Investigating the Efficacy of Online Text Reconstruction Exercises for Facilitating the Use of Metadiscourse Markers in First-Year Japanese University Students’ Argumentative Writing

Thesis

How to cite:

For guidance on citations see FAQs.

© 2022 David Allen Gann

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.21954/ou.ro.000156a2

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.
Investigating the Efficacy of Online Text Reconstruction Exercises for Facilitating the Use of Metadiscourse Markers in First-Year Japanese University Students’ Argumentative Writing

David A. Gann
MA (Literature), MA (TESOL)
Doctor of Education (Ed.D)
The Open University

Education

November 25th, 2022
Abstract
This thesis examines pedagogical merits of online text reconstruction exercises (OTREs) and examines their efficacy in teaching argumentative writing (AW) and students’ experience of learning through them. It looks at a specific type of OTRE called WebSequiturs and that application’s unique pedagogical affordances. Key research questions ask to what extent OTREs can influence EFL university students to use select metadiscourse markers (MDMs) in written argument; and to what extent those exercises can guide students to use those MDMs appropriately. A third research question asks about students’ perceptions of their experiences completing OTREs. This study was conducted at Tokyo University of Science and involved nine students in a first-year undergraduate compulsory English course.

This study strives for an interpretive understanding of participants’ experiences. Hence, I pursue this topic through action research. I also take a mixed-methods approach, using both quantitative and qualitative data. Quantitative data shows that following OTRE sessions, frequency of the use of selected MDMs in participants’ writing increased. The variety of MDMs having similar meanings likewise increased. There was also an increase in the range of communicative functions across which MDMs were used. Moreover, the unique uses of select MDMs also increased. Finally, there was an increase in appropriacy of use.

Qualitative data showed that during the OTREs participants sometimes reached metalinguistic levels of awareness, as evidenced by their utterances. These utterances were frequently followed by increased levels of appropriacy of MDM use in participants’ writing. During semi-structured interviews, participants responded, evincing the view that OTREs had played an important role in their developing use of MDMs in their AW.

The findings of this study suggest that OTREs are helpful in influencing English learners to become autonomous users of selected MDMs and that they can also guide learners to use MDMs appropriately. The findings also show that participants understood the purpose of the OTREs and valued the exercises for their linguistic instruction and for their collaborative qualities. In my conclusion, I recommend that universities implement programs using similar OTREs and train their teachers in how to maximise their efficacy in teaching AW.
Acknowledgements

The first person I would like to dedicate this work to is my wife, Rika, who suffered through these years with me often absent from home, and often absent when home.

I would like to thank my supervisors Dr. Prithvi Shrestha and Dr. Nathaniel Owen, both of whom I was extremely lucky to draw as my advisors.

Thank you, also to my colleague Dr. Stephen Jennings who preceded me in this Ed.D programme and whose view ahead was invaluable.

Last but not least, thank you to the students who agreed to participate in this study. They not only happily participated; they also added to the quality of this thesis with their insightful comments.
Table of Contents

Abstract .................................................................................................................................................................................................................................................. 3
Acknowledgements ......................................................................................................................................................................................................................... 4
List of Tables ......................................................................................................................................................................................................................... 11
List of Figures ................................................................................................................................................................................................................... 13
List of Abbreviations .................................................................................................................................................................................................................. 17

Chapter 1: Introduction .................................................................................................................................................................................................................. 18
1.1. Formative motivations ........................................................................................................................................................................................................ 18
1.2. What is L2 writing? What processes do L2 writers engage in? ........................................................................................................................................................................ 19
1.3. Challenges to Japanese students developing AW skills ........................................................................................................................................................................ 21
1.4. The challenges and problems of adopting CLT in Japan ........................................................................................................................................................................ 21
1.5. Institutional responses ........................................................................................................................................................................................................... 22
1.5.1. Metadiscourse markers and AW .................................................................................................................................................................................. 22
1.5.2. Taxonomy of MDMs ................................................................................................................................................................................................................ 23
1.5.3. A bridge between instruction and performance ........................................................................................................................................................................ 23
1.6. Specific context of this study ............................................................................................................................................................................................................ 24
1.7. Aims and research Questions .................................................................................................................................................................................................. 25
1.8. Overview of this thesis ........................................................................................................................................................................................................... 26

Chapter 2: Literature review: Metadiscourse markers and argumentative writing .................................................................................................. 28
2.1. Introduction ......................................................................................................................................................................................................................... 28
2.2. Critical views of CT instruction .................................................................................................................................................................................................. 28
2.3. CT skills instruction and AW ................................................................................................................................................................................................... 29
2.3.1. MDMS and AW ............................................................................................................................................................................................................. 29
2.3.2. Problems in Japanese use of English metadiscourse markers in AW ......................................................................................................................... 30
2.3.3. Discourse community MDM use ............................................................................................................................................................................ 33
2.3.4. Text-centred and user-centred use of MDMs related to AW ................................................................................................................................................................ 34
2.3.5. Critical issues in MDM taxonomy .............................................................................................................................................................................. 35
2.4. Methodological considerations and influential studies ......................................................................................................................................................................... 36
2.5. Challenges to learning AW in Japan ................................................................................................................................................................................. 37
2.6. CT and AW pedagogy at national, institutional and classroom levels .................................................................................................................. 39
2.7. Situating students’ writing within EFL .................................................................................................................................................................................. 40
2.7.1. Computer mediated communication ............................................................................................................................................................................. 41
2.7.2. Forums and chat rooms ...................................................................................................................................................................................................... 42
2.7.3. CMC’s impact on form ........................................................................................................................................................................................................... 44
2.7.4. Motivation in CMCs ............................................................................................................................................................................................................ 45
2.7.5. Imitation, internalisation, interlanguage and social constructivism ......................................................... 46
2.7.6. Learning transfer between ACMC and SCMC environments ........................................................................... 48
2.8. Podcasting and instruction .......................................................................................................................... 48
2.8.1. Text reconstruction exercises ............................................................................................................... 49
2.8.2. Background of text reconstruction ........................................................................................................ 49
2.8.3. Text-reconstruction as a trend within CALL .............................................................................................. 50
2.9. Semi-structured interviews ......................................................................................................................... 51
2.10. Limits of available research ..................................................................................................................... 52
2.11. Research questions ..................................................................................................................................... 54
2.12. Summary ..................................................................................................................................................... 54

Chapter 3: Methodology: design and methods .................................................................................................. 56

3.1. Introduction ................................................................................................................................................... 56
3.2. Theoretical position: An interpretivist approach .......................................................................................... 56
3.3. Research rationale ....................................................................................................................................... 58
3.3.1. Introduction ............................................................................................................................................. 58
3.3.2. Action research ...................................................................................................................................... 58
3.3.3. Advantages of mixed-methods in action research .................................................................................... 58
3.4. Research context ......................................................................................................................................... 60
3.4.1. Participants ............................................................................................................................................. 60
3.4.2. MDM categories used in this study and Hyland’s taxonomy ................................................................. 63
3.4.3. Interactive and interactional MDMs ....................................................................................................... 67
3.5. Research design ........................................................................................................................................... 69
3.6. Initial study .................................................................................................................................................. 71
3.7. Findings of initial study ............................................................................................................................... 72
3.7.1. Data sources and collection instruments ................................................................................................. 72
3.7.2. Data collection and research questions ................................................................................................ 74
3.7.3. Master MDM List .................................................................................................................................... 77
3.7.4. Podcasts and research considerations .................................................................................................. 79
3.7.5. WebSequiturs ......................................................................................................................................... 81
3.7.6. WebSequitur generation ......................................................................................................................... 82
    Using WebSequiturs as an innovative MDM pedagogical tool ................................................................. 87
3.7.7. Merits of collaborative TRE work .......................................................................................................... 88
3.8. Forum readings ........................................................................................................................................... 89
3.9. The Monty Hall Problem ........................................................................................................................... 92
3.10. Summary: Research tools that relate to RQs 1, 2 and 3 ........................................................................... 94
3.11. Methods .................................................................................................................................................... 95
3.11.1. Initial lessons and participant selection ............................................................................................... 95
3.11.2. Online Text Reconstruction Exercise sessions ..................................................................................... 96
Chapter 4: Data analysis

4.1. Quantitative analysis of MDM use .......................................................... 115
4.1.2. Forums as a data collection tool ......................................................... 115
4.1.3. Chat rooms as a data collection tool summary .................................. 116
4.1.4. Group Written Report as a data collection tool ................................. 116
4.1.5. In-class Rewrite from Memory assignment as a data collection ............ 117
4.1.6. Confirming MDM genre ...................................................................... 117
4.1.7. Types of quantitative data collected .................................................... 119
4.1.8. Categories of argument indicators ..................................................... 121
4.1.9. Calculation of MDM use ................................................................. 123
4.1.10. Charts displaying MDM use ............................................................ 126
4.2. Qualitative analysis ............................................................................. 127
4.2.1. Qualitative analysis of MDM use ...................................................... 127
4.2.2. Establishing a scale of appropriacy of MDM use ............................... 127
4.2.3. Independent assessment of appropriacy of MDM use ......................... 129
4.2.4. Thematic analysis ............................................................................ 131
4.2.5. Thematic analysis of transcripts of student-student talk during OTRE sessions .............................................................. 131
4.2.6. Semi-structured interviews .............................................................. 136
4.2.7. Themes presented in the interview ..................................................... 137
4.2.8. Establishing a scale of appropriacy of MDM use ............................... 138
Chapter 5: MDM (token) use by frequency

