will now be asked to write all of the lexical items down (list provided on screen)—then, the teacher talks about the lexis in the context of the upcoming presentation theme. While that is unfolding, the students simply circle Y/N under each word. This will produce much less cognitive overload. And, there will be much fewer spelling issues and misheard words. This small change will likely make this activity much more effective as a teaching opportunity, and more affective for the students L2 development.

Modification 10 → Emotional valence bubbles added.

Modification 11 → (Type 1: Students did not understand the task.) In the interviews, it became clear that the purpose of this Master List was lost on some students. Here I have changed the wording in English and in Japanese to reinforce the nuance that these sentences are for the upcoming presentation. In this way, it helps reinforce the idea of future planning in the L2, helping make the idea of using L2 in the future more of a reality for the students.

Modification 12 → Added emotional valence bubbles.

Modification 13 → (Type 1: Sidebar was not always used effectively) Added emotional valence bubble into the yellow side-bars.

Modification 14 → To help this page become more aesthetically pleasing, welcoming, and to help raise motivation, the colors have been changed to a warmer color, and pleasant characters' faces have been added to the page.

Modification 15 → The graphic from the opening of the chapter has been carried over to here, the presentation assessment page, to bring this study to a "full circle".

Modification 16 → (Type 1: Underused/misused by students) I have added explicit instructions to this page.

Modification 17, 18. 19 → New page, "Level Up Time!". Discussed in detail in above section.

Modification 20 → Gold color has been added to the title for horizontal coherence.

Modification 21, 22 → (Type 1: Wasted time. Test design autonomy limited by 50%.) Mandated Section 1 to be a “Vocabulary Section”. Mandated Section 3 to be a “Grammar Section”. Sections 2 and 4 are still fully autonomous.

Modification 23 → Added friendly Japanese subtitles.

Appendix B: Lectical analysis test

RFJ001: (conducted with online version of this content)



Some scientists think that all violent TV shows are **bad** for children, because they can teach children to hurt other people. Other scientists disagree. They think it is **okay** for children to watch some violent TV shows.

Some scientists think that violent TV shows are bad for children. Others think some violent TV shows are okay.

a. Which group of scientists do you think is right? (Check the box next to your answer.)

o The scientists who think violent TV is bad for children

o The scientists who think some violent TV is okay

o Neither group

o Both groups

a. Explain your choice.

_____

_____

_____

How would you decide which group of scientists was right?

b. If you were trying to decide which group of scientists was right, what would you need to know? Why?

_____

_____

_____

_____

_____

c. If you were trying to decide which group of scientists was right, where would you look for more information? Why?

_____

_____

_____

_____

_____

d. If you were trying to decide which group of scientists was right, how would you know if you could trust what you learned?

_____

_____

_____

_____

_____

e. If you were trying to decide which group of scientists was right, what else could you do to help you decide? Why would that help?

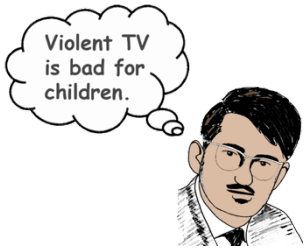
_____

_____

_____

_____

_____



If you were one of the scientists who thought violent TV was **bad** for children, what could you do to convince the other group of scientists that you were right?

a. Describe what you would do to convince the other scientists.

_____

_____

_____

_____

_____

b. Explain why.

_____

_____

_____

_____

_____



If you were one of the scientists who thought some violent TV was **okay** for children, what could you do to convince the other group of scientists that you were right?

a. Describe what you would do to convince the other scientists.

_____

_____

_____

_____

_____

b. Explain why.

How is it possible that the two groups of scientists have such different ideas?

c. Explain your answer.

Is it possible to know for sure if violent television is bad for children?

a. Check the box next to your answer.

θ Yes θ No θ Maybe

b. Explain your answer.

Appendix C: Lectical analysis, collected student responses

Each line is the data field that emerged per test. This was the exact data that was coded. The responses in each string correspond to the following questions:

First name, Last name, test time (1/pre or 2/post), RFJ test number, date, word count, lectical score, Lectical phase, spelling, capitalization, punctuation, other grammar, vocabulary, relevance, repetition, coherence, persuasiveness, judgement justifications, LEC probe 1, LEC main response

Keiichi Sono 1 RFJ003 4/24/14 2:03 294 10.12 10a 8 9 9 5 8 8 8 6 5 8 the scientists who think bullying is always bad Bullying is bad for children; because it is not good for children's body and heart. It is bad for children's health. And children have their future. If children will be old; they should think about bullying; and move; themselves.
Keiichi Sono 2 RFJ001 7/31/14 1:32 227 10.03 10a 8 9 9 6 8 8 7 6 4 8 the scientists who think violent TV is bad I think the scientists who think violent TV is bad is right. Violent TV gives children knowledges of violences. I have heard this story. Some violent story by children are comes from childres's mistake that they couldn't understand different between game world and real world.
yuuna kanmura 1 RFJ001 4/24/14 1:59 203 10.08 10a 5 2 6 4 7 6 7 3 4 6 the scientists who think violent TV is bad i agree the scientists who think violent tv is bad.because children are flue by around invilonment.so;children that watch the violent tv might be violence to other people.i think pearents must be watch the tv show that children have a good impact.
yuuna kanmura 2 RFJ003 7/31/14 1:32 329 10.33 10b 6 8 8 3 7 7 8 5 5 8 the scientists who think a little bullying can be okay I think little bullying is can be okey for childlen. Because sometime children do bad things. Adult must teach the thing that children do now is bad thing. but;children deaf the adult's tallk. So;if children don't listen the adult's talk;need a little bullying.
yukiko kajimaru 1 RFJ001 4/24/14 1:56 230 10.01 10a 6 8 8 3 7 8 7 5 3 6 the scientists who think violent TV is bad I think it is light that violent TV is bad. Because violent TV has bad scene for children. children should learn things peace is very important.
yukiko kajimaru 2 RFJ003 7/31/14 1:29 321 10.15 10a 6 9 9 5 6 6 8 6 3 7 the scientists who think bullying is always bad I thinkã€€bullying is always bad. Because it makes children sad and link to bad grow. So children who bully should be scolded by teacher and their parents.
MAKOTO SAKAMAKI 1 RFJ001 4/24/14 2:02 226 10.08 10a 6 9 9 4 7 7 6 5 4 6 the scientists who think some violent TV is okay I think that violent TV shoows are bad for children because children has a delicate heart. Most of children don't know terrible violent. Children has nice sympathy and can't refuse it. It is going to became trauma. It is important that understand terrible real. But it shows for growing children.
MAKOTO SAKAMAKI 2 RFJ003 7/31/14 1:32 359 10.37 10b 6 9 8 3 7 8 8 6 4 8 the scientists who think a little bullying can be okay I think that little bullying can be okay.Because little bullying makes them growth. I experienced it when I am junior high school students. I talk to friends other friend's bad point before I experienced little bullying. But;I stopped to talk friend's bad point after little bullying.And I can understand people's mind.I think that little bullying makes me growth. But;heavy bullings injure mind.This is bad. So I think that little bullying can be okay.
SHUJI KURODA 1 RFJ001 4/24/14 1:52 114 10.10 10a 7 9 8 6 8 8 7 5 4 7 the scientists who think violent TV is bad Because;children become more viorent. For exanple;parents watch violent TV. Children are watch violent movie with their parents. So;This is bad education.
SHUJI KURODA 2 RFJ003 7/31/14 1:00 100 9.78 09d 5 9 9 3 6 8 6 5 3 8

both groups Bullying is important for children. Because they can't undersutand good thing and bad thing. So adult's lecture are important.
SENA EBII 1 RFJ003 4/24/14 1:32 151 10.32 10b 8 9 9 5 8 7 8 8 4 8 the scientists who think bullying is always bad Bullying is very very bad because people hurt someone.
SENA EBII 2 RFJ001 7/31/14 1:21 373 10.45 10b 7 9 9 6 8 7 5 7 5 9 the scientists who think violent TV is bad I think that the scientists who think violet TV is bad because the violet TV give childrens only bad effect. For example; to hurt others and angry or fear. If children watch the violet TV; children do bad behavior. So; children don't have to watch the violet TV.
HARUKA NAGATA 1 RFJ003 4/24/14 1:53 159 10.14 10a 8 9 9 6 9 8 7 6 3 8 the scientists who think a little bullying can be okay It is important for children to bully a little .Bullying is harder for them;but bullying makes children stronger and they may be kind everyone.
HARUKA NAGATA 2 RFJ001 7/31/14 1:23 355 10.05 10a 7 9 9 5 8 8 6 6 4 8 the scientists who think violent TV is bad Iã€€think violent TV is bad for children because violent TV makes childeren violent pople . Children think viorent is okay by looking violent TV; but all violent TV is not bad.
YUURI YOSHINO 1 RFJ003 4/24/14 1:54 351 10.33 10b 8 9 9 5 7 7 8 7 5 8 the scientists who think bullying is always bad I think that bullying is always bad for children.If there is a little bullying in class;it must become a stronger bullying.Children are still weak.They cannot distinguish a good thing and a bad thing.Teachers should teach that bullying is one of the bad thing to injure for other people.So I think the scientists who think bullying is always bad is right.
YUURI YOSHINO 2 RFJ001 7/31/14 1:28 432 10.18 10a 7 9 9 5 8 7 7 6 4 8 the scientists who think violent TV is bad I think that the scientists who think violent TV is bad is right. These days;I often see very bad news occuring by children. Sometimes now age's children is said that becomes more even than before. But I think if violent TV shows hadn't been allowed;such bad news aren't occured by children.Watching violent TV becomes children exciting to violent.So I think violent TV is bad.
NATSUKO MORODOMI 1 RFJ003 4/24/14 1:57 233 10.41 10b 8 9 9 6 7 8 9 6 6 8 both groups I think that we cannot say whether bullying right or not; in each positions. Because I have an idea that bullying is a way of expressions from childen's emotions. we sometimes think that it is bad not bully people but bullyed people. But; I think that we need have many views in my mind.
NATSUKO MORODOMI 2 RFJ001 7/31/14 1:21 284 10.60 10c 6 9 9 5 7 8 6 5 4 8 both groups I think that violent TV show effect children. But the perspnal characters not only made by the effection of that TV show. Personal character made by culuture ; environment and education and so on. Because I believe that violent TV shows are okay.
SAKI SEGAWA 1 RFJ001 4/24/14 2:00 231 9.94 09d 7 9 9 7 8 8 6 6 5 7 the scientists who think some violent TV is okay I think the group of the scientists who think some violent TV is okey is right because children are able to learn what they hurt other people some time from the TV. On other words they can learn to kind for other people.
SAKI SEGAWA 2 RFJ003 7/31/14 1:31 406 10.07 10a 5 9 9 4 7 8 7 4 3 8 the scientists who think bullying is always bad I think that the scientists who think bullying is always bad is right. It is because bullying does harm to children. And the children may be agrressive. This is very scarely. Plus bullying is said bad in generally. So bullying is always bad for children.
TOMOHA HAYASHI 1 RFJ003 4/24/14 1:59 141 10.24 10a 6 9 9 5 8 8 8 7 4 6 the scientists who think bullying is always bad Bullying is most bat things . Children is very delicate. So; if children is bullied;they rereceive much damage in their hart.
TOMOHA HAYASHI 2 RFJ001 7/31/14 1:30 385 10.27 10b 6 9 8 5 7 8 7 5 5 8 the scientists who think some violent TV is okay I think that some violent TV shows are okey for children. Bcause it is chance for children to terch that a