5.1. Introduction ................................................................. 141
5.2. Data Sources and frames .................................................. 141
5.2.1. Participants’ reported understanding of the purpose of the OTREs ........................................... 144
5.2.2. Forums I, II and V ......................................................... 145
5.2.3. Forums III and IV ......................................................... 147
5.2.4. Chat rooms I and II ....................................................... 148
5.2.5. Group Written Report and In-Class Rewrite ................................................... 150
5.3. MDM (token) use ............................................................... 151
5.3.2. Frame 1: Forums I, II and V ............................................ 152
5.3.3. Frame 2: Forums III and IV ............................................. 154
5.3.4. Chat rooms I and II ....................................................... 156
5.3.5. Written Report and Individual In-Class Rewrite ................................................... 158
5.3.6. Quantitative Summary .................................................... 160
5.4. Unique MDM (token) use ................................................... 161
5.4.1. MDM (token) use across communicative functions ................................................... 163
5.5. Explicitly connective premises and conclusion indicators ................................................... 165
5.5.1. Implicitly connective and explicitly connective MDMs (tokens) in Forums I, II and V .......... 166
5.5.2. Implicitly and explicitly connective premises and conclusion indicators in Forums III, IV .... 171
5.5.3. Implicitly and explicitly connective hypothetical premises and conclusion indicators in Forums III and IV ................................................... 176
5.5.4. Implicitly and explicitly connective premise and conclusion indicators in Chat Rooms I and II .......... 181
5.5.5. Implicitly and explicitly connective conditional statements and hypothetical premise indicators in Chat Rooms I and II ................................................... 183
5.5.6. Implicitly and explicitly connective premises and conclusion indicators in the GWR and ICR ........................... 185
5.5.7. Implicitly Conditional statements in the GWR and ICR ................................................... 189
5.6. Participants’ perception of the efficacy of OTREs ................................................... 191
5.6.1. Participants’ understanding of the OTREs purpose ................................................... 191
5.6.2. Findings of thematic analysis of OTRE transcripts ................................................... 192
  Interrater reliability of thematic analysis of OTRE session transcripts and interview transcripts ............. 193
  Discussion of procedure for beginning an exercise ................................................... 194
  Negotiation of answers ................................................... 194
  Disagreement ................................................... 195
  Correction of misunderstandings ................................................... 195
5.6.3. Prior knowledge and instructions in the use of MDMs (tokens) in AW ................................................... 196
5.6.4. The influence of grammatical cues vs argument indicators ................................................... 200
5.6.5. The influence of argument indicators ................................................... 203
Chapter 6: Appropriacy of MDM (token) use ........................................................................ 209

6.1 Introduction ............................................................................................................. 209

6.2. Independent assessment of appropriacy of token use ........................................... 209

6.2.1. Examples of appropriacy assessment .................................................................. 211

6.2.2. Efficacy of OTREs over printed handouts ......................................................... 220

6.2.3. Participants’ perception of technological and social constructivist advantages of OTREs ........................................................................................................ 223

6.2.4. Assessment of levels of appropriacy for conditional statements and hypothetical premises (tokens) ... 231

6.4.1. Participants’ High Point assessments in Forums III, IV, the GWR and the ICR ... 232

6.4.2. Summary of appropriacy levels in conditional statements and hypothetical statements ... 249

6.5. Participant interview responses ............................................................................. 249

6.5.1. Participants’ perceptions of the general effectiveness of OTREs ......................... 250

6.5.2. Efficacy of OTREs over printed handouts .......................................................... 252

6.5.3. Participants’ perception of technological and social constructivist advantages of OTREs ........................................................................................................ 254

6.6. Participant interview responses ............................................................................. 259

6.7. Discussion ............................................................................................................. 271

6.8. Summary .............................................................................................................. 274
Chapter 7: Conclusions, implications, limitations and future research .................................................. 276

7.1. Introduction ..................................................................................................................................... 276
7.2. Review of study aims .................................................................................................................... 276
7.3. Findings .......................................................................................................................................... 278

7.3.1. Research Question 1: OTREs and frequency of MDM (token) use in AW ............................ 280
7.3.2. Research Question 2: OTREs and appropriate use of MDMs (tokens) in AW ...................... 282
7.3.3. Research Question 3: Participant perceptions of OTREs ..................................................... 284

7.4. Implications for faculty and teachers, administration, and policy makers ................................. 285
7.4.1. Implications for faculty and teachers ...................................................................................... 285
7.4.2. Implications for administration .............................................................................................. 287
7.4.3. Implications for national policy makers ................................................................................... 288

7.5. Limitations and future iterations .................................................................................................. 288

References ........................................................................................................................................... 292

Appendices ........................................................................................................................................ 311
List of Tables

Table 2.1: Premises, conclusions and argument indicators
Table 2.2: Hypothetical premises, antecedents, consequents, and indicators
Table 2.3: Future predictive and present counterfactual conditional forms
Table 3.1: Participants’ past English study experience, etc.
Table 3.2: Sample of premise and conclusion related MDMs
Table 3.3: Four data sources
Table 3.4: Research tools as data sources
Table 3.5: RQs and data sources
Table 3.6: Interview questions
Table 3.7: Main themes of student-student talk during OTRE sessions
Table 3.8: Interview questions and RQs
Table 4.1: Number of posts for each participant during each forum
Table 4.2: Chat Rooms’ durations and number of turns for each participant
Table 4.3: Raters’ distribution of MDM List items as spoken, written or both
Table 4.4: Cohen’s Kappa for raters’ categorisation of MDMs as spoken or written
Table 4.5: Premise and Conclusion Indicators
Table 4.6: Conditional statements and hypothetical premises
Table 4.7: MDM Category titles and descriptions
Table 4.8: MDM Appropriacy Scale
Table 4.9: Scale for assessing the appropriacy of hypothetical premise constructions
Table 5.1: Frames and data sources
Table 5.2: Number of words by each participant as an overall measure of Forum engagement
Table 5.3: Number of words written by each participant in forums
Table 5.4: Number of words written by each participant in Chat Rooms I and II
Table 5.5: Number of words written by each participant in the GWR and ICR
Table 5.6: Average token count of nine participants
Table 5.7: Token use in Forums I, II and V, by P1-P9
Table 5.8: F I, P1-P9 token counts, frequency, total wordcount
Table 5.9: F II, P1-P9 token counts, frequency, total wordcount
Table 5.10: F V, P1-P9 token counts, frequency, total word count
Table 5.11: Average token use for nine participants, Frame 2
Table 5.12: Table 5.12: Token use in Forums III and IV, by P1-P9, Frame 2
Table 5.13: F III, P1-P9 token counts, frequency, total word count
Table 5.14: F IV, P1-P9 token counts, frequency, total word count
Table 5.15: Average token use for nine participants, Frame 2
Table 5.16: Token use in Chat Rooms I and II, by P1-P9, Frame 3
Table 5.17: CR I, P1-P9 token counts, frequency, total word count
Table 5.18: CR II, P1-P9 token counts, frequency, total word count
Table 5.19: Average token use for nine participants, Frame 2
Table 5.20: Token use in the GWP and ICR, by P1-P9, Frame 4
Table 5.21: GWR, P1-P9 token counts, frequency, total word count
Table 5.22: ICR, P1-P9 token counts, frequency, total word count
Table 5.23: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 1
Table 5.24: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 2
Table 5.25: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 2
Table 5.26: P6’s and P7’s token use for premises and conclusion indicators
Table 5.27: P6’s and P7’s token use for conditional statements and hypothetical premises
Table 5.28: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 3
Table 5.29: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 3
Table 5.30: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 4
Table 5.31: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 4
Table 6.1: Averages of nine participants’ appropriacy use assessment by output sources and raters
Table 6.2: Number of assessments for each level by raters 1, 2 and 3
Table 6.3: Utterance counts of participants by category
Table 6.4: Participants’ utterance counts grouped (categories condensed)
Table 6.5: Participants’ utterance count groupings converted to L, M and H, and compared with appropriacy of token use ratings
Table 7.1: Findings organised by RQs 1-3
List of Figures

Figure 2.1: ELT methodologies, Ikeda, 2012
Figure 3.1: Initial Study
Figure 3.2: Section I of the Master MDM List: premises and conclusions
Figure 3.3: Section XI of the Master MDM List: conditional statements and hypothetical premises
Figure 3.4: Critically Minded Podcast episodes on argument organisation
Figure 3.5: Critically Minded Podcast episodes on conditional statements and hypothetical premises
Figure 3.6: A WebSequitur being constructed in the Sequitur application
Figure 3.7: The WebSequitur from above in its completed online form
Figure 3.8: WebSequitur as it appears after completion
Figure 3.9: Break points, before and after argument indicators
Figure 3.10: A WebSequitur modelling conditional statements and hypothetical premises
Figure 3.11: A WebSequitur text reconstruction
Figure 3.12: Short narrative and discussion questions
Figure 3.13: First forum
Figure 3.14: Cups and Coins, Part One, handout
Figure 3.15: The class project
Figure 3.16: The second of the three OTREs focusing on conditional statements and hypothetical premises
Figure 3.17: A teacher’s post introducing the second forum
Figure 3.18: A teacher’s post introducing the third forum
Figure 3.19: An illustration from the second half of the Cups and Coins activity
Figure 3.20: A chat between three students
Figure 3.21: Task brief for Group Report
Figure 3.22: One participating triad’s final Group Report with the MDMs highlighted
Figure 4.1: Raw data for one participant’s MDM use
Figure 4.2: An example of an independent reader’s assessment
Figure 4.3: Transcript of student-student talk during an OTRE session
Figure 4.4: List of themes distributed to each rater
Figure 4.5: Tool for interrating OTRE transcripts
Figure 5.1: Average of nine participants’ unique MDM (token) uses
Figure 5.2: Average of nine participants token use and unique token use
Figure 5.3: Average for Unique tokens and Communicative Functions in Forums I, II and V
Figure 5.4: Average for Unique tokens and Communicative Functions in Forums III, IV
Figure 5.5: Average for Unique tokens and Communicative Functions in Chat Rooms I, II
Figure 5.6: Average for Unique tokens and Communicative Functions in GWR, ICR
Figure 5.7: Average of P1-P9s’ frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.8: P2’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.9: P3’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.10: P’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.11: P6’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.12: P9’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.13: Average of P1-P9s’ frequency of Implicitly and Explicitly Connective token use for Forums III and IV
Figure 5.14: P1’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV
Figure 5.15: P6’s frequency of token use in conditional statements and hypothetical premises in Forums III and IV
Figure 5.16: P2’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV
Figure 5.17: P4’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV
Figure 5.18: P5’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV
Figure 5.19: Average of P1-P9s’ frequency of token use in conditional statements and hypothetical premises in the Forums III and IV
Figure 5.20: P2’s frequency of token use in conditional statements and hypothetical premises in the Forums III and IV
Figure 5.21: P3’s frequency of token use in conditional statements and hypothetical premises in the Forums III and IV
Figure 5.22: P4’s frequency of token use in conditional statements and hypothetical premises in the Forums III and IV
Figure 5.23: Average of Participants I-9s’ frequency of Implicitly and Explicitly Connective token use for Chat Rooms I and II
Figure 5.24: Average of P1-P9s’ frequency of token use in conditional statements and hypothetical premises in Chat Rooms I and II
Figure 5.25: P4’s frequency of token use in conditional statements and hypothetical premises in the Chat Rooms I and II

Figure 5.26: P6’s frequency of Implicitly and Explicitly Connective token use for Chat Rooms I and II

Figure 5.27: P7’s frequency of Implicitly and Explicitly Connective token use for GWP and IC

Figure 5.28: P8’s frequency of Implicitly and Explicitly Connective token use for the GWP and ICR

Figure 5.29: Average of P1-P9s’ frequency of token use in conditional statement indicators and hypothetical premises in the GWP and ICR

Figure 5.30: P1’s frequency of token use in conditional statements and hypothetical premises in the GWP and ICR

Figure 5.75: Inter-rating tool used in thematic analysis of OTRE transcripts of participant talk

Figure 6.1: Cronbach’s Alpha for raters assessment of appropriacy of token use

Figure 6.2: Average of nine participants’ token use by levels of appropriacy

Figure 6.3: Three examples of participants’ written output assessed for appropriacy levels 1-4

Figure 6.4: Levels of appropriacy of token use for P1 in Forums I, II and V

Figure 6.5: Levels of appropriacy of token use for P2 in Forums I, II and V

Figure 6.6: Levels of appropriacy of token use for P8 in Forums I, II and V

Figure 6.7: Levels of appropriacy of token use for P4 in Forums I, II and V

Figure 6.8: Levels of appropriacy of token use for P6 in Forums I, II and V

Figure 6.9: Levels of appropriacy of token use for P2 in Forums III and IV

Figure 6.10: Level of appropriacy of token use for P6 in Forums III and IV

Figure 6.11: Level of appropriacy of token use for P5 in Forums III and IV

Figure 6.12: Level of appropriacy of token use for P1 in Chat Rooms I and II

Figure 6.13: Level of appropriacy of token use for P2 in Chat Rooms I and II

Figure 6.14: Level of appropriacy of token use for P6 in Chat Rooms I and II

Figure 6.15: Level of appropriacy of token use for P9 in Chat Rooms I and II

Figure 6.16: Level of appropriacy of token use for P1 in the GWR and the ICR

Figure 6.17: Level of appropriacy of token use for P3 in the GWR and the ICR

Figure 6.18: Level of appropriacy of token use for P2 in the GWR and the ICR

Figure 6.19: Level of appropriacy of token use for P7 in the GWR and the ICR

Figure 6.20: Average of assessment of appropriacy for nine participants in Forums III and IV and the GWR and the ICR

Figure 6.21: Average of assessment of appropriacy for P2 in Forums III and IV

Figure 6.22: Average of assessment of appropriacy for P1 in Forums III and IV
Figure 6.23: Average of assessment of appropriacy for P4 in Forums III and IV
Figure 6.24: Average of assessment of appropriacy for P3 in Forums III and IV
Figure 6.25: Average of assessment of appropriacy for P9 in Forums III and IV
Figure 6.26: Average of assessment of appropriacy for P4 in the GWR and the ICR
Figure 6.27: Average of High Point assessment of appropriacy for P6 in the GWR and the ICR
Figure 6.28: Average of High Point assessment of appropriacy for P1 in the GWR and the ICR
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACMC</td>
<td>Asynchronous Computer Mediated Communication</td>
</tr>
<tr>
<td>AW</td>
<td>Argumentative writing</td>
</tr>
<tr>
<td>CF</td>
<td>Communicative Function</td>
</tr>
<tr>
<td>CLT</td>
<td>Communicative Language Teaching CI: Conclusion Indicator</td>
</tr>
<tr>
<td>CMC</td>
<td>Computer Mediated Communication</td>
</tr>
<tr>
<td>CS</td>
<td>Conditional Statement CT: Critical Thinking</td>
</tr>
<tr>
<td>DM</td>
<td>Discourse Marker</td>
</tr>
<tr>
<td>EAP</td>
<td>English for Academic Purposes</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a Foreign Language ELF: English as Lingua Franca</td>
</tr>
<tr>
<td>ESP</td>
<td>English for Specific Purposes</td>
</tr>
<tr>
<td>GWR</td>
<td>Group Written Report</td>
</tr>
<tr>
<td>HP</td>
<td>Hypothetical Premise</td>
</tr>
<tr>
<td>IC</td>
<td>Implicitly Connective</td>
</tr>
<tr>
<td>ICR</td>
<td>In-Class Rewrite</td>
</tr>
<tr>
<td>LC</td>
<td>Explicitly connective MDM: Metadiscourse Marker</td>
</tr>
<tr>
<td>MEXT</td>
<td>Ministry of Education, Culture, Sports, Science and Technology</td>
</tr>
<tr>
<td>OTRE</td>
<td>Online Text Reconstruction Exercise PC: Premise and Conclusion</td>
</tr>
<tr>
<td>PI</td>
<td>Premise Indicator</td>
</tr>
<tr>
<td>SCMC</td>
<td>Synchronous Computer Mediated Communication</td>
</tr>
<tr>
<td>TBL</td>
<td>Task Based Learning</td>
</tr>
<tr>
<td>TESOL</td>
<td>Teachers of English to Speakers of Other Languages Trans: Translation</td>
</tr>
<tr>
<td>TUS</td>
<td>Tokyo University of Science</td>
</tr>
<tr>
<td>VLE</td>
<td>Virtual Learning Environment</td>
</tr>
<tr>
<td>ZPD</td>
<td>Zone of Proximal Development</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1. Formative motivations

This study comes as a result of a deep and long-time interest in critical thinking instruction and a sustained effort to understand the forms that critical thinking instruction might take in an English as a Foreign Language (EFL) context in a first-year university English course. This pursuit later led me to narrower research into the basic skills involved in argumentative writing and the research questions introduced in Chapter 2.

In 2008, I was teaching first and second-year university students in several suburban cities north of Tokyo. As a result of exposure to podcasts centred around critical thinking and scepticism, I became interested in creating similar content appropriate for an EFL learning context. Since as early as the late 1990s, critical thinking (CT) instruction has been considered an important part of essential reading skills such as skimming-and-scanning, comparison-and-contrast, expressing temporal sequence, cause and effect, reasons and conclusions, and critical awareness of academic discourse conventions (Pally, 1997). In the 2000s, Japan’s Ministry of Education began mandating educational reform, including critical thinking instruction (Stroupe, 2006).

CT skills are regularly categorised in tiers. The first of these is the ability to detect arguments and to identify premises and conclusions by recognizing premise indicators (e.g., because and since) and conclusion indicators (e.g., hence and therefore) referred to collectively as argument indicators (Fisher, 2001). In addition, CT skills instruction and argumentative writing skills instruction are closely connected, and one skill set can transfer to the other (Golpour, 2014; Dabaghi, Zabihi and Rezazadeh, 2012; Hashemi and Zabihi, 2012). I noted at that time, however, that among EFL textbooks which purport to present CT skills, almost none included the use of common argument indicators (Chen and Baker, 2010). Adverbial linking clauses, for example, because and therefore, were largely untreated, and this has continued to be the case as late as 2017 (Phoocharsoenil, 2017) often used as logical connectors (Hyland, 2005a). Instead, a more concise style of expressing opinions is modelled, in which the reader is expected to infer the presence and structure of arguments (Gann and Bufton, 2012a).