viorent is bad thing. When they show some violent TV;they show a hart people. Then; they can sure that violent is not good and they have to have kind hart to other people. If we ban to show some violent TV for children;then they don't learn "violent is bad". But;when they show some violent TV;adults must terch "this doing is wrong. you must not hart other people."
HIROMI OGITA 1 RFJ001 4/24/14 1:57 270 9.84 09d 7 9 9 4 6 7 7 5 3 5 the scientists who think violent TV is bad I think violent TV is bad for children. First of all;violent TV teachs bad behavier for children.So;children hurt friends. Second;violent TV is bad for children's education;because that TV inform bad imfomation. That of all;I think violent TV is bad for children.
HIROMI OGITA 2 RFJ003 7/31/14 1:25 283 10.26 10b 7 9 9 4 8 8 7 8 2 8 the scientists who think a little bullying can be okay I think that a little bullying can be okay.Because a little bulling give children grouth.Plus a little bullying give children many experience.So;I think that a little bullying can be okay.
MITSUKI TAKAKURA 1 RFJ003 4/24/14 1:52 114 9.79 09d 7 9 9 4 8 7 8 4 3 7 the scientists who think bullying is always bad I think bullying is always bad; because bullying injure friends. freinds become very sad.
MITSUKI TAKAKURA 2 RFJ001 7/31/14 1:15 197 9.61 09c 8 9 9 5 8 8 6 6 4 9 the scientists who think violent TV is bad I think that the scientists who think violent TV is bad because violent TV teach children violent. Children may hurt other people like the violent TV. And; I think that not TV teach children violent is bad but teacher and perents.
Yamamoto Nao 1 RFJ003 4/24/14 1:48 319 10.37 10b 8 9 9 6 8 8 7 7 6 8 the scientists who think bullying is always bad I think bullying is to hurt someone. I think it is important for children to know about frend"s feeling. If children bully since small ; they may not judge good and bad after they become adult. That is why I think children must not do bullying. In addition teachers have to look at children carefully not to exist children who are bullied by someone.
Yamamoto Nao 2 RFJ001 7/31/14 1:15 425 10.33 10b 8 9 9 6 8 8 8 7 6 9 the scientists who think violent TV is bad I think viorent TV shows are bad for children. Because children tend to imitate everything. For example;if TV shows dorama about bullyings; then children may have an influence on dorama; and some children can bully others. It is true that children often wach bad TV;but their mother or fahter must be attention not to make children wach bad TV.
EMIKO SHODA 1 RFJ001 4/24/14 1:59 214 10.01 10a 7 9 9 4 7 8 8 5 4 7 the scientists who think violent TV is bad I think that's bad. Because children immitate them who they looked.They don't learn what is right and wrong. So;if children are infulense by some violent TV;many trouble may happen.That's why I think some violent TV shows are bad for children.
EMIKO SHODA 2 RFJ003 7/31/14 1:31 338 10.21 10a 6 9 9 4 6 8 8 6 4 8 the scientists who think a little bullying can be okay I think a little bullying can be okay. Because there are things which bullying tells us. Of course;bullying is never good action. Also;bullying have made many people sad. But some people have learn feeling of people who are bullying;through bullying. That's why I think a little bullying can be okay.
RYOSUKE MAEDA 1 RFJ003 4/24/14 1:56 497 10.41 10b 8 9 9 6 7 8 8 7 6 8 neither group I think some children should be bullied;other children should not be bullied. Children have diversity. I think teacher and parents should observe their children carefuly and then; they should decide to bully them or not.
RYOSUKE MAEDA 2 RFJ001 7/31/14 1:22 518 10.43 10b 8 9 9 6 8 8 7 6 5 9 the scientists who think some violent TV is okay I think some violent TV shows are okey for children.Because;they teach children how weak our lives.By the way;I have head that children who didn't killed some insects when they ware young children could come to cluel.And;I think children should have a lot of experiences to become clever adult.
TAKUYA MATSUI 1 RFJ001 4/24/14 1:53 351 10.13 10a 6 9 9 4 6 8 6 5 5 7 both groups I think both of groups is right Because the violent TV

is bad effect for children but they can learn what is right from the TV. Childrens can't learn how to hurt other peoples if they don't see violent TV; but they can't learn what accions for hurt others. If they don't know that acctions; they hurts someone without their will. It's dengerous. So; they should know what accion can hurt other peoples and don't do that things. That is why I think so.
MANAMI SAEKI 1 RFJ003 4/24/14 1:47 143 9.88 09d 7 9 9 5 7 7 8 6 4 7 the scientists who think a little bullying can be okay Bullying is very important for children. So a little bulling can be okay.
MANAMI SAEKI 2 RFJ001 7/31/14 1:29 292 10.05 10a 5 9 8 6 8 7 7 5 4 8 the scientists who think violent TV is bad I think that violent TV is bad for children. Because it is for children to have bad influence. If they watch it; then they may think that violent want to do. Plus bad craim may be inclease. So violent TV is bad for children.
KOTA SASAKI 1 RFJ003 4/24/14 1:52 192 9.83 09d 9 9 9 4 9 6 6 5 4 7 the scientists who think a little bullying can be okay I think that the teacher a little bullying is ok.Because school is a place children learn good thing and bad thing.For example; If chind do violence other peaple; teacher shold teach the thing is bad thing. To do this; it is important that he teaches a little titly.
KOTA SASAKI 2 RFJ001 7/31/14 1:30 339 10.18 10a 8 9 9 5 8 7 7 6 4 8 the scientists who think violent TV is bad I think that violent TV shows are bad for children because violent TV make violent children.If children show violent seen for example;kick;punch and violent words then the children could speak violent word and do bad thing.Plus;bacause of see them; children have bad society rule and bad judgement skill. If the children see violent seen in school then the children don't think the thing is bad.
YUUKI FUKUDA 1 RFJ003 4/24/14 1:50 162 10.16 10a 8 9 10 5 7 8 8 5 4 8 the scientists who think bullying is always bad I think that bullying is always bad.Because;bulling made us sad.Bullying hurt people;so I should reduce bullying completely. If someone bully; I would tell a teacher.It is most peaceful behavior.
YUUKI FUKUDA 2 RFJ001 7/31/14 1:13 213 9.92 09d 8 9 9 5 8 8 7 5 5 8 the scientists who think some violent TV is okay I think violent TV is okay.Because;there are many violences in the world.So;chirdren have to know it.When they grown up; they behave right way without confusing.
hiroto yamamoto 1 RFJ003 4/24/14 1:48 183 10.39 10b 7 9 9 3 7 8 8 5 4 7 the scientists who think bullying is always bad If boy strike other boy;he may injured. Young boy should learn right.
hiroto yamamoto 2 RFJ001 7/31/14 1:23 435 10.54 10c 8 9 9 6 7 6 2 4 3 8 the scientists who think violent TV is bad I think violent TV shows bad for children because children have not jugdement skill. In society; many people violent TV is bad. If children watch violent TV shows; they may injure other people. Also in feelings; children may have bad emotion. Other people may have same idea.
RIKU MURAKAMI 1 RFJ001 4/24/14 1:52 351 10.61 10c 9 8 9 6 7 4 6 6 6 7 the scientists who think some violent TV is okay I think violent TV is okay. Because children have to learn?socialproblem. children?is thinking about something. And peace;kind;are not all. And violentproblem is one of the most important problem all over the world. It is important for children to think about that problem. I think so.
RIKU MURAKAMI 2 RFJ003 7/31/14 1:33 641 10.69 10c 6 8 9 3 8 4 5 5 4 8 the scientists who think bullying is always bad I think bullying is always bad for children. Because bullying makes people unhappy;and makes bad feelings.So we should stop bullying.Also bullying can be big problem.Because children do not know badthings or goodthings sometime.It is very important things. I think we have to think feelings of children who in bullying. And we have to think how to solve the problem of bullying. We should save children.But on the other hand we should give children power of think by ownself. I think it.
taro hamamoto 1 RFJ001 4/24/14 1:51 226 10.48 10b 6 7 7 4 8 8 7 5 4 7 the scientists who think some violent TV is okay I think children to watch

violent TV is okay. becouse violent TV is not influence to children. I think when children watch violent TV is fear. so children is not play violence action. therefore cildren is watch violent TV. If someone cildren play violence action;but is this action influence on violent TV? IT`s don`t say violence TV is bad. IT has oppotunity that cildren is crazy.
taro hamamoto 2 RFJ003 7/31/14 1:22 464 10.05 10a 5 4 8 3 7 6 8 4 3 8 both groups I think bullying is always bad; because child is too easy break them heart. I think bullying is craim. children had bullied is become bad adalt. plus bad adalt possibully do craim. but I think bullying have good point. we can bully to others. because we smart to other animal. we can bullying is evidence for people is too smart. If we too fool people;we can`t bully to oters.
KENGO KAWAI 1 RFJ001 4/24/14 1:52 150 10.47 10b 6 9 8 5 8 8 7 5 5 7 the scientists who think some violent TV is okay Violent TV is bad for children because it can give children some bad effect. For example; when the children who watched violent TV grow up; they may kill;hurt and injure other people.
KENGO KAWAI 2 RFJ003 7/31/14 1:24 309 10.48 10b 5 9 9 4 7 6 8 6 5 8 neither group I think bulling is not always bad for children; but cannot be okey. Exactly;the children who ware bullied hurted in one's heart. And the children who ware bullying may serious prpblem in future. But it isn't 100%. A boy was bullied feel the power of justise; and become police officer. A girl was bullying become the fighter. So; I think neither group is right.
YURIKA NIRE 1 RFJ003 4/24/14 1:25 150 10.13 10a 6 9 9 4 7 5 8 5 3 8 the scientists who think a little bullying can be okay Bullying is not bad;because it is always happen. Chirdlen are bullying and growing up.
YURIKA NIRE 2 RFJ001 7/31/14 1:16 363 10.07 10a 6 9 9 4 7 8 6 5 4 8 the scientists who think some violent TV is okay I think that some violent is bad. Because;violent TV is very shocking for childlen.And;if we allowed to see violent TV for childlen; thay can violent man or woman. I don`t want to be those human.
YURI MUKAI 1 RFJ001 4/24/14 1:51 257 9.93 09d 8 9 7 5 8 7 6 6 7 8 both groups I think that violent TV shows have good points and bad points for children. For example; the animation superman fights badman tell children they don't have to bad things.Otherside;children should have to learn that all things don't finish by biolences.
YURI MUKAI 2 RFJ003 7/31/14 1:19 439 10.34 10b 7 9 9 5 8 8 7 5 3 8 the scientists who think a little bullying can be okay I think that a little bulling can be okay because we sometimes learn our fellings by crashing to other people. For example; kitties learn their power by fighting with their friends or brothers. Like this; We can't think about ourselves without other people. If we don't learn sad feeling about bulling; we can't understand other people's sad feeling. So; Ithink alittle bulling can be good for childlen.
RIKU TAKUSHIMA 1 RFJ001 4/24/14 1:54 222 9.99 09d 5 9 9 4 9 8 4 5 3 7 the scientists who think violent TV is bad I think that showing violent TV is bad for little children. Becouse when little children show that TV long time ; they will become to think `violences are forgived.`But when children glowed up;I think violent TVs are not so bad. Becouse they can think about violence and they rearise those TVs are fiction.
RIKU TAKUSHIMA 2 RFJ003 7/31/14 1:32 406 10.11 10a 6 9 8 4 7 6 7 5 3 8 the scientists who think a little bullying can be okay I think about littlr bullying is okay. Becaus children must learn about Don't hurt someone. So little bullying is needed children. But little bullying become hard bullying when adult don`t help. So adult have to look at children`s behave.
kanji ifuku 1 RFJ001 4/24/14 2:01 157 9.60 09c 8 7 7 7 8 7 6 4 4 5 the scientists who think violent TV is bad I think that violent TV shows are bad for children. Because;children can not between good and bad. If vieorent TV shows are bad contents; childrens might think good contents. Children imitate violent TV show;then children bad behavier. this reason; I think that TV shows are bad for children.