This approach skips the first tier of critical skills development. I found that students often were not able to differentiate between arguments and non-arguments, although the language used signposted premises and conclusions (Gann, 2013). It was for that reason that I began producing a podcast that would meet my needs, and potentially, the needs of other EFL
instructors. In 2010, production of *Critically Minded Podcast* (Gann and Bufton, n.d.) began. In short, the approach was to teach CT skills by covering the argument indicators used to express and employ them, rather than the issues-based approach I had formerly taken. Moreover, the instruction on these argument indicators was done, via podcast, outside of class. Many students initially responded by dutifully listening to the podcasts, which was evidenced by their printing and even marking up the scripts. However, after some time, comprehension quiz scores indicated that podcast listening had fallen off. The reason, I gathered, was that the podcast content was not perceived by the students as being meaningfully connected to what was happening during class meetings.

At that time, during my master’s degree studies in Teaching English to Speakers of Other Languages (TESOL) at University of Manchester, I began to approach the instruction of argument indicators through an EFL perspective. As a result of the educational technology component of that programme, I became interested in online text reconstruction exercises (OTREs) as a possible resource for teaching argumentative writing.

OTREs involve the reconstruction of short written passages performed on a computer. I collected phrases through various EFL resources, and when appropriate, the written output from students themselves. By these means, the content of Critically Minded Podcast was determined and during the next four years the OTREs were designed, arranged, and implemented.

Historically, OTREs have been used largely for vocabulary and grammar training (Čechovičová, 2010; Rodríguez, 2010). This study extends that utility, showing first, the influence OTREs can have in increasing student writers’ use of AW related phrases and secondly the guidance OTREs can provide in student writers’ appropriate use of those phrases. In the context of this study, those skills involve the use of those phrases to signal basic argumentative structure related to premises and conclusions. This thesis looks into OTREs efficacy in helping EFL learners to develop argumentative writing skills. It fills a gap in this field by drawing attention to a unique and useful computer application which seems to have not been employed by other educators to introduce metadiscourse markers in student AW. At its farthest reaching, it aims at providing support for my colleagues, modelling good practices for teaching students AW.

1.2. What is L2 writing? What processes do L2 writers engage in?
This study takes place in an EFL learning context, where second language (L2) writing is the writing of persons learning to write in a language other than their native tongue (Ellis, 1994).
Recognizing the fundamental difference between ESL (where the L2 is the dominant language) and EFL (where the L2 is a foreign language) is an important distinction and informs the teaching approaches used in either context (Manchón, R. M., 2009). Historically EFL emerged from the field of English as a Second Language (ESL) and a greater body of research exists in ESL (Reichelt, 2011). The term L2 may be applied to both ESL and EFL contexts (Storch, 2013; Cimasko and Reichelt, 2011). In this study, L2 is used to refer to the EFL context in which it was situated. Some of the relevant processes which L2 learners work through in their language learning are lexical retrieval (Schoonen, et. al, 2009), solving language-related problems concerning rhetorical and discourse issues, i.e., appropriacy of language use and backtracking (referring back to the previously written text segment to confirm one’s position), formulation (putting one’s ideas into words), planning (considering possible pragmatic decisions before writing); and these are often carried out in the learner’s L1.

Writing practices are also situated in culture. Like all argumentative writing, writing for EFL students also involves an effort to achieve coherence (Basturkmen and Randow, 2014; Hyland, 2009). L2 learners of argumentative writing must learn, not only the grammar and vocabulary of the L2, but also the culturally accepted structures and the phrases used to make those structures salient to readers (Hyland, 2005b). Each EFL learner must work from their own native culture’s ethos and preconceptions about rhetoric. In Japan, most high school students undergo an intensive training in preparation for entrance exams and as Rinnert and Kobayashi (2009) found, L1 writing training had a positive effect on students’ later L2 writing efforts with students who had received L1 writing training making more use of argumentative metadiscourse markers. Still, that training does advance a rhetorical argumentative structure that is inductive and specific-to-general, the opposite of what is generally taught in western schools and students must later learn to write in their L2 in a way that is contrary to their L1 training (Rinnert and Kobayashi, 2009). In the next section I briefly discuss these and other challenges facing Japanese students.

The students in a first-year English composition course, such as the one in this study, work toward reorienting their attitudes and preconceptions about writing, and about writing in English. Having been placed in university, they can now consider writing in English in terms outside the pressure of the university entrance examination (Seargeant, 2008). They are now challenged to think of English as a tool for communication of ideas, opinions, and in a university context, reasons for their opinions. They may do so within the context of writing
assignments, as with the Group Written Report and written communications, as with the use of forums and chats, all important parts of this study.

1.3. Challenges to Japanese students developing AW skills
One problem which may cause difficulty for Japanese students learning AW is the influence of Confucian philosophy (Stanford, 2012). Confucianism has much to do with how the educational system has become one in which emulation of the teacher (Stanford, 2012) and rote learning from exemplary texts has long been the dominant practice (Gray, 1998). Atkinson (1997) maintains that CT instruction may be incompatible with an Asian mindset and societal norms and argues for caution in its inclusion in curricula in Japanese universities. Among these societal norms is a reticence for engaging in argument. Confucianism prioritises social harmony and an avoidance of discord (Stanford, 2012). In terms of AW, Confucianism has played a strong role in the development of Japanese rhetoric and the ki-shoo-ten-ketsu style of writing which is inductive, placing a greater responsibility on the reader to infer meaning, and tending to suggest logical flow, rather than state it clearly (McKinley, 2013). For my part, I eschew essentialising students as thinkers dominated by Confucian principles (Cutrone, 2013). To this point, Stapleton (2002) found strong evidence that Japanese undergraduate students understood the importance of supporting their opinions with reasons. The extent of Confucianism’s present-day influence in students’ willingness to speak out, express opinions and even disagree with the teacher is questionable as recent research suggests that many university students do not fit into this generalised characterisation (Stapleton, 2002). However, its historical influence on the Japanese educational system cannot be discounted.

Another challenge is the culturally entrenched top-down yakudoku, i.e., teaching to the test (Bao, 2014; Kohro, 2009) approach. Yakudoku is the result of an educational system geared towards preparing students to pass university entrance examinations. It is characterised by the memorisation of grammatical rules at the expense of fostering of communicative competence (Seargeant, 2008). Due to various factors, such as the pressure on Japanese high school EFL teachers to prepare students for university exams, yakudoku seems no less prevalent in recent years (Thompson and Yanagita, 2015) than it did at the beginning of the century (Gorsuch, 2001).

1.4 The challenges and problems of adopting CLT in Japan
The response to these challenges has been to adopt the teaching practices of Communicative
Language Teaching (CLT). CLT, a movement which originated in the 1980s, prioritises communicative competence over structuralism and grammatical knowledge. It seeks to create genuine communication opportunities, even in the classroom, by drawing on the available resources in each given context (Hiep, 2007). As communicative competence is emphasised, the focus of CLT pedagogy is on the role of the learner (Savignon, 2007). While CLT has been lauded as an improvement over the formerly dominant grammar-translation method, it has been criticised as an import from ESL, imposed on the EFL context in Asia and in Japan, without sufficient consideration for how it would work with national and local societal norms (Lochland, 2013; McKinley, 2013; Ikeda, 2012). Adaptions to CLT have been made by researchers and practitioners working within EFL contexts. This is discussed in greater detail in Chapter 2.

1.5. Institutional responses
In the latter half of this decade, while the Ministry of Education set goals, methods for achieving those goals are unclear. Programs are being implemented to introduce CT skills, such as in high schools (Suzuki, 2015a). Universities must prepare to satisfy demands of the coming wave of students who will expect coursework to extend beyond what they have already learned. At Tokyo University of Science (TUS), where this study was conducted, many students still generally lack AW skills (Jennings, 2018). Although TUS is enacting institutional policy aimed at improving students’ CT and AW skills (see Appendices 1-5), no clear means of achieving this has been recommended or taken up. This study aims to help fill that gap.

1.5.1. Metadiscourse markers and AW
Metadiscourse markers (MDMs) are words or phrases which cohere the relationship between one segment of a text and preceding one (Hyland, 2005a; Fraser, 1999). They are reflexive (Mauranen, 2010), enabling a text to discuss itself, within single, short sentences, across long paragraphs or entire essays (Halliday, 2009). Coherence is an essential quality in AW (Bax et al., 2019; Fung and Carter, 2007; Intaraprawat and Steffensen, 1995).

Argument indicators are MDMs that allow writers to regulate logical information flow (Mauranen, 2010). They create cohesion and maintain coherence within systems of meaning (Halliday, 2009) by drawing together premises and conclusions in order to make the whole of an argument salient for the reader. In this study, I am focusing on first-tier critical thinking skills, the ability to identify premises and conclusions (Fisher, 2001), and to put them to use
in AW as *transitions* (Hyland, 2005a).

I use the term MDM to refer to (1) premise and conclusion indicators and (2) other discourse markers, which, for example, signify agreement or disagreement, express levels of conviction, or create focus in dialogue, and which also signal or indicate the presence of a premise or a conclusion.