kanji ifuku 2 RFJ003 7/31/14 1:30 298 10.21 10a 7 1 6 2 4 8 7 3 3 7 the scientists who think bullying is always bad i think bulling is always bad. because; i think bulling gives bad effect for children. if children allow a little bulling ; children would input bad behavier. plus; children isnot understand good and dad. because; we should say that bulling is bad for children. therefore; i think bulling is always bad.
IKUE TANIGUCHI 1 RFJ001 4/24/14 1:59 161 9.88 09d 7 8 4 4 6 7 7 5 3 7 both groups Violent TV shows make children scarery; becase they think that when we are violented. But they will be able to learn to kind heart; because they know that violent makes other people sad. So;I think that both groups are right.
IKUE TANIGUCHI 2 RFJ003 7/31/14 1:27 334 10.18 10a 6 9 9 3 8 7 8 6 4 8 the scientists who think bullying is always bad I think that bulling is always bad;because bulling always attack somebady. Bulling is not thing to allow. If my family bull somebady;then I will be shock. Also;if my family is bulled somebady; then I will be anxious and angury. When I think about my family;we know bulling is always bad.
RYOTARO ICHIKI 1 RFJ003 4/24/14 1:52 207 10.20 10a 7 9 9 5 9 8 5 7 4 8 the scientists who think a little bullying can be okay I choice the scientists who think bullying is always bad is right. because bullying may never disapear. bullying is bad. but children will grow up in fight to bullying.
RYOTARO ICHIKI 2 RFJ001 7/31/14 1:32 298 10.26 10b 8 9 9 6 9 7 7 7 4 8 the scientists who think violent TV is bad I think violent TV shows are bad for children. because children can`t judge right or bad. parents have to teach there children ;"violence is not good". and violent TV is for adult; not for children.
SHUNSUKE SAKAE 1 RFJ003 4/24/14 1:51 174 10.59 10c 6 9 6 5 8 7 8 6 4 7 the scientists who think bullying is always bad I agree former scientists. Becouse bullying dosen`t have merit. I have two ideas . First;bullying has possibility to kill for people. I know bullying have students suiside in japan. It`s very sad for japanese people. Scond;bulling rob happy. Becouse I`m bullyed by my classmate. I thought I wnat to suicide ;life is very cheap. Bullying has this sad and bad problem By two ideas ;I agree former scientists
SHUNSUKE SAKAE 2 RFJ001 7/31/14 1:31 245 10.51 10c 6 9 9 4 8 6 5 5 3 8 the scientists who think violent TV is bad I think that violent TV is bad. Becouse it give bad effect for children.Today; bully is society probrem in Japan.It is two kinds.One is body attack;another is mental attack.body attack is violence.I think that violent TV help to increase to bully.By their ideas ;I disagree violent TV
RINA KATAHIRA 1 RFJ003 4/23/14 23:57 460 10.58 10c 8 9 9 7 8 9 9 7 6 9 the scientists who think a little bullying can be okay Bullying can sometimes be okay as long as it is not so serious. It is important for those who are bullied to try to solve the problem by themselves at first step and if it become difficult to do that; they should ask their teachers or parents for help. It could be possible that we can learn how to make good relationship or deal with such problems from that. What we could learn from such things would be very important to live in the society especially when they start working.
RINA KATAHIRA 2 RFJ001 7/31/14 0:26 485 10.30 10b 7 9 9 6 8 8 8 8 5 8 the scientists who think some violent TV is okay I think some violent TV is okay for children because it is possible that such TV teaches them moral that violence is bad. For instance; most bad man is violent and they aren't loved by people in such TV. Thus children learn that most violence is bad. And they come to think that they will be hated if they become violent.
Sakiko Takemoto 1 RFJ001 4/23/14 23:33 412 10.13 10a 7 9 9 6 7 9 9 7 5 9 the scientists who think violent TV is bad I think violent TV shows are bad for children because when they watch violent TV shows; they want to try to do same things that they watch in TV shows. Children have a lot of curiosity. They will be interested in violent things if they watch violent TV shows. So; I do not agree to show children violent TV shows.
Sakiko Takemoto 2 RFJ003 7/30/14 23:42 557 10.36 10b 7 9 9 6 8 7 7 7 5 8 the

scientists who think bullying is always bad I agree with the scientists who think bullyuing is always bad. If I am bullied; I wii not able to go to school because I cannot study confortably in the situation. Also; my friend said that she could not go to school when she was bullied in junior high school. She could not feel cpnfortable in her classroom; so she could not go. When children start to be bullied; they often cannot go toschool. This leads childrend's low scores. It is not fair that children cannot study because of bullying; so I think bullying is always bad.
NARUMI OKAZAKI 1 RFJ003 4/23/14 23:42 459 10.45 10b 8 9 9 7 9 9 6 8 7 9 the scientists who think a little bullying can be okay I think a little bullying can be a good experience for young children because they learn how to deal with people in the society from that. Some famouse actors or musicians nowadays make confessions about their bullied experience. They often said their bullied experience made them much stronger.
NARUMI OKAZAKI 2 RFJ001 7/30/14 23:15 771 10.50 10c 8 9 9 7 9 9 9 9 7 9 the scientists who think violent TV is bad I think "the scientists who think violent TV is bad" are right. Children are very sensitive and they tend to easily get affected by what they saw or what they heard. For example; some of killers mentioned that they have got their violent idias from their favorite TV shows. Also; I remember that there were some boys in my kindergarden who played acting a super hero and their enemies from a TV show; which is sometimes very violent. As long as children have a sense of good or bad; they should stay away from violent TV shows because they may misrecognize that viloence is cool.
IKUMI HONDA 1 RFJ003 4/23/14 23:37 755 10.47 10b 8 9 8 7 9 8 8 7 6 8 the scientists who think bullying is always bad I sincerely agree with the scientists who think bullying is always bad. Bullying hurts people's feeling. In fact; in my country; many students kill themselves because of that. Also; bullying is bad even for bulliers. They do not know social rules or they are lack of the skills of understanding other people's feelings. If people allow them to bully someone; they may have any illegal troubles in their future. To protect both bullies and poor children who are bullied; people must never allow anyone to bully someone else.
NANAYO OKAMOTO 1 RFJ001 5/28/14 15:55 181 9.97 09d 9 9 9 6 7 9 9 8 7 9 neither group no answer
NANAYO OKAMOTO 2 RFJ003 8/5/14 2:34 380 10.35 10b 6 8 9 4 7 7 7 5 4 8 both groups If there is no bullying; no one need to be felt sad. This is the best situation. But I think people can learn many things from bullying.
RYO MIYATA 1 RFJ003 5/2/14 3:15 163 10.25 10b 8 9 9 5 6 8 8 6 5 7 the scientists who think bullying is always bad Bullying is bad behavior because bullied people become unhappy for mental.Bullied people can't fight back to bullying people because bullied people isn't strong.Bullying people have considerable to bullied people.
RYO MIYATA 2 RFJ001 8/5/14 2:53 174 10.23 10a 8 9 9 5 8 6 6 5 3 8 both groups Violence is bad because the people who take violence is very sad and unhappy.But violence is permited to the people who do bad things.The police use violence when arrest the people who commit the crime.
YAMATO KOGA 1 RFJ001 5/28/14 16:01 113 10.13 10a 9 7 5 4 6 8 8 6 5 7 neither group no answer
YAMATO KOGA 2 RFJ003 8/5/14 2:25 176 9.99 09d 5 5 7 3 7 7 8 4 5 7 the scientists who think bullying is always bad I think bullying is bad thing.If someone bully my friend;it gradually go bad more.so bulling must not do.
MIKI HAYASHI 1 RFJ001 5/30/14 8:41 107 9.66 09c 9 8 8 4 7 6 8 5 3 5 neither group no answer
MIKI HAYASHI 2 RFJ003 8/5/14 2:50 219 9.88 09d 7 8 4 4 6 7 7 5 3 7 the scientists who think bullying is always bad I think building gives a dream to children. For example; Tokyo-Sky-Tree is the best building. Many children likes big thing like large building. When I was young; I like big things too. I liked Fukuoka-Tower very much. I went there with my family many times. It is very interesting my memory.
seiichirou simoyama 1 RFJ003 5/2/14 3:06 151 9.96 09d 7 9 9 5 7 8 8 6 5 7 the

scientists who think bullying is always bad I think that bullying is always bad because bullying will connect bad case. For example; suicide and illness of mind. That's why; I think bad.
seiichirou simoyama 2 RFJ001 8/5/14 2:29 161 10.09 10a 7 9 7 6 8 7 7 7 4 8 the scientists who think some violent TV is okay I think that violent TV is bad for children ; but I also think that violent TV is tell important things to children .
taisei shimizu 1 RFJ003 5/2/14 3:13 287 10.13 10a 8 9 9 4 7 8 8 6 5 7 the scientists who think bullying is always bad I think bullying is always bad;because bullying makes childlen sad. Even if you do not like someone;that is not reason why you bully him. I am sure that bullying can not make happiness. So we should not admit bullyihg.
taisei shimizu 2 RFJ001 8/5/14 2:52 379 10.02 10a 9 9 9 7 9 9 7 7 5 9 the scientists who think violent TV is bad I also think violent TV is bad for children;because children may think violence is okay by watching it.So;I think violent TV needs limit of old to watch it.
ANDO NORIFUMI 1 RFJ001 5/28/14 16:21 105 9.94 09d 9 9 9 6 8 8 8 7 3 7 neither group no answer
ANDO NORIFUMI 2 RFJ003 8/5/14 2:40 225 10.13 10a 3 8 8 3 6 8 7 4 3 8 the scientists who think a little bullying can be okay A little bullying is good effect for children.For example;They may make good frinds after bulling.And I can wider relationsip than not take bullying.
YUNA UMEDA 1 RFJ003 5/2/14 3:16 202 9.57 09c 9 9 7 5 9 8 7 7 4 8 the scientists who think bullying is always bad I think bullying is always bad;because children who were bullied are very sad. And bullying is not right. There are not people who feel good when they see bullying. So;other children who see this suituations should help bullying children or tell teachers and parents. And;adluts should help!
YUNA UMEDA 2 RFJ001 8/5/14 2:33 230 10.13 10a 8 9 9 6 9 8 7 6 4 8 the scientists who think violent TV is bad I think violent TV is bad. Because it is not good for children to learn about society. All people to live happily;peace is No.1.
MIZUKI YANO 1 RFJ003 5/2/14 3:13 252 10.50 10c 8 9 8 5 9 9 7 6 4 7 the scientists who think bullying is always bad I think bullying is always bad because if the children continue bullying others until adult;they don't live society with people.We live others; our family;friend;and community and so on.So we sometimes happen troubles to need help. If they don't understand helping others and only know bullying them;they couldn't consider others' heart; and others look them as "no-social human".We should teach children to be kind to others.If this idea spreed all over the world;we will have our better life style.
MIZUKI YANO 2 RFJ001 8/5/14 2:31 128 10.03 10a 8 9 9 5 8 8 8 4 4 8 neither group I think that some violent TV have a bad impact for children ;the other violent TV have a good impact for children.
AKARI MAEHATA 1 RFJ001 5/30/14 8:34 113 10.04 10a 9 9 9 4 7 7 8 5 4 8 neither group no answer
AKARI MAEHATA 2 RFJ003 8/5/14 2:33 201 10.26 10b 6 9 9 4 7 7 6 3 3 7 the scientists who think bullying is always bad We must not bully. Because bullying is killing other people's heart. I think it is very very hard and sad for people to bully. Bullying children will not be able to think other people's heart. Some of people kill ourselves because of bullyng. So bullyng is big problem in our life. So I think that the scientists who think bullyng is always bad is right.
RYUTARO HATABE 1 RFJ001 5/28/14 16:12 177 10.22 10a 9 9 8 5 8 8 8 6 6 8 neither group no answer
RYUTARO HATABE 2 RFJ003 8/5/14 2:57 282 10.42 10b 5 8 9 4 7 8 8 6 4 8 the scientists who think bullying is always bad I think bullying is always bad for children. bullying has no good pointfor children.
TAIGA SHINKAI 1 RFJ001 5/28/14 16:06 83 9.92 09d 9 6 9 6 7 8 7 7 4 7 neither group no answer
TAIGA SHINKAI 2 RFJ003 8/5/14 2:17 263 10.09 10a 7 6 9 5 8 8 8 7 5 8 the

scientists who think bullying is always bad bullying must be bad for all children because bulling can make children cry;damaged and so on.maybe all people think so.few people think that bulling can be ok.i think they are crazy because they don't think of the children bulliedby others.I never follow bullying children.
KOICHIRO HOSOKAWA 1 RFJ003 5/2/14 3:13 259 10.01 10a 8 9 9 7 9 9 8 7 8 9 the scientists who think bullying is always bad Because bullying hurt many people not only children but also their parents. And bullying do not make anypeople happy. Forexample children who are damaged by bullying can not forget their experience forever.
KOICHIRO HOSOKAWA 2 RFJ001 8/5/14 2:38 265 10.21 10a 8 9 9 6 8 9 8 7 5 9 the scientists who think violent TV is bad I think violent TV shows don't give any profit for children. If children watch violent TV shows; they may not think to hurt someone is bad and hurt someone who is around them.
rikako yoshioka 1 RFJ001 5/28/14 16:24 107 9.61 09c 9 9 9 4 6 5 8 6 3 7 neither group no answer
rikako yoshioka 2 RFJ003 8/5/14 2:34 197 9.91 09d 6 9 9 3 7 7 7 4 2 8 the scientists who think bullying is always bad I think bullying is bad that is not only children but also our people .
Shinji Wada 1 RFJ003 5/2/14 3:12 177 10.34 10b 7 8 7 5 8 8 8 6 4 8 the scientists who think bullying is always bad I think memory in childhood is more important for children; because people are based in memory when they were young or children ; and they grow up with them . So I don't think a little bullying can be okey.
Shinji Wada 2 RFJ001 8/5/14 2:36 251 10.24 10a 7 9 9 5 8 8 7 6 4 9 the scientists who think violent TV is bad I think that Violent TV is give children nothing; Because CHildhood is very important times. They study anything aroud them and their heart made up of everything they studied .Violense give nothing; don't makes us happy; just hurt you deeply. So I disagree that thing.
YUUKI SIGETOMI 2 RFJ003 8/5/14 2:37 311 10.32 10b 7 7 9 5 8 8 8 6 4 9 the scientists who think bullying is always bad I think bullying is always bad. First;children should be friendly each other. If we have some friendly people;they would help us when we need helps. Second;bullying have possibility what children become mad adult. Children should learn good communication with other people. So;I think the sientists who think bullying is always bad are right.
HARUNA NAKAO 1 RFJ003 5/2/14 3:18 311 10.38 10b 7 9 9 5 8 9 8 6 5 8 the scientists who think bullying is always bad I think bullying is not good for all people. In the world;many kinds of people are living. A man have good points and bad points. A woman have good points and bad points;too. All people have bad point. We should be friendly each other. So;bullying is bad.
MIWA KOMORI 1 RFJ001 5/30/14 8:37 154 9.78 09d 9 9 9 6 9 9 8 8 5 8 neither group no answer
MIWA KOMORI 2 RFJ003 8/5/14 2:11 141 9.74 09c 7 9 9 5 8 6 6 5 2 8 the scientists who think bullying is always bad It is because;he must sad who was bullyed.
MAMI HIOKI 1 RFJ003 5/2/14 3:14 207 10.23 10a 7 9 9 5 7 5 5 6 4 8 the scientists who think bullying is always bad Because I think that bullying makes everyone bad feeling. Bullying hurt people who is bullied and bully. And I think that we should tell children that bad thing must not do! So if they are children ; bullying is bad thing.
MAMI HIOKI 2 RFJ001 8/5/14 2:47 298 10.39 10b 7 9 9 7 9 7 7 7 5 9 the scientists who think violent TV is bad I think that childlen should not watch violent TV shows.Because If young children watch violent TV shows;some people who intend contents of the TV shows may appere. This is bad thing society.
KOSUKE NAKAMURA 1 RFJ003 5/2/14 3:06 210 10.11 10a 8 9 9 5 7 6 8 6 4 7 both groups I think that both group are right. That is because bullying makes bad affection. But a little bullying can be okay too. That is because a little bullying makes people strong. It is very difficult to think about bullying. But it is very big probrem to think. So I want know more about bullying probrem.

KOSUKE NAKAMURA 2 RFJ001 8/5/14 2:28 275 10.09 10a 6 8 8 4 7 8 6 5 4 8 the scientists who think violent TV is bad I think that violent TV is bad. That is because violent TV teach bad thing to children. For example; children became bad human to hurt othe people. And children be have bad mind.
KEIKICHI NAKANISHI 2 RFJ003 8/5/14 2:45 259 10.11 10a 4 9 8 2 7 6 6 4 2 7 neither group I think little bullying can be okay. However; my little bullying doesn't have injured. This sentence doesn't have expression of little bullying. Therefore; I can't agree neither group.
AIRI NAKAMURA 1 RFJ003 5/2/14 3:18 234 10.29 10b 7 9 9 4 8 9 6 6 3 8 the scientists who think a little bullying can be okay I agree with the scientists who think a little bullying can be okey. Because a little bullying give us strengh breaking probrems. Flower that we didn't give it lot of water bring up better than it that we give lot of water . So people too. Right quantity is good but bullying may be bad thing .
AIRI NAKAMURA 2 RFJ001 8/5/14 2:38 264 9.95 09d 7 9 9 5 8 8 7 5 4 8 both groups I think that both idea are right. One ideas violent TV is bad for children is right.Because violent is bad thing and danger. Another ideas some violent TV is okay for children is right.Because in order to teach children to violent never doing.
HAYATO HIROO 1 RFJ001 5/28/14 16:28 129 9.75 09d 9 9 9 5 8 8 8 7 4 8 neither group no answer
HAYATO HIROO 2 RFJ003 8/5/14 2:41 261 10.06 10a 5 8 9 3 8 8 8 6 4 8 the scientists who think bullying is always bad Sometimes;Bullying can make children suicide. It musn't be happen. So;I think bullying is always bad.
RIO OKAMOTO 1 RFJ003 5/2/14 3:10 238 10.08 10a 9 9 8 5 8 8 8 6 5 7 the scientists who think bullying is always bad I think bullying is always bad. This is because I have been told not to bully other people for a long time. If I bully some people;no one feel happy. These days;every school challenges not to bully to enjoy school life for students.
RIO OKAMOTO 2 RFJ001 8/5/14 2:25 194 10.13 10a 8 9 9 3 8 8 6 4 3 8 the scientists who think violent TV is bad I think the scientists who think violent TV is bad. It is impact on children to watch violent TV.
RIINA UEHARA 2 RFJ001 8/5/14 2:30 237 9.92 09d 7 9 9 5 9 8 7 6 4 8 both groups I think that child who has never watched violent TV will be peason without information of violent. It's very dangerous because he can't deal with such people. I think people need some information of violent. But if we have too much information of it; we will be violent peason. So Ithink both groups are right.
YUKIKO INOUE 1 RFJ003 5/2/14 3:17 257 10.46 10b 8 9 9 5 7 6 8 8 4 8 the scientists who think a little bullying can be okay I agree that a little bulling can be okay; because the boy who is bullied is not always right;and he may have bad points. Of course;the boys who is bulling are most bad. But; he should think about what he did;and he should think about how to be recieved by everyone;or how to be good friends.So;evereyone will be happy!!
YUKIKO INOUE 2 RFJ001 8/5/14 2:37 185 9.88 09d 9 9 9 6 8 6 4 4 3 7 the scientists who think some violent TV is okay I think violent TV is okey;because violent TV is important for children to learn good thing. For example; Anpanman is help every one!
MAIKO ONIZUKA 1 RFJ001 5/30/14 8:30 124 10.09 10a 9 8 8 4 8 8 7 6 4 8 neither group no answer
MAIKO ONIZUKA 2 RFJ003 8/5/14 2:16 162 9.83 09d 7 8 8 3 7 7 7 6 2 8 the scientists who think bullying is always bad I think bullying is very bad!!
TAKUMI SODA 1 RFJ003 5/2/14 3:03 183 10.71 10c 6 9 8 6 6 9 8 6 7 8 neither group I can not judge bullying is bad or right; because it is delicate problem; For example; sometimes bullied people caused bullying. On the other hand;bullied people do not harm anyone; but they are bullied. Its judge is so sivior. Which the bullying is bad change by condition that time.

TAKUMI SODA 2 RFJ001 8/5/14 2:29 202 9.94 09d 6 8 8 4 7 8 5 4 4 7 the scientists who think some violent TV is okay I think it is not bad to show violent TV. But Children's palents shoud tell them that violent is bad action.
student1 1 RFJ003 4/23/14 23:45 365 10.09 10a 7 9 8 5 9 8 8 6 5 9 the scientists who think a little bullying can be okay I agree the idea that a littele bulling can be okay. Beacause a littele bulling can sometimes makes children strong and realize about the feeling of being bullied people. Actually; bulling is bad because children who are bullied get hurt. But I think a littele bulling needs for children to learn people's feeling.
student2 1 RFJ001 4/23/14 23:06 304 10.25 10b 8 9 9 8 9 8 7 7 6 8 the scientists who think some violent TV is okay It is okay for children to watch violent TV. When they watch such a TV show; they might be able to understand and learn how dangerlous the violent things are. Also; if it's famous Tv show and everybody watches the show; they should watch it; otherwise they cannot catch up when their friends are talking about it. That'd be boring. I mean I won't be agree with watching all of the violent TV. Some are okay to watch.
student3 1 RFJ003 4/24/14 0:14 530 10.60 10c 10 10 10 7 9 9 8 8 6 9 neither group I think bullying is always bad; so I don't agree with the scientists who think a little bullying can be okay. However; I think the existing of the idea that a little bullying can be okay is allowed because having an idea is always allowed; whatever it is. Therefore; I can't say which group is right or wrong.
student4 2 RFJ001 7/30/14 22:40 366 10.32 10b 8 9 9 5 8 8 7 6 4 8 the scientists who think some violent TV is okay I think some violent TV show are okay because children have to learn how dangerlous the violent is. if they go into the society in the future without knowing it; then they might fight with violent. That is not good. Also when you watch such a show; your stress will be released.
student5 2 RFJ003 7/30/14 23:41 643 10.40 10b 8 9 9 6 7 9 8 7 5 9 the scientists who think bullying is always bad I think that bullying is always bad for children; because bullying leads children to kill themselves. Recently; the number of children who kill themselves because of bullying has been increasing. I don't want them to kill themselves. Plus sometimes the children who are bullied kill the other children who bully them. That is so sad thing. So I think that bullying is always bad.