### 1.5.2. Taxonomy of MDMs

How MDMs are grouped and classified has been an area of debate and several taxonomies exist. The choice of taxonomy informs how we understand text reflexivity. Kopple (1985) places premise and conclusion indicators in the category of *textual connectives* and in the subgroup of *text connectives*. Crismore et al. (1993) similarly group argument indicators as *textual markers* and the subgroup, *logical connectives*. Both view argument indicators as limited to text regulation and as distinguished from *interpersonal* metadiscourse.

Hyland (2005a) departs from both earlier models by dispensing with the dividing line between textual and interpersonal. Instead, Hyland views all metadiscourse as interpersonal. He categorises argument indicators as *interactive* and within the subcategory of *transitions*. Hyland’s dispensing with the line between textual and interpersonal becomes important later in Section 2.3.5, when I discuss how some other categories of metadiscourse can also function as argument indicators. It is for those reasons that I have made the critical decision to follow Hyland (2005a) regarding this issue.

### 1.5.3. A bridge between instruction and performance

OTREs are a potentially effective way to address issues concerning AW instruction (Warschauer and Healey, 1998; Thornbury, 1997) and since 2011, I have used OTREs to create a bridge between out-of-class podcasted instruction and in-class tasks requiring AW. In particular, *WebSequitur*, an online text reconstruction application made available by the same authors as the more well-known Hot Potatoes (accessible at [http://www.cict.co.uk/textoys/](http://www.cict.co.uk/textoys/)) includes features which Rodríguez (2010) used to teach paragraph composition and which this study found to be effective in promoting the skilled use of premise and conclusion indicators. Note, that early on in the study the distinction between Hot Potatoes and WebSequitur which is licensed under TexToys, whose site is promoted at the Hot Potatoes page, was not clear. For that reason participants sometimes referred to the OTREs as Hot Potatoes when

Because there is nothing, in principle, preventing a teacher from reproducing the
WebSequitur design on an alternate platform, such as Moodle, I have chosen to avoid the trade name WebSequiturs and for the most part will refer to these exercises as OTREs. I have developed a series of OTREs which provide students with the opportunity to practice what they have learned through explicit instruction, via podcast, about the use of MDMs in AW.

1.6. Specific context of this study
The students involved are first-year students at a high-ranking science university. Typically, such students, while high performing in maths and the sciences, are not highly proficient in English. However, given that they have been accepted into a university with somewhat high entrance examination demands, it may be supposed that, overall, they have the capacity to make great improvements. The syllabus for the course in which this study is being conducted states that students should learn to ‘read critically’, and later, to ‘think logically and critically’ (see Appendix 7).

This study took place in a first-year second term course, called ISE-A. ISE is an acronym for Integrated Skills in English. The -A following the ISE indicates that this is the first of three ISE courses, which follow Freshman English (FE-1) in the first semester with ISE-A and in the second year with ISE-B and -C. These four required courses are central to the TUS Noda Campus English programme.

In ISE-A students are expected to advance from commonplace spoken expressions in the first term Freshman English course (FE-1) to an emphasis on paragraph writing skills and the expression of reasoned opinions (see Appendix 8). They are also expected to advance from the construction of single sentences to the level of writing short paragraphs (see Appendix 7). Students are given the goals of learning to communicate ideas, thinking logically and critically, and becoming proficient at writing short paragraphs about general topics, for example those presented in the class textbook.

The writing tasks involved in this study are framed as students writing to their peers at their university campus, or in a professional or civic context, and the words and phrases that students are instructed to use included the polite written communication which a reader would likely infer was that of persons who have achieved an undergraduate level of higher education or higher or had undergone some similar professional training. These tasks are in accordance with the goals set out by the university mandate asking that instructors teach writing skills that include a ‘sense of internationalization / cross-cultural communication / knowledge of foreign languages / knowledge of other peoples / knowledge of foreign affairs’ (Tokyo University of Science, 2019b) (see Appendix 7). The writing in the first half of the term is interactive, as it
is done within online forums (see Section 3.11.3) and online chat rooms (see Section 3.11.8). That communication involves the discussion of issues presented in an EFL textbook and prepares students for ISE-B in which they are further tasked to participate in written discussions of topics at paragraph-length exchanges using the AW skills they have studied in the class (see Appendix 8). In the latter half of the term, while continuing those discussions, students also begin working in small groups, writing a short paper (see Section 3.11.9) for details), which also involves AW.

In the EFL context of this study, the teaching approach used falls within communicative language teaching (CLT). To this end, the second half of the term moves towards a weak form of task-supported instruction and greater autonomy from students (Butler, 2011) (see Section 2.7). Within the specific context in which this study was carried out, students wrote in three online locations. Throughout the term, students wrote argumentatively in forums (asynchronously) and in chat rooms (synchronously). At the end of the term, they wrote a group written report; and during the last class meeting, they rewrote that report from memory.

1.7. Aims and research Questions

There is a body of research on the use of MDMs in AW, some in L2 writing contexts. These studies all have some commonalities with my own, but they also differ in important ways. They involve students with high-intermediate English skills (e.g., Dabaghi, Zabihi, and Rezazadeh, 2012) and others are in an ESL context (e.g., Bax et al., 2019) or the language learners are preteen non-native EFL students (Kisby, n.d.).

Text reconstruction also has received mention by leading scholars in the field of language instruction like Brett (1994) and Thornbury (1997). However, there are few empirical studies in the specific study of text reconstruction and MDMs. Kameen (1978) provides insight into the value of collaboratively completed sentence-combining exercises and the experiential component of the activity, much of which speaks to the experience of the students in my study. However, Kameen’s exercises are paper-based rather than online, and the author is not concerned with MDMs’ role in creating coherence and cohesion. Rodríguez (2010) is concerned with using Hot Potatoes applications, including WebSequitur, to aid students in developing better paragraph writing skills. Her approach is to teach classical paragraph structure and she does not address the issue of MDMs and cohesion. Čechovičová (2010) and Rodríguez (2010) cover text reconstruction, but their interest is in students’ basic vocabulary or grammar skills.
This study addresses multiple gaps in the literature and points out a potentially important and hitherto unresearched technique for repurposing a teaching tool for better teaching practices, and helping universities reach internally set (Tokyo University of Science, n.d.; 2016) and governmentally mandated goals (Suzuki, 2015b).

To investigate these issues, I have set the following research questions:

1. To what extent does completing OTREs and particularly WebSequiturs influence language learners to use selected MDMs in their written arguments?
2. To what extent does completing OTREs and particularly WebSequiturs guide language learners to use MDMs appropriately in such a way that indicates skilful signalling of argumentative structures in written discourse?
3. What are participants’ perceptions of their experiences with OTREs and particularly WebSequiturs?

1.8. Overview of this thesis

In Chapter 1, I explained the experiences in my teaching practice that led to my interest in this subject. I framed my specific interest in CT instruction as an AW skills development issue. I defined L2 writing and discuss two major impediments to Japanese students’ developing AW skills. I situated the students learning in this study within as a weak form of CLT, leaning toward an EFL context. I also discussed national and institutional policy in response to these challenges. Next, I discuss MDMs and their relevance to AW skills. I explained the issues or MDM taxonomy. I then introduced the theme of OTREs as a resource for teaching the more frequent and skilled use of MDMs in AW. After that, I provided a description of the context in which students in this course are situated: the four-term programme and the aims of this first-year, second-term course. Lastly, I discussed the aims and rationale of my research. I concluded the chapter by noting some of the epistemological limitations and how I mediated those difficulties through a mixed-methods (Denscombe, 2008) use of quantitative and qualitative data.

In Chapter 2, I discuss the literature on the subject of AW and posit it as distinctive from a mere subset of CT. I note some of the problems of teaching university level AW skills in Japan. I discuss MDMs and their role in AW. I discuss the major issues of competing taxonomies and the methodological issued related to that. I discuss the affordances, constraints and influences of asynchronous and synchronous environments within computer mediated communication (CMC) and I then introduce text reconstruction exercises in detail, focusing on the pedagogical affordances of online text reconstruction. I note gaps in the
literature around the pedagogical affordances of using online text reconstruction exercises to teach appropriate use of MDMs in AW.
Chapter 2: Literature review: Metadiscourse markers and argumentative writing

2.1. Introduction

In this chapter, I review the literature relating to the major themes of this thesis. There are three main sections. In the first section, I note some of the critical views of CT instruction, including ones which question the validity of including CT in Asian educational curriculum. In the second section, I describe the relationship between CT, AW, and MDMs. Next, I discuss MDMs in greater detail. I then discuss methodological considerations central to this study. I examine challenges specific to teaching AW skills in university in Japan, and how mandates from Japan’s Ministry of Education are reflected in recent curriculum changes at TUS. I narrowly situate the writing of participants in this study within Communicative Language Teaching (CLT). Following this, I discuss Computer Mediated Communication (CMC). I define it in terms of its two main divisions, asynchronous and synchronous. I explain the differing pedagogical affordances of each of the forums and chat rooms and I relate those to the issue of AW and CLT. I then discuss the literature regarding the use of online language learning applications relevant to MDM use. I discuss OTREs, their history, and I then narrow my focus to WebSequiturs, and explain their unique and powerful affordances for MDM instruction in EFL contexts. I acknowledge the gaps in the research in the relevant areas of this study: the use of OTREs for teaching MDM use; and instruction in AW and MDM use in L2 writing. Finally, I list and clarify the research questions which emerged from my engagement with the literature and my prior experience in the classroom.