Appendix D: University questionnaire results

Translations for the 21 questions on the standard questionnaire:

1. Attendance
2. Deeper understanding
3. Course satisfaction
4. Difficulty/level
5. Pace
6. Course content
7. Teacher's enthusiasm
8. Visual aids
9. Coursebook
10. Teacher's voicing
11. Pre-read syllabus
12. Read syllabus during semester
13. Content-syllabus match
14. Empowerment
15. Own dilligence
16. Study outside of class
17. n/a
18. n/a
19. n/a
20. n/a
21. n/a

		回答数				
No.	質問	1	2	3	4	5
1	(1)出席	0	0	0	26	0
2	(2)理解の深まり	0	0	1	13	12
3	(3)授業の満足	0	0	2	12	12
4	(4)レベル	0	1	13	10	2
5	(5)進度	0	0	24	1	1
6	(6)授業内容	0	0	2	10	14
7	(7)教員の熱意	0	0	0	6	20
8	(8)視覚情報	0	0	3	10	13
9	(9)教科書	0	2	3	9	12
10	(10)話し方	0	0	0	7	19
11	(11)受講前ｼﾗﾊﾞｽ	4	22	0	0	0
12	(12)受講中ｼﾗﾊﾞｽ	1	25	0	0	0
13	(13)ｼﾗﾊﾞｽと一致	0	0	0	0	2
14	(14)力の修得	0	0	0	1	2
15	(15)授業態度	0	0	1	9	16
16	(16)授業外学習	0	0	1	10	15
17	(17)学科1	0	0	0	0	0
18	(18)学科2	0	26	0	0	0
19	(19)学年	23	1	0	0	0
20	(20)追加質問1	0	0	0	0	0
21	(21)追加質問2	0	0	0	0	0

Appendix E. Survey Monkey questionnaire

Survey Monkey online questionnaire to all participating students.

1. What is your first name? ローマ字で「名」*(例: TARO)*

**2. What is your last name?* ローマ字で「姓」*(例: YAMADA)*

**3. Your full name in Kanji*
漢字等でフルネーム *(*例：山田　アレックス）

**4. Are you male or female?*

Male　男
Female　女

**5. Student Number:*

**6. Which class are you in?*

火Tue 1st period
火 Tue 2nd period
火 Tue 3rd period
火 Tue 4th period
木 Thu 2nd period
木 Thu 3rd period
木 Thu 4th period
木 Thu 5th period

**7. email address (smartphone/cellphone)* 「確実」にパソコンからのメールが受信できるスマホ、携帯のメルアド。（自分の「拒否設定」要確認*!!)*

**8. email address (home computer or university email address)* 「確実」にパソコンからのメールが受信できる自分の大学メルアド又は普段のパソコンのメルアド。（自分の「拒否設定」要確認*!!)*

**9. How old are you?*

17
18
19
20
21
22
23
24
25
26
26+

**10. How long have you spent in an English speaking country?*
英語圏（留学・観光等）総合滞在期間

ZERO
1-4 weeks
1-3 months
3-6 months
6-12 months
1-2 years
2-3 years
4-5 years
6 years or more

11. If you did spend time in an English speaking country 海外滞在経験者のみ
how much time did you spend in English? 使った英語の量

ZERO
less than half
about half
most of the time

100% English
12. If you have studied English/Eikaiwa outside of regular school, when did you? 学校の授業以外で英語／英会話の経験　*(*複数選択可*)*

幼児の頃、英塾CRAM SCHOOL before elementary school age
幼児の頃、英会話 English Conversation School before elementary school age
小学生の頃、英塾 during elementary school
小学生の頃、英会話 English Conversation School during elementary school
中学生の頃、英塾 during junior high school
中学生の頃、英会話 English Conversation School during junior high school
高校生の頃、英塾 during high school
高校生の頃、英会話 English Conversation School during high school
現在、英塾 now
現在、英会話 English Conversation School now
13. Have you ever taught English? 塾や英会話で英語を教えた経験　*(*複数選択可*)*

塾　Cram School
英会話 Conversation School
個人レッスン Private Lesson
14. If yes, then how many years did you teach in total?（英語／英会話を教えたことのある人のみ）

Less than 6 months
Less than 1 year
1-2 years
2-3 years
more than 3 years
15. 全員*: When did you enjoy English?* 英語が好きだったと言える頃　*(*複数選択可*)*

Before Elementary school
Elementary School 1-3
Elementary School 4-6
Junior High School 1st year
Junior High School 2nd year
Junior High School 3rd year
High School 1st year
High School 2nd year
High School 3rd year
University 1st year
University 2nd year
University 3rd year
University 4th year
16. 英語が好きだったと言える頃と嫌いだったと言える頃があれば、その変化の理由を詳しく（英語でも日本語でも*ok!)*
If you flip-flopped between liking and disliking English, please explain in detail (English or Japanese is ok)

**17. Reading* 自己評価 *(*本音で*!)*　★が多い程得意。

★
★★
★★★
★★★★
★★★★★
**18. Writing* 自己評価 *(*本音で*!)*　★が多い程得意。

★
★★
★★★
★★★★
★★★★★
19. Listening 自己評価 *(*本音で*!)*　★が多い程得意。

★
★★
★★★

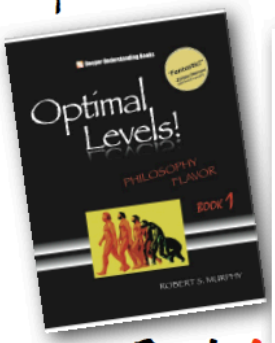
★★★★
★★★★★
*20. Speaking 自己評価 (本音で!)　★が多い程得意。
★
★★
★★★
★★★★
★★★★★
*21. Grammar 自己評価 (本音で!)　★が多い程得意。
★
★★
★★★
★★★★
★★★★★
*22. Vocabulary 自己評価 (本音で!)　★が多い程得意。
★
★★
★★★
★★★★
★★★★★
*23. Spelling 自己評価 (本音で!)　★が多い程得意。
★
★★
★★★
★★★★
★★★★★
*24. Presentations プレゼンテーション 自己評価 (本音で!)　★が多い程得意。
★
★★
★★★
★★★★
★★★★★
*25. Overall Fluency　総合的な流暢さ 自己評価 (本音で!)　★が多い程得意。
★
★★
★★★
★★★★
★★★★★
*26. What is your major/department at this university? 学部、専攻
*27. Do you use English regularly outside of the classroom? If so, when/where? 教室外で英語を使う人:どこで／いつ英語を使っていますか？　(複数選択可)
NEVER
I sometimes speak English with my family
I often speak English with my family
I sometimes speak English with my teachers (outside of class)
I often speak English with my teachers (outside of class)
I sometimes speak English with my foreign friends
I often speak English with my foreign friends
I sometimes speak English with my Japanese friends
I often speak English with my Japanese friends
I use English at ESS, or other English club activities
Facebook, sometimes
Facebook, often
Twitter, sometimes
Twitter, often
LINE, sometimes
LINE, often
Online chat (like Skype, etc), sometimes

Online chat (like Skype, etc), often
Other (please specify)
*28. English in your future! How likely do you think you will use English with your future family?　将来の自分と英語：将来の自分の家庭で英語を使う「可能性」（頻度ではなく、可能性の高さを考える）
★
★★
★★★
★★★★
★★★★★
*29. English in your future! How likely do you think you will use English with your future friends?　将来の自分と英語：将来の自分の友達と英語を使う「可能性」（頻度ではなく、可能性の高さを考える）
★
★★
★★★
★★★★
★★★★★
*30. English in your future! How likely do you think you will use English with your future work/ business?　将来の自分と英語：将来の自分の仕事（ビジネス）で英語を使う「可能性」（頻度ではなく、可能性の高さを考える）
★
★★
★★★
★★★★
★★★★★
*31. English in your future! How likely do you think you will use English on the internet?　将来の自分と英語：将来、インターネットで英語を使う「可能性」（頻度ではなく、可能性の高さを考える）
★
★★
★★★
★★★★
★★★★★
*32. English in your future! How likely do you think you will use English as a 'hobby', or out of fun (music, movies, etc)?　将来の自分と英語：英語を趣味で活かす可能性(音楽や映画等)。頻度ではなく、可能性の高さを考える。
★
★★
★★★
★★★★
★★★★★
33. How many times have you taken Mr. Murphy's English class before? マーフィ先生のクラスはこれで何回目？
今年が初めて　this is my first time
前に一度　I took his class once before
前に二度　I took his class twice before
前に三度　I took his class three times before
前に四度　I took his class four times before
Other (please specify)
*34. Which textbook are you using? テキスト名

New Philosophy Flavor (Blue/Black book)
Fun Flavor Book 2
Philosophy Flavor Book 2
NEW Fun Flavor BOOK 2
New Fun Flavor (Red book)
Optimal Levels Business Flavor (Black book)

Appendix F. OL1 Sample chapter

Optimal Levels! Philosophy Flavor

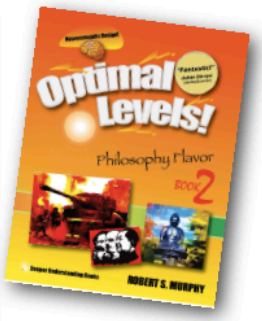


Modules	Themes	Optional
Module 1	Love	The Blossom
Module 2	Justice	Guilty
Module 3	Humanity	Evolution
Module 4	Economy	Manhattan
Module 5	Art	The Jazz Bar
Module 6	Freedom	On the Wings...

Book 1

Title Optimal Levels!: Philosophy Flavor Book 1
Level STUDENT CENTERED: High Beginner to Advanced. Very popular at universities!
ISBN 978-1451553284
Price ¥2,980 (Include TAX ¥3,129)
Philosophy in English class? Yes! Enjoy watching your students grow with great themes from Philosophy. These six modules are designed to foster a deep understanding in: Love, Justice, Humanity, Economy, Art, and Freedom. Teachers and student have great fun with this *Philosophy Flavor*. Very popular at universities! Works best when combined with *Optimal Levels! Project Book*.

Modules	Themes	Optional
Module 1	Happiness	The Couple
Module 2	Religion	What is Fair?
Module 3	War	Humanity
Module 4	Education	Getting Paid
Module 5	Democracy	Artistic Freedom
Module 6	Communism	In Chains?



Book 2

Title Optimal Levels!: Philosophy Flavor Book 2
Level Beginner STUDENT CENTERED: High Beginner to Advanced. Very popular at universities!
ISBN
Price ¥2,980 (Include TAX ¥3,129)
Students hungry for more Philosophy? Book 2 is the answer! Enjoy watching your students grow even further with these big themes from Philosophy. These six modules are designed to foster a deep understanding in: Happiness, Religion, War, Education, Democracy, and Communism. Heavy topics for your best students! Very popular at universities! Works best when combined with *Optimal Levels! Project Book*.

SELF-ASSESSMENT SHEET (PRE)

NAME: DATE:

	LEVEL (1-5 stars)	COMMENT:
Reading	☆ ☆ ☆ ☆ ☆	_____
Writing	☆ ☆ ☆ ☆ ☆	_____
Listening	☆ ☆ ☆ ☆ ☆	_____
Speaking	☆ ☆ ☆ ☆ ☆	_____
Grammar	☆ ☆ ☆ ☆ ☆	_____
Vocabulary	☆ ☆ ☆ ☆ ☆	_____
Spelling	☆ ☆ ☆ ☆ ☆	_____
Presentations	☆ ☆ ☆ ☆ ☆	_____
Overall Fluency	☆ ☆ ☆ ☆ ☆	_____

Module 1

Love



Section 1. Introduction

Look at the picture above. Ask and answer "who, what, when, where, why, how" questions.

Your questions	Your answers	Partner's answers
Q1. What do you see?		
Q2. _____		
Q3. _____		
Q4. _____		

Feelings/emotions from the picture:

_ _ _ _ _ _ _ _ _ _

Performance of Understanding Time!

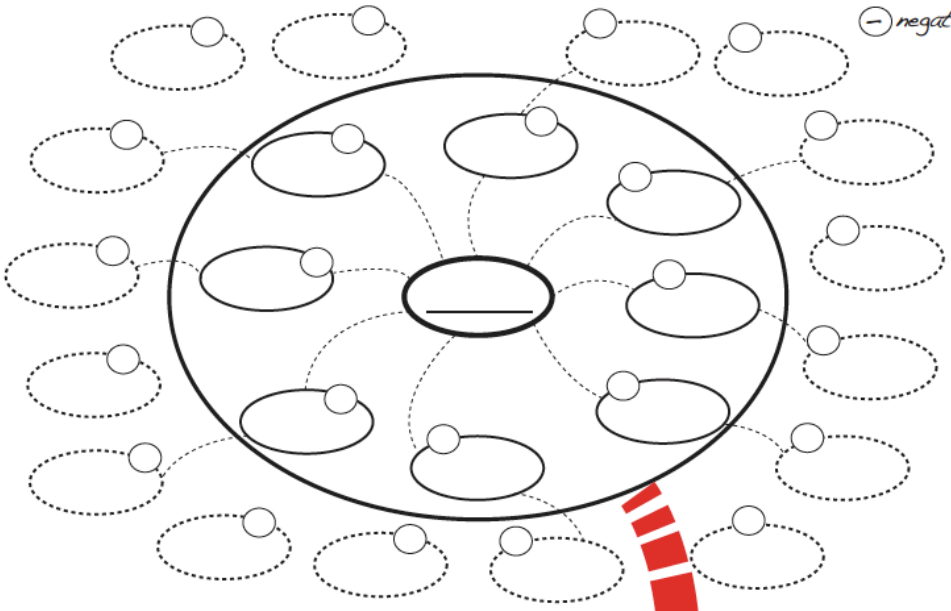
Discussion:

"This picture makes me feel..."

Section 2. Word map to expression!

Choose your own Root Word from the picture in Section 1 and create a word map.

⊕ positive
⊖ negative



Write three sentences from your word map ideas!
Draw pictures!

Sentence A

Picture A

Sentence B

Picture B

Sentence C

Picture C

Self-assessment of sentences: low 1 2 3 4 5 high

Section 3. Expansion



Work with a partner. Write your partner's sentences below. Discuss.

Partner's Sentence A

Partner's Sentence B

Partner's Sentence C

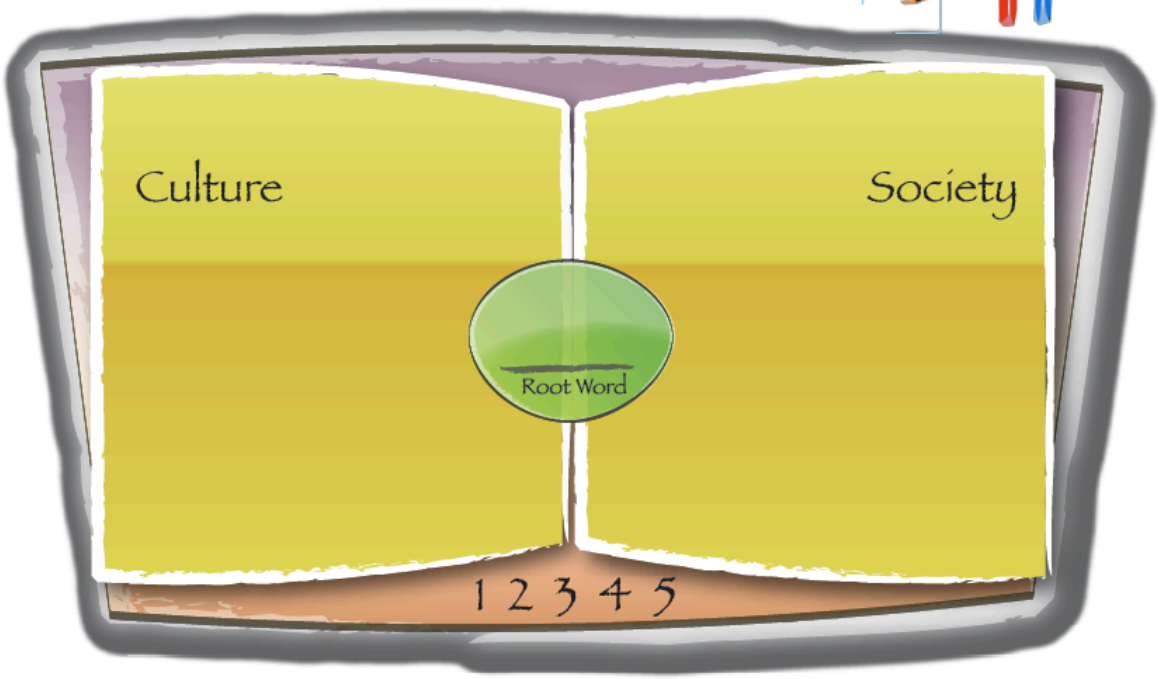
low high

Give your partner a score! 1 2 3 4 5

Performance of Understanding Time!

Discuss how the Root Word connects to culture and society. Write in the chart below.
Receive a Deep Understanding Score from your partner.





Section 4. Lexical Analysis!



Find coordinates, collocations, synonyms, and hyponyms of the Root Word. Write them on the cards.

A Coordinates with Root Word

salt - pepper - ketchup - mustard

B Collocations of Root Word

big house, birds fly, spicy curry

Root Word

C Synonyms of Root Word

hot-spicy, lost-missing, tall-high

D Hyponyms of Root Word

COLOR (red, blue), MONEY (yen, euro)

low high

Self-assessment: 1 2 3 4 5

Section 5. Lexical Performance Time!



Write sentences using the connections you made in Section 4.

A Sentence using "coordinates" of Root Word

My score for this sentence: 1 2 3 4 5

B Sentence using "collocation" of Root Word

My score for this sentence: 1 2 3 4 5

C Sentence using "synonym" of Root Word

My score for this sentence: 1 2 3 4 5

D Sentence using "hyponym" of Root Word

My score for this sentence: 1 2 3 4 5



Exchange your book with your partner.
Discuss your sentences. Make corrections if necessary.
Give yourself a final score for each sentence.

Section 6. High Performance!

Write a short story using your ideas from this module.



Title:

Date:

Feelings/Emotions discussed:

Socio-culture discussed:

Deep Understanding Scores:

	shallow				deep
Partner	1	2	3	4	5
Teacher	1	2	3	4	5

Fill in the box above.
Discuss. Receive scores and advice from your partner and teacher!

Section 7. Final Performance of Understanding!

Make your story even better!

Title

Date

Receive scores from your partner and teacher:

shallow deep shallow deep

1 2 3 4 5 1 2 3 4 5

Partner Teacher

Was your partner helpful?
Score your partner's help level: [1 2 3 4 5]

OPTIONAL

THINK AGAIN

Think about the unit you just finished. Answer these questions.

1. What did you enjoy about this unit?

2. How can you connect what you learned to the real world?

3. What have you become interested in because of this unit?

4. Ideas for improving your skills:

low high
Self-assessment: 1 2 3 4 5

Tutorial for Book 1 (it's all about building skills!)

Teachers, adjust the following instructions to match your students' levels.

FREE ONLINE TUTORIALS too!
DeeperUnderstandingBooks.com

Module 1
Love



Section 1. Introduction

Look at the picture above. Ask and answer "who, what, when, where, why, how" questions.

Your questions	Your answers	Partner's answers
Q1. What do you see?		
Q2. ____		
Q3. ____		
Q4. ____		

Feelings/emotions from the picture: _ _ _ _ _ _ _ _ _ _

Performance of Understanding Time!

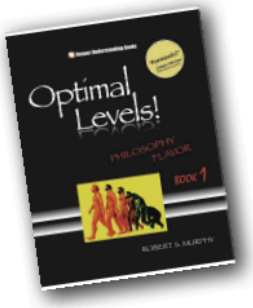
Discussion:
"This picture makes me feel..."

Section 1. Introduction

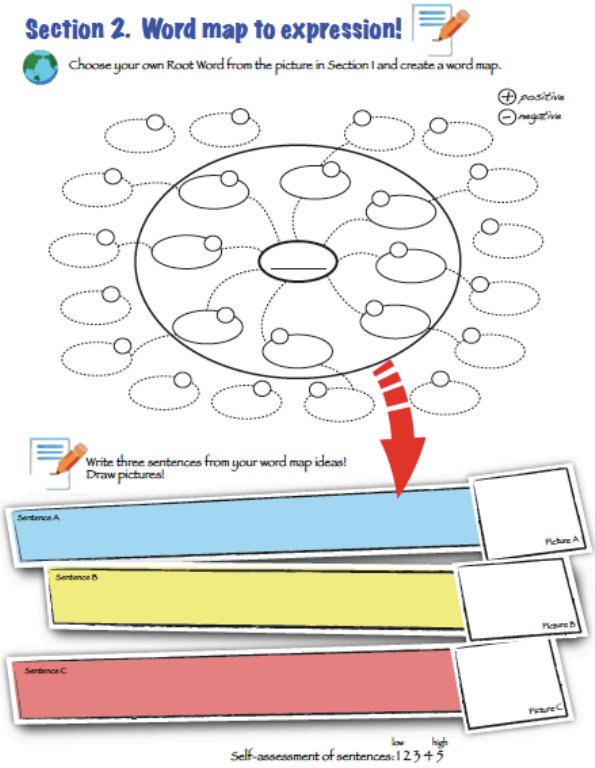
Look carefully at the provided picture. This picture represents the theme for this entire unit. What questions can you come up with? Think and then write down three more *"who, what, when, where, why, and how"* questions. After you have written the questions in the boxes, think of answers to your own questions and write them down too. Find a partner. Ask your questions to your partner and then write your partner's answers in the boxes next to your own answers. *How different were your answers?*

Feelings/emotions from the picture: Write down one or two feelings/emotions that you get from the picture.