2.2. Critical views of CT instruction

When the discussion of CT instruction came to the fore in the mid-nineties, not all theorists in EFL believed that teaching CT skills in Asia was appropriate. Most notably, Atkinson (1997) recommended that caution should be exercised when considering introducing CT skills into Asian curricula. Atkinson argues that CT is as much a social practice as a skill and these social skills are culturally embedded and not easily transplanted; and that the values of some cultures are opposed to CT. He notes that CT instruction can sometimes lead to the marginalisation of some groups. Atkinson also argues that although some EFL learners may learn the lesson being taught, they do not carry that learning onward to new and different contexts. The first point Atkinson raises is the lack of clear definition of CT, and it is this point that Davidson (1998) of Gakuen University in Tokyo first counters, writing that the lack of a clear understanding of critical thinking suggests that teachers should receive better
training. He writes that much of Atkinson’s claim that CT is essentially an individualistic act is not true. Davidson notes that Ennis has written of CT as both an individual and a group act, and that all societies practice some form of critical thinking, simply by questioning the truthfulness of information presented to them. Davidson also argues that teachers have a responsibility to prepare students to interact with native English speakers, particularly in academic contexts, where critical thinking is part of the discourse.

I have highlighted these two articles, because they are early examples, and because so many of the subsequent arguments on the subject of CT instruction in Asia and Japan derive from same or similar arguments. While it is important to consider these caveats and their rebuttals, it is my critical position that the concerns about CT in Asian EFL curricula is not greatly relevant to this study. CT involves a broad range of skills and Atkinson’s concerns may be applicable to some of these. However, teaching university students to identify and signpost premises and conclusions is as much an area of AW as it is a CT concern. It is non-controversial. The subject of this study falls outside the scope of Atkinson’s argument and firmly within the purview of AW.

2.3. CT skills instruction and AW

Basic critical thinking skills of building arguments are necessary for students to become competent argumentative writers (Golpour, 2014). Nearly all academic works on the subject of CT distribute skills across three hierarchised areas in the following order: (1) analysis, i.e., identifying premises and conclusions; (2) evaluation of premises and soundness of reasoning; and (3) ability to infer and to generalise conclusions, i.e., knowledge transfer (Paul and Elder, 2014; Wright, 2013; Morrow and Weston, 2011; Hughes, Lavery and Doran, 2010; Porter, 2002; Fisher, 2001). These sources typically refer to words and phrases that inform readers of the presence of premises and conclusions as indicators (Fisher, 2001; Ennis, 1996), or signals (Cottrell, 2011; Barnet and Bedau, 2011). There is evidence of much association between these basic CT skills and skills development in AW (Golpour, 2014; Dabaghi, Zabihi and Rezazadeh, 2012; Hashemi and Zabihi, 2012).

In the following sections, I discuss premise and conclusion indicators in an EFL context by considering them in terms of metadiscourse commonly employed in AW.

2.3.1. MDMS and AW

In the following sections, I define metadiscourse and MDMs. I explain the effects of MDMs on the structure of the texts. I frame argument indicators as MDMs. I also discuss how MDM
use creates cohesion and coherence through skilled and appropriate use which comports with the expectations of readers. I discuss the varieties of reflexivity, citing Crismore, Markkanen and Steffensen’s (1993) textual and interpersonal and interactional categories. I then introduce Hyland’s (2005a) taxonomy and discuss the resulting narrow and broad categorisations. I discuss how Hyland’s departure spotlights an issue of thin and thick research methods and the critical decision I made to use Hyland’s taxonomy.

2.3.2. Problems in Japanese use of English metadiscourse markers in AW

A discourse marker is a word or phrase that indicates or describes a communication act (Blakemore et al., 2002) for the reader. Metadiscourse markers (MDMs) are words and phrases that cohere discourse, between one segment of a text and a preceding or subsequent segment (Fraser, 1999). In Chapters 5 and 6 of this study, as I discuss MDMs treated in this thesis, I use the word token to denote those metadiscourse markers. Metadiscourse may be described as the way in which a text references, not only the world outside the text, but also the world inside the text (Blakemore et al., 2002). An important point in describing AW is the attention given to metalinguistic markers, specifically discourse markers (Rahimi, 2011), which writers employ when anticipating how certain language usages are normatively interpreted within a given discourse community (Hyland, 2017). As students develop skills constructing arguments that span sentences, paragraphs, and texts, they begin to employ metadiscourse (Blakemore et al., 2002; Ross and Rossen-Knill, 2016) or ‘discourse across a text’ (Eggins, 2013, p. 28). The problems of Japanese students’ writing suggest a similar theme. Their writing is grammatical; however, at the sentence and paragraph level, as I discuss below in Section 2.5, signposting appropriate to AW in a university context, is often lacking.

Conversely, metadiscourse leads skilled readers to understand a text’s structure or how the writer feels about the topic or content (Basturkmen and Randow, 2014). MDMs like because and so indicate logical connections in textual structure. These can operate locally as with the logical markers used in propositional statements, i.e., statements of fact and opinion.

Propositional statements are statements, explicit or implied, which can be agreed or disagreed with, subjected to scepticism, falsified, or verified. In argumentation the corollary to propositional statements is the premise. A premise alone offers no clue as to where it fits in the overall structure of a text, but propositional statements, when posited within an argument, are often signalled by premise indicators. These provide cues to the reader about the text’s overall structure (Hyland, 2017). For example, it is not immediately clear whether It’s likely
to rain today is a premise or a conclusion. It is a problem that the reader must work out by using logic, their understanding of argument indicators and their real-world knowledge. However, when discourse markers like because and so are introduced, the distinction between premise and conclusion becomes more salient. Even if the indicator is attached to the conclusion rather than the premise, e.g., so, I should take an umbrella, it can more easily be inferred that It’s likely to rain today is a premise than if there were no indicator (see Table 2.1). In this argument, other indicators, such as furthermore, moreover, or simply and could be introduced to indicate additional reasons. While these do not explicitly regulate logical flow, they often do implicitly indicate premises. Such an example demonstrates, as Bax et al. (2019) point out, how discourse markers cannot be pinned down as either text referential or outside world referential; they often play a dual role. The same can be said of other kinds of markers that function as statement connectors like although, of course and even so and a wide variety of other commonly used discourse markers. Such markers not only implicitly regulate the logical flow of the writer’s argument; they also convey the writer’s attitude. In this study, I have termed these as implicitly connective indicators.

To avoid confusion with Hyland’s (2005a) use of logical connectors, which refers specifically to transition markers in the sub-category within frame markers, I opted for explicitly and implicitly connective indicator. These two terms are used throughout this thesis.

<table>
<thead>
<tr>
<th>Argument component</th>
<th>Sentence content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propositional statement</td>
<td></td>
</tr>
<tr>
<td>Premise</td>
<td>It’s likely to rain today, Conclusion</td>
</tr>
<tr>
<td>Conclusion indicator</td>
<td>I should take an umbrella</td>
</tr>
<tr>
<td>Indicator</td>
<td></td>
</tr>
<tr>
<td>Conclusion indicator</td>
<td>so Conclusion</td>
</tr>
<tr>
<td>Conclusion</td>
<td>I should take an umbrella Conclusion</td>
</tr>
<tr>
<td></td>
<td>because</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Premise it’s likely to rain today,</td>
</tr>
</tbody>
</table>

Table 2.1: Premises, conclusions and argument indicators

An example of premise-conclusion signalling which also conveys attitude is the use of phrases, I think and I believe that fall within the category of epistemic markers (Hyland, 2005a). When used argumentatively, these are typically conclusion indicators. In addition, they also tell us how convinced the speaker or writer is of the certainty of the proposition. Another example of premise signalling, and an important one in the second half of this study, is the hypothetical premise, a type of conditional statement also called the future predictive
(Ishikawa and Suzuki, 2016). Hypothetical premises, like all two-clause conditional statements, consist of an antecedent (the proposed conditional statement) and the consequent (the consequence of the conditions being realised) (Moore and Parker, 2012).

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Antecedent</th>
<th>Consequent</th>
</tr>
</thead>
<tbody>
<tr>
<td>If / When / In the event that / Unless (antecedent)</td>
<td>~ then (consequent) modal, e.g. will / is going to</td>
<td></td>
</tr>
<tr>
<td>Propositional statement</td>
<td>If there is air in the jar, then the candle will burn.</td>
<td></td>
</tr>
<tr>
<td>Unless there is air in the jar, then the candle will not burn.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2: Hypothetical premises, antecedents, consequents, and indicators

Some studies show that the structures of conditional statements are difficult for Asian students, including Japanese students, because of the variety of forms of conditional statements typically found in the languages of that region (Shintani and Aubrey, 2016). In their study of errors made by Thai students when constructing conditional statements, Katip and Gammper (2016) found that 16% of the errors involved hypothetical premises, i.e. future predictive conditional statements and that problems with subject verb agreement were involved in the five most prevalent of the error types. Bryant (1984) identified conditional statements as a trouble source among Japanese ESL learners. However, his study examined only conditional statements involving imagined situations (such as a differently experienced past, or a future event, such as winning the lottery and did not specifically examine future predictive conditional statements. In a study involving 53 Japanese university EFL learners, Ishikawa and Suzuki (2016) found that the participants tended to mistakenly confuse the verb patterns of the Future Predictive with those of the Present Counterfactual (see Table 2.3).

Conroy and Cupples (2013) also found a high frequency of errors among Japanese students in the composition of counterfactual conditional statements.

The literature indicates that the use of MDMs used in both premises and conclusions and in conditional statements are problematic for Japanese EFL learners. The lack of research into the hypothetical premises form suggests that there is a need for research in the instruction of conditional statements and particularly, the future predictive form.

<table>
<thead>
<tr>
<th>Conditional Forms</th>
<th>Argument component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future Predictive</td>
<td>If there is air in the jar, then the candle will burn</td>
</tr>
<tr>
<td>Present Counterfactual</td>
<td>If there were air in the jar, then the candle would burn.</td>
</tr>
</tbody>
</table>

Table 2.3: Future predictive and present counterfactual conditional forms
2.3.3. Discourse community MDM use

MDM use is community-based. Within any discourse community, a ‘shared repertoire’ (Wenger, 1998, pp. 73) of styles and discourse develops. The rules that constrain and describe those structures and inter-relations emerge from a ‘community of practice’ (Lave and Wenger, 1991), where, ideally, novice writers working within a ‘network of members’ develop skills that help them become competent writers in a given genre (Schmied, 2011, p. 5). Accordingly, instruction in argumentative writing should involve teaching relevant structures (Connor, 1990).

For learner writers to achieve that acceptance within their discourse community both frequency and skilfulness of MDM use are important. In Williams (1992), although some non-native speakers of English used MDMs as or more frequently than native speakers, their writing was sometimes scored as less comprehensible than native speakers. Unskilled use or overuse of cohesive devices, such as connectives, e.g., *besides* and *furthermore* can be deleterious to a text’s overall coherence (Basturkmen and Randow, 2014). In their studies of EFL pedagogy in Basturkmen and Randow found instruction in MDMs in the foreign language teaching curriculum to be generally absent. MDMs were sometimes used inappropriately or in ways that differ from the usage in their discourse community.

Relevantly, Fung and Carter (2007) found MDMs receive the most instruction in terms of their functions as *connectors*, e.g., *and moreover*, *in addition*, whereas their function in relation to propositional statements, i.e., premises, receive much less attention. Hyland (2008) found, similarly, a greater use of MDMs was not always a sign of greater proficiency in the writing of research level writers. Chen and Baker (2010) note differences between Hyland’s data sets which might compromise the validity of his conclusions. They found that when the problems of statistical validity in Hyland’s study were precluded in their own study, their results were supportive of the claim that increased approximation of learner English writing to native expert writing in terms of lexical bundles such as discourse organisers, i.e., metadiscourse markers, was associated with assessments of higher proficiency. In a larger and more recent study, (Bax et al., 2019) found that a greater use of MDMs did not always result in higher scoring writing. However, in the case of eight of those 13 categories, an above average frequency of MDM use did result in a higher-than-average assessment. Logical connectives were not among those eight. However, among the 24 phrases tracked in Bax et al. (2019), 14 of those, all of which were used by participants in this study as implicitly connective argument indicators, showed an increase in usage among the higher scoring texts.
2.3.4. Text-centred and user-centred use of MDMs related to AW

The use of MDMs is one way which typifies what Halliday and Hasan (1976) have called *texture*. Texture is the property which confers wholeness to clauses and can be approached analytically in terms of cohesion and coherence (Eggins, 2014). Noting that the presumption of an audience and a will to communicate with it is present in most or all AW, Cheng and Steffensen (1996) include cohesion and coherence among seven attributes of effective texts. Cohesion and coherence are classified as text-centred. There are also five user-centred attributes. All seven contribute to quality of writing. Cohesion enables meaning-making at the surface and coherence draws texts together at a deeper level (Basturkmen and Randow, 2014). While cohesion describes meaning generated by grammar and syntax, coherence is the result of relations within a text being made sufficiently explicit primarily through the distribution of MDMs. User-centred attributes include the paired *intentionality* and *acceptability*. The first two of these derive from an intent to communicate with an audience willing and able to make sense of the communication. This is achieved by engaging the attributes of *situationality*, *intertextuality* and *informativity*. In short, situationality involves selecting language appropriate to the context and relationship between the writer and reader. Intertextuality is concerned with a presumed shared knowledge of other texts and other text genres which the writer may reference and thereby aid and shape readers’ understanding. Informativity refers to the amount and level of information included in a communication. That amount and level should be appropriate to audiences’ intellectual capacity, education, and desire or willingness to read, i.e., acceptability. In AW, one way in which textual coherence (Halliday, 2009) is reached is through skilled use of MDMs. Skilled MDM use is reached through writers’ appropriate intentionality and readers’ sense of appropriate use. To this point, Cheng and Steffensen (1996) found that as students became more aware of the rhetorical and pragmatic utility of MDMs, their writing improved, then became more intentional, measured and skillful in their MDM use.

Student writers may lack the meta-linguistic awareness needed to understand that readers play a large role in making writing comprehensible when they cohere a text by applying background knowledge of how other texts of various rhetorical functions and genres that they have encountered have been organised (Basturkmen and Randow, 2014, p. 15). They need to know that they have both the responsibility (Hyland, 2009) and the means (Basturkmen and Randow, 2014) to signpost the organisation of their texts and thus guide readers’ inferences and conclusions.
A major problem in composition classes is the difficulty in conceptualising a meaningful audience other than the teacher. In the absence of a clearly defined audience, it becomes difficult to teach students to consider these attributes in their writing. Small groups and class forums can be used by teachers to address this problem (Cheng and Steffensen, 1996).

I will return to this discussion of attributes and signposting in Chapter 3, when I explain my rationale for determining appropriacy of MDM use and the role forums and chat rooms played in this study.

2.3.5. Critical issues in MDM taxonomy

In this section, I discuss basic taxonomical terms related to metadiscourse. I then note how Hyland’s (2005a) departure from the accepted taxonomies of Fraser (1999) and Crismore, Markkanen and Steffensen (1993) affects my own taxonomy and my position on reflexivity.

The term *metadiscourse* may be understood as referring to qualities of a text that fulfil the latter two of three macro-functions: ideational, textual and interpersonal (Halliday, 2009). *Ideational* corresponds to the content, or subject matter that is being referenced and is not considered to be metadiscoursal. *Textual* refers to the capacity for metadiscourse to function across a text to comment on structure and organization. This macro-function concerns how the writer embeds discourse markers that work as cues to inform the reader about the rhetorical structure, e.g., the cause-and-effect pattern or the logical flow of an argument (Kopple, 2012). *Interpersonal* refers to how the author predicts readers’ expectations and interpretations in the writer-reader relationship (Gee and Green, 1998). There has been a shift in the divisions of some of these terms over time from the taxonomy originally developed by Crismore, Markkanen and Steffensen (1993). Fraser’s taxonomy (1999) describes a somewhat different system with three main types of metadiscourse: contrastive, elaborative and implicative. However, both taxonomies maintain a distinction between textual and interpersonal discourse. Hyland (2005a) breaks from his predecessors by no longer separating textual and interpersonal metadiscourse. Hyland views textual metadiscourse as interpersonal, largely dispenses with the use of the term textual, and instead differentiates between *interactive* and *interactional* metadiscourse (Hyland, 2005a). This shift causes a major change in the way *reflexivity* can be understood. For reasons discussed below, I am working with Hyland’s taxonomy, rather than Fraser’s.

While most researchers in this area agree that reflexivity is a central point in understanding how writers and readers mediate understanding, not all agree that these two
qualities should both be defined as reflexive. The term narrow approach has been used to refer to the approach which accepts only interactional metadiscourse as reflexive and the term broad approach has been used to refer the approach which accepts both interactive and interactional metadiscourse as having the quality of reflexivity (Zhang, 2016). This view is also called integrative as it allows interactive categories to be integrated into metadiscourse (Zhang, 2016).

The two camps in this disagreement are located along largely separate paths which lead to incompatible methodologies. Whether I align this study with the broad or narrow view of metadiscourse informs the selection of research methods relevant to my research questions and the data being sought out or produced. In short, the narrow perspective leads to a thin research approach favouring a quantitative framework; and a broad perspective leads to a thick research approach allowing for both quantitative and qualitative frameworks. That is because, with the thick approach, the researcher looks at specific instances of language use. Language use is seen as highly contextualised cases of situated meaning. In my research context, the more contextualised the understanding of metadiscourse, the more accurate the data analysis is likely to be. As I explain in more detail in the following section, MDMs often operate on both an interactive and interactional level. For that reason, I have adopted a broad approach, favourable toward the integrative camp, for which, not only interactional but also interactive language use is reflexive. This has led me to a thick approach to data analysis (Ädel and Mauranen, 2010).

2.4. Methodological considerations and influential studies
In order to assess the influence of OTREs on the frequency of students’ MDM use and the extent to which OTREs guide students to use MDMs appropriately, data in the form of student writing from forums, chat rooms, and student reports needed to be analysed quantitatively and qualitatively. In developing a methodology for this small-scale action research study, several problems involving collection and interpretation of quantitative and qualitative data had to be considered. These were first, how to quantify frequency of students’ use of MDMs and secondly, how to qualitatively assess the appropriacy of students’ MDM use.