Performance of Understanding Time! Find a partner. Discuss how the picture makes you feel and why. Compare answers with your partner. *How different were your answers?* Discuss with your teacher!



from Philosophy Flavor Book 1



Section 2. Word map to expression!

Choose one word (or one phrase) as the *Root Word* for the word map. The Root Word can be anything that you think matches this module's theme. Write the Root Word in the middle of the word map. Use your imagination and connect as many first generation links to other words as possible. Decide if the connections are positive or negative. Write a *plus* (+) or a *minus* (-) mark in the small circles. Next, use your imagination and create second generation connections. Write a *plus* (+) or a *minus* (-) mark in the small circles. There are eight 'floating connections'. Use these as you wish. You can make them first, second, or even third generation connections.

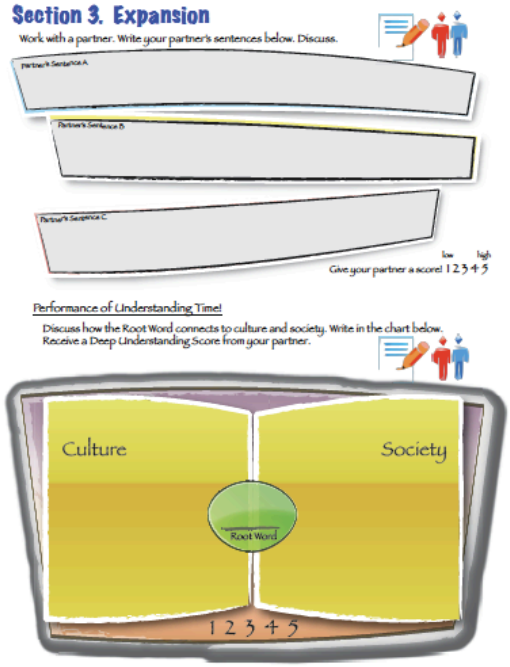
Write three sentences... Look carefully at your completed word map. You can probably think of many sentences. Write down three good sentences: Sentence A, Sentence B, and Sentence C. Next, draw one picture about each sentence next to each sentence.

Self-assessment of sentences: Think about your sentences. Choose what you want to assess about them. (Are they well connected to the theme? Do your sentences make sense? Are they grammatically correct?) *Did you do well?* Give yourself a score! [Scoring: 1 is low and 5 is high.]

Section 3. Expansion

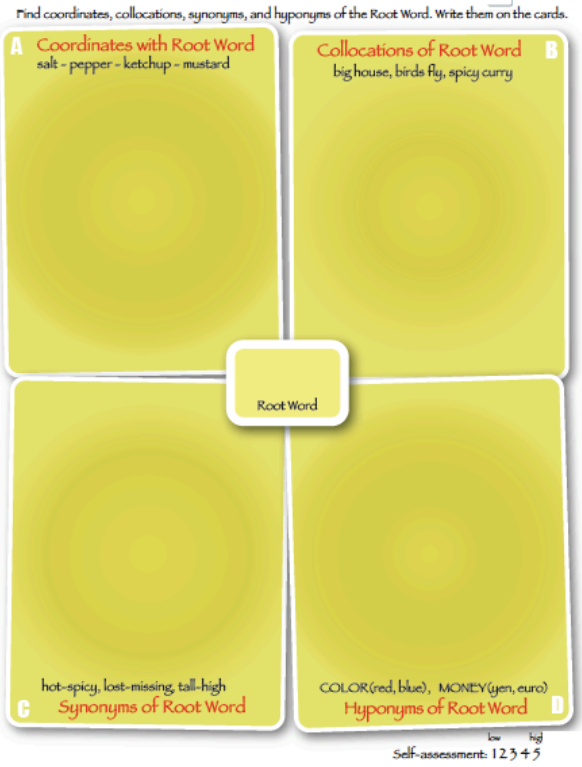
Find a partner. Ask your partner to dictate his/her three sentences to you. Write them down in the boxes. Think about your partner's sentences. Choose what you want to assess about them. *(Are they well connected to the theme? Do the sentences make sense? Are they grammatically correct?)* Give your partner a score! [Scoring: 1 is low and 5 is high.]

Performance of Understanding Time! Find a partner. Discuss your Root Words and how they connect to *culture (old culture and new/pop culture)*. Write your ideas down. Next, discuss how your Root Words connect to *society (any society)*. Write your ideas down. When you are finished, look at your partner's book. *How deep are your partner's ideas?* Give your partner a Deep Understanding Score (and receive a score from your partner). Discuss your ideas with your teacher.



Section 4. Lexical Analysis!

Find coordinates, collocations, synonyms, and hyponyms of the Root Word. Write them on the cards.



low high
Self-assessment: 1 2 3 4 5

Section 5. Lexical Performance Time!

Card A: Coordinates Write down a sentence using *coordinates* for the Root Word. A good example is: "I like to put *salt* and *pepper* on steak, but a hot dog needs *ketchup* and *mustard*!"

Card B: Collocations Write down a sentence using *collocations* for the Root Word. A good example is "Today I live in a *small* house, but some day I want to live in a *big* house!"

Card C: Synonyms Write down a sentence using *synonyms* for the Root Word. A good example is "I thought the Indian curry would be very *hot*, but it was not as *spicy* as I had expected!"

Card D: Hyponyms Write down a sentence using *hyponyms* for the Root Word. A good example is "I bought a new set of cutlery - *forks, knives, and spoons!*"

Exchange your book... Find a partner. Exchange your book with your partner. Discuss your sentences. Ask questions. Are there any mistakes? Correct the mistakes that you find. Return your partner's book.

Self-assessment: Choose what you want to assess. How well did you do? Give yourself a self-assessment for each sentence (1 to 5).

Section 4. Lexical Analysis

Write your Root Word in the Root Word box in the center of the page.

Card A: Coordinates Write down *coordinates* for the Root Word. Coordinates are words that are in the same general group, on the same level, and have a similar function. A good example is: "salt - pepper - ketchup - mustard".

Card B: Collocations Write down *collocations* for the Root Word. Collocations are words that fit next to the Root Word. Good example are: "*big* house", "*red* house", "*expensive* house" and "house *guest*".

Card C: Synonyms Write down *synonyms* for the Root Word. Synonyms are words that have the same or a similar meaning. Good examples are: "hot - *spicy*", "hot - *boiling*", and "hot - *sexy*".

Card D: Hyponyms Write down *hyponyms* for the Root Word. Hyponyms are words that belong to a set of words descriptive of, but are "*below*", the root word. Good examples are "Color *(red, blue, green)*", "Money *(dollars, yen, euro)*", and "Computer *(CPU, RAM, HDD)*" (If you cannot think of hyponyms, you can use *hypernyms* - words that are "above" the Root Word. *Color* is a hypernym of red.)

Self-assessment: Choose what you want to assess. How well did you do? Give yourself a self-assessment (1 to 5).

Section 5. Lexical Performance Time!

Write sentences using the connections you made in Section 4.



Exchange your book with your partner.
Discuss your sentences. Make corrections if necessary.
Give yourself a final score for each sentence.



Section 6. High Performance!

Look though the unit. Look at what you have done. Imagine a story using your ideas from this unit. Think of a title. Write the title down. Write the date down. Write your story down!

Feelings/Emotions discussed: Read your story to yourself. Think about the feelings/ emotions that are in the story. Write them in the box.

Socio-culture discussed: Read your story to yourself. *How does it connect to society and culture*? Write your a few words in the box.

Deep Understanding Scores: Exchange books with a partner. Read your partner's story. Think about your partner's story. Give and receive *Deep Understanding Scores* to each other. Show your story to your teacher. Receive a score from your teacher, too!



Section 7. Final Performance of Understanding!

Think about the scores that you received from your partner and from your teacher. Think about how your can improve your story. Write down a newer title. Write down the date. Write down your improved story!

Deep Understanding Scores: Exchange books with a partner. Read your partner's new story. Think about your partner's story. Give and receive new *Deep Understanding Scores* to each other. Show your new story to your teacher. Receive a score from your teacher too!

Was your partner helpful? Think about your partner's advice. How helpful was your partner? Give your partner a score (in your own book -you don't have to show your partner this score.)

Appendix G. New Optimal Levels (OL2) Revised version, Sample chapter

There are known *PDF-Word conversion errors* in Appendix F that do not exist in the published and bound version of coursebook.

(The OL2 book sample begins on the next page)

ROBERT S. MURPHY

NEW PHILOSOPHY FLAVOR

BOOK 2

Optimal Levels!

Deeper
Understanding
Books



About this series

Welcome to the NEW **OPTIMAL LEVELS!** series. This series is probably unlike any series of textbooks that you have encountered in the past. These textbooks have been designed to maximize student thinking and foster the construction of cognitive skills through the usage of the English language.

"Deeper Understanding" is really about the ability to solve puzzles and problems in the real world. Rather than *presenting* to students what must be learned, this series proposes motivating themes and scaffolded tasks that are designed to build skills systematically and dynamically, from the bottom-up. By doing the theme-based student-centered tasks, students learn and understand language usage by creating the *skills* necessary to negotiate meaning and build upon what they already know.

Enjoy exploring the creation of these dynamic skills that will lead to the *deeper understanding* of English and beyond!

Robert S. Murphy, series author
Deeper Understanding Books

NEW OPTIMAL LEVELS! PHILOSOPHY FLAVOR, BOOK 2

Modules	Themes	Grammar Focus
Module 1	People of the Past!	Past
Module 2	People of the Future!	Future
Module 3	People of Today!	Present



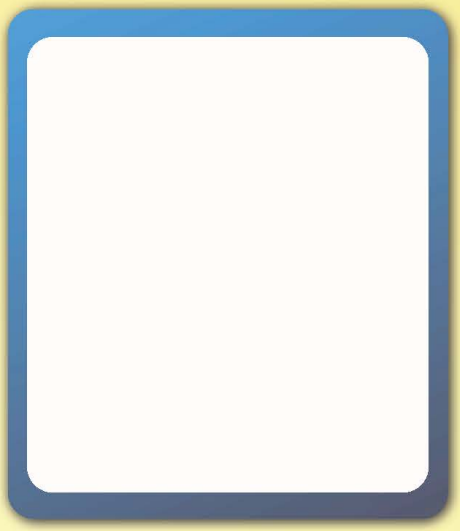
SELF-ASSESSMENT SHEET (PRE)

NAME: DATE:

	LEVEL (1-5 stars)	COMMENT:
Reading	☆ ☆ ☆ ☆ ☆	_____
Writing	☆ ☆ ☆ ☆ ☆	_____
Listening	☆ ☆ ☆ ☆ ☆	_____
Speaking	☆ ☆ ☆ ☆ ☆	_____
Grammar	☆ ☆ ☆ ☆ ☆	_____
Vocabulary	☆ ☆ ☆ ☆ ☆	_____
Spelling	☆ ☆ ☆ ☆ ☆	_____
Presentations	☆ ☆ ☆ ☆ ☆	_____
Overall Fluency	☆ ☆ ☆ ☆ ☆	_____

About me!