Concerning the first issue, the use of several categories of MDMs needed to be tracked. These included explicitly connective and implicitly connective premise and conclusion indicators, and several sub-categories thereof (see 4.1.8). Also, beyond frequency, other aspects of use needed to be tracked. These included the number of categories and the
Khedri, Heng and Ebrahimi (2013) provided an example by following the categorisation of interpersonal markers in Hyland (2005a) and then running a search application necessary for the number of documents involved. This facilitated the quantification of markers present in student written abstracts which they were analysing. I explain later how I developed a system for using successive word searches within MS Word, which proved to be adequate for the comparatively lower amount of text analysed in this study. This process involved tagging each MDM according to its communicative category and this became useful later during the qualitative analysis.

Concerning qualitative analysis, Intaraprawat and Steffensen (1995) provided a useful method of assessing appropriacy of use. In their study, six academically qualified judges independently assessed 47 placement test essays written by EFL students, several of them Japanese, at a university in the United States. Using this means for assessment, it was found that, among poor-scoring essays, there was a sudden drop-off in the number of markers used to anticipate readers’ comprehension; and cases in which MDMs were used more frequently received higher scores in areas reflecting the effectiveness of arguments. Cheng and Steffensen (1996) also utilised three raters with similar results. Sanford (2012) also employed raters to assess student written texts in her study and found that, texts by novice writers using MDMs such as transition markers, including connectives such as because and so, were judged to be more skilled. The three preceding studies’ use of raters became a template for my own analysis and owing to the training of the raters, assessments were gathered could lead to meaningful interpretations.

2.5. Challenges to learning AW in Japan
In Japan, two challenges face first-year students in their development of AW skills. First, students entering university commonly have difficulty in recognizing and forming argumentative structure (Wingate, 2012). Japanese high school students are trained to write in the ki-shoo-ten-ketsu pattern (Maynard, 1998), consisting of an introduction, support, unexpected turn, and main point, revealed in the conclusion. For this reason, Japanese rhetoric is heavily inductive (Maynard, 1998). Readers bear more responsibility to make sense of texts than with western writing (Hirose, 2003; Kubota, 1998; Maynard, 1998; Hinds 1987).

Hyland (2017, p. 17) coined the term ‘recipient design filter’ to describe stylistic decisions writers make to employ MDMs to fit in with common usage within the target
discourse community. Skilled writers specifically employ discourse markers in light of how certain writing styles and language item usages are normatively interpreted within the target discourse community (Hyland, 2017). The writing of learning writers tends to progressively approximate expert writing through the appropriate use of discourse markers (Chen and Baker, 2010). Using the ki-shoo-ten-ketsu form to write argumentatively in a university context is an example of a recipient design filter playing a counter-productive role, as novice writers overgeneralise its utility. Writers switching between languages should ‘anticipate differences from how thoughts progress’ in one system or another (Maynard, 1998, p. 38).

One perspective, therefore, is to consider each language (and culture) as having preferred rhetorical functions. Conversely, it is also argued that educational background and meta-knowledge in the writing process of texts is one among several major factors determining the learning of rhetorical form (Kohro, 2009). Kobayashi and Rinnert (2008) comment on how Japanese students with L1 writing training tended to write in a more expository mode and to use the unexpected turn of the ki-shoo-ten-ketsu pattern. Conversely, as Kubota (1998) points out, it is widely believed that a major problem for novice Japanese writers is L1 interference (Ellis, 1994), that is when a ‘negative transfer of L1’ results in ‘inappropriate usage of L1 linguistic knowledge during L2 production’ (Yossatorn et al., 2022, p. 10). Japanese English learners’ writing is typically grammatical, across paragraphs, but often lacks the logical structure, and signposting western readers generally expect (Thornbury, 2005). Kubota (1998) concludes in her study of Japanese university students’ argumentative and persuasive writing that, although L1 interference cannot be discounted, the point most contributing to problems in students’ writing was simply lack of experience in argumentative and persuasive writing. To this point, Kobayashi and Rinnert (2008) found that Japanese students in their study who had also received second language (L2) training in writing, showed a marked difference in rhetorical mode and were better able to write argumentatively with clear premise-conclusion structures.

A second major impediment to developing AW skills is societal. In a first-year English class at TUS, responding to a writing assignment on cultural identity, one student wrote, ‘In Japan, people are taught not to be expressive [sic]. They tell their children that to show everything you think straightly is not good thing’ (Anon., 2016a). Another wrote, ‘Expressive culture is not encouraged. If you made expressive communication, people would think you strange and might think you impolite person’ (Anon., 2016b). Bao (2014) likewise found Japanese university students associated speaking up with rudeness.

This constraining societal factor has become pedagogically institutionalised. One
full-time TUS English teacher interviewed by Jennings (2018, p. 94) speaks of a ‘very deep gap’ between traditional, grammar translation instruction methods (yakudoku) focused on passing entrance examinations (Bao, 2014; Kohro, 2009) and student-centred teaching approaches focused on communicative competence. Relying on grammatical analyses and translations to and from Japanese as primary methods, English lessons have historically consisted of mechanical exercises designed to develop knowledge that can be accessed mentally at the time of a test (Humphries and Burns, 2015; Bao, 2014; Kobayashi, 2013; Hu and McKay, 2012). One first-year university student in Kobayashi and Rinnert (2008) said that during her pre-university first-language (L1) training, she had learned that, when writing an essay, the main reason for her opinion should be withheld until near the end in order to maintain readers’ interest. Thus, the teacher of first-year university English must often teach against the grain of students’ earlier training.

We have then, two ideas about why Japanese undergraduate students have difficulty writing argumentatively. In one view, societal pressures caused by cultural differences ingrained into rhetorical practice factor largely in the reasons why students often cannot easily express, in language appropriate to university discourse, what they believe and why they believe it. In the other view, the difficulty is seen largely as a problem that could be solved by effective teaching methods. This thesis is mainly concerned with the latter of the two views. The former view, however, is relevant as perceived contradictions arise when students are tasked to express agreement, disagreement, and other critical views of other students’ opinions in forums and chat rooms.

2.6. CT and AW pedagogy at national, institutional and classroom levels

The problem of improving university students’ CT and AW skills has been a greatly discussed topic in education in Japan since the mid-2000s. At the level of national policy, recently, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) mandated that CT skills be taught at university level (MEXT, n.d; Stroupe, 2006).

In 2011, MEXT launched a plan to open two hundred International Baccalaureate programmes by 2018. Among the core components are critical dispositions and CT skills (MEXT, 2011b). Also, MEXT’s 2011 Five Proposals and Specific Measures for Developing Proficiency in English for International Communication calls for the development of students’ ability to express and argue using ‘logical and reasoned explanation of one’s own views’ (MEXT 2011a, p. 3) and there are numerous mentions of the need for debate skills.
throughout the document. More recently, MEXT has also announced its intent to intensify this plan by 2030 (Suzuki, 2015a). There is a trend, globally, as in Driver et al. (2000), to cultivate argument as a human and social practice, and MEXT is responding to a perceived change, globally, in the importance of higher order thinking skills. In 2015, the Organisation for the Economic and Cooperation and Development (OECD) reported that from 1960 there was a swift increase in the demand for workers capable of autonomous decision making in novel contexts, and that this trend has increased steadily since the 1990s (Suzuki, 2015b). There has also been an increase since the early 2000s in the demand for workers with high level problem-solving skills (Suzuki, 2015b). The OECD has advanced an overarching curriculum redesign involving a focus on the overlapping area of factual knowledge, critical thinking, critical dispositions and collaborative abilities (Suzuki, 2015b). In addition, high schools are now implementing collaboration between peers, enquiry into global issues and the development of foundational CT skills, the ability to work cooperatively, think scientifically and logically (see Appendices 9-10), and to express oneself clearly in order to communicate with other professionals and academics (Suzuki, 2015b).

2.7. Situating students’ writing within EFL

In this section, I posit the writing of students in this study within a weak Communicative Language Teaching (CLT) context. CLT is a movement in EFL which began in the 1970s as a response to the perceived inadequacies of traditional methods and a movement towards teaching English as tool for carrying out communicative acts as contrasted with teaching English as a subject (Littlewood, 2014). CLT has also been criticised for being a vaguely defined collection of differing approaches gathered within commonly held notions about language teaching (Littlewood, 2014). While CLT has been seen as an improvement over the formerly dominant grammar-translation method, it has been criticised as an import from ESL, imposed on the EFL context in Asia and in Japan, without sufficient consideration for how it would work with national and local societal norms (Lochland, 2013; McKinley, 2013; Ikeda, 2012). Bax (2003) argues that EFL should now move beyond CLT, with its prioritizing of specific teaching methods and exclusion of others. He argues for a prioritization of contextual concerns. It is my position that, although these reservations and criticism of CLT are applicable in some or many cases, to the extent that the teaching methods used in this study were consistent with CLT, they were likewise not at odds with the context in the classroom and at the university. Moreover, adaptations to CLT have been made by researchers and practitioners working within EFL contexts. Figure 2.1 shows ELT methodologies on a
There are strong and weak forms of CLT. The strong form is closer to the ESL pole, where language acquisition is typified as occurring naturally. The weak form is closer to the EFL pole, typified by a learning style that is more structure-based. The weak form is widely seen by teachers as leaning towards students’ later learning in a less teacher