Name:



Module 1

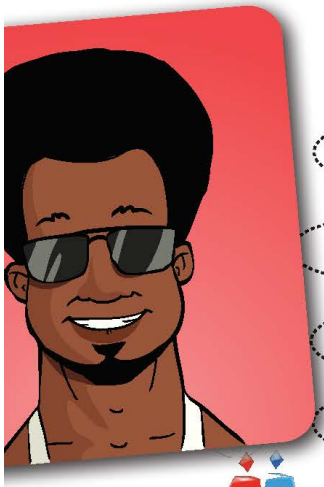
People of the Past!

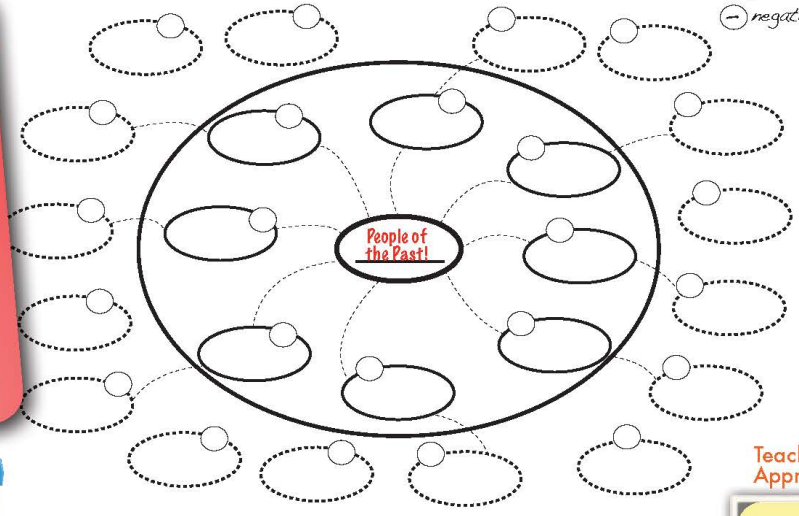


Cycle 1. Production: Word Map

Create a word map. Connect words/phrases you would need to discuss this topic.
ヒント：特にディスカションで使いそうな単語・熟語をどんどん繋げよう！

⊕ positive
⊖ negative





Teacher's Approval

許可マーク

Cycle 2. Exchange: Partner's advice

Exchange books with a partner. Think of better words and give advice.
ヒント：特にディスカションに必要そうな単語を考えてレベルアップさせよう！

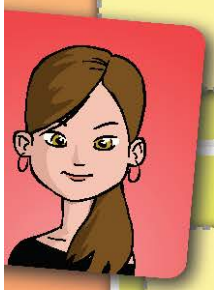
Cycle 3a. Production: Word categories

Group work. Use your word maps to create four word lists on this topic.
ヒント：グループワークで助け合い、最強のワードリストを作ろう！



A **Cultural Connections**
文化的なつながり

Feelings & Emotions
感情のつながり



People of the Past!
Root Word

C **Connections to Society**
社会的なつながり

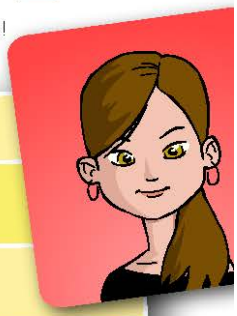
Phrases & Idioms D
フレーズ・熟語



Partner's Approval
許可マーク

Cycle 3b. Production: Context

Copy the words from 3A. Create good sentences as "Grammar Templates".
ヒント：文法テンプレートの作成。3aの単語から「基本」となる文章を作ろう！



Cultural Connections 文化的なつながり

Word/phrase	Sentence with context!

Feelings & Emotions 感情のつながり

Word/phrase	Sentence with context!



Partner's Approval
許可マーク

Cycle 3b. Production: Context

Copy the words from 3A. Create good sentences as "Grammar Templates".
ヒント：文法テンプレートの作成。3aの単語から「基本」となる文章を作ろう！

Connections to Society 社会的なつながり

Word/phrase	Sentence with context!



Phrases & Idioms フレーズ・熟語

Word/phrase	Sentence with context!

Partner's Approval
許可マーク

Cycle 4a. Exchange: Lexical ideas from the teacher

View teachers lists. Copy them here! Think and compare with your list.
先生のリストを写して、自分のリストとの違いを探ろう！

A Cultural Connections
文化的なつながり

on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO

B Feelings
感情のつながり

on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/N

People of the Past
Root Word



C Connections to society
社会的なつながり

on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO

D Phrases & Idioms
フレーズ・熟語

on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO
on my list YES/NO | on my list YES/NO

Partner's Approval
許可マーク

Cycle 4b. Exchange: Grammar Ideas from the teacher

View teachers "Grammar Template" examples. Copy them here!
Think and compare with your list.
先生のリストを写して、自分のリストとの違いを探ろう！

People of the Past!

Word/phrase	notes
Word/phrase	notes
Word/phrase	notes
Word/phrase	notes
Word/phrase	notes
Word/phrase	notes



Partner's Approval
許可マーク

Cycle 5a. Great ideas for the Presentation!

Get into your presentation group. Discuss and write down presentation content ideas!
ヒント：グループでプレゼンのコンテンツや流れをおおまかに決めよう！

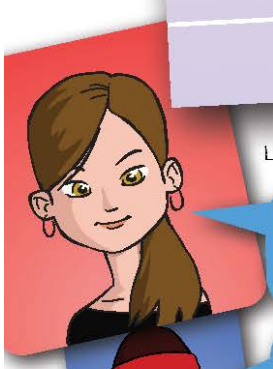
People of the Past!



Cycle 5b. Production - Group Master List!

Create a Master List with your group! Mix your ideas with your teacher's ideas.
ヒント：２つのリストを比べて、グループで話し合い、最強のリストを作ろう！

Phrases	Culture	Feelings	Society



List three of your best "Grammar Template" sentences for this topic. 文法の例文となるsentences!

A



B

C

Show this group list to the class with your teacher's help (on screen).

Partner's Approval
許可マーク

Cycle 6. Group-owned: Presentation!



In groups decide presentation content. Practice outside of class, too!
ヒント：プレゼンの内容をよく話し合って、楽しく教室外で何度も練習をしよう！

Slide 1

Theme: introduction

Word/phrase list

Slide 2

Theme:

Word/phrase list

Cycle 6. Group-owned: Presentation!



In groups decide presentation content. Practice outside of class, too!
ヒント：プレゼンの内容をよく話し合って、楽しく教室外で何度も練習をしよう！

Slide 3

Theme:

Word/phrase list

Slide 4

Theme:

Word/phrase list

Cycle 6. Group-owned: Presentation!



In groups decide presentation content. Practice outside of class, too!
ヒント：プレゼンの内容をよく話し合って、楽しく教室外で何度も練習をしよう！

Slide 5

Theme:

Word/phrase list

Slide 6

Theme: Conclusion!

Word/phrase list

Cycle 7. Presentation Feedback

Listen to the group presentations. Assess performance. Score fairly! Add your learning points from them.
ヒント：他のグループのプレゼンの良い/悪い所を細かくチェック！学べる所もあるはず→Learning Point

	Vocabulary 単語レベル	Fluency 流暢レベル	Grammar 文法レベル	Captivating 魅了的内容	Logical Flow 論理的流れ
Group A	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Learning point					
Group B	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Learning point					
Group C	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Learning point					
Group D	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Learning point					
Group E	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Learning point					
Group F	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Learning point					

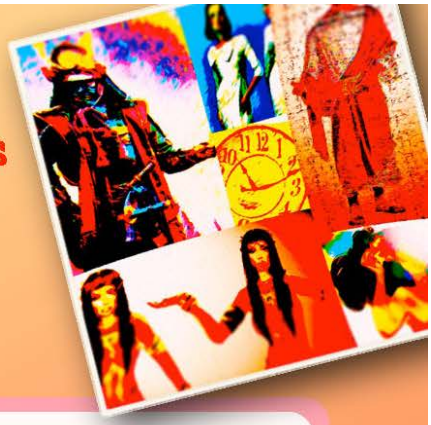


People of the Past!



Cycle 7b. Presentation: Feedback notes

Use this area to write down notes about the presentations!



Today's best performance!

Group

Cycle 8. Group work: Design a test! (Page 1)

People of the Past!



In your groups, design a test! Consider your Master Lists, your Grammar Boxes, and the Learning Points!

Name_____ Student Number_____ Date_____ Class_____

MODULE 1. TEST TITLE: _____



Section 1:

Section 2:

Cycle 8. Group work: Design a test! (Page 2)



ヒント： 今までの学習内容を中心に、自分を高めるためのテストをデザイン！

Name_____ Student Number_____ Date_____ Class_____

MODULE 1. TEST TITLE: _____

People of the Past!

Section 3.

Section 4.



Show this group test to the class with your teacher's help (on screen). The best test becomes the official test!

ヒント：選ばれたテストが実際にクラステストとして使われる！自分のテストが選ばれたら、絶対にトク！

Cycle 9. Choice: Best Test Design Award!





Use these charts to choose the best test for the class! 審査と受賞！

	Word choice	Difficulty	Design
Group A	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group B	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group C	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group D	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group E	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5
Group F	1..2..3..4..5	1..2..3..4..5	1..2..3..4..5

	Comments about the test:
Group A	
Group B	
Group C	
Group D	
Group E	
Group F	

Cycle 10. Testing, Reflection, and New Plans!

PART A: Take the Master Test that your class has chosen as the best test for you.
ヒント：みんなで選んだテストを受けよう。

People of the Past!

My test score!

Score

PART B: Think about your presentations and the test. Write down advice for raising your skills further.
ヒント：プレゼンとテストの結果をみて、さらなる実力アップのためにできることを考えよう！

Learning point

Learning point



Learning point

Learning point

Learning point

PART C: Discuss your ideas in class with your teacher. Write down new ideas. Make a learning plan.
ヒント：新たなラーニング プランを作成！次のレベルへステップアップ！

New Plan for Study	What do I need to do?

Teacher's Approval
許可マーク

OPTIONAL

THINK AGAIN



Think about the module you just finished. Interview a partner. Write down your combined answers.
ヒント：パートナーと自分の考えをここでまとめよう！

1. What did you enjoy about this unit?
   Partner:

   Me:

2. How can you connect what you learned to the real world?
   Partner:

   Me:

3. What have you become interested in because of this unit?
   Partner:

   Me:

4. List any new ideas for improving your skills:
   Partner:

   Me:

low high
Self-assessment: 1 2 3 4 5

SELF-ASSESSMENT SHEET (POST)

NAME: DATE:

	LEVEL (1-5 stars)	COMMENT:
Reading	☆ ☆ ☆ ☆ ☆	_____
Writing	☆ ☆ ☆ ☆ ☆	_____
Listening	☆ ☆ ☆ ☆ ☆	_____
Speaking	☆ ☆ ☆ ☆ ☆	_____
Grammar	☆ ☆ ☆ ☆ ☆	_____
Vocabulary	☆ ☆ ☆ ☆ ☆	_____
Spelling	☆ ☆ ☆ ☆ ☆	_____
Presentations	☆ ☆ ☆ ☆ ☆	_____
Overall Fluency	☆ ☆ ☆ ☆ ☆	_____

Positive Thoughts Journal Day 15

I am happy because:



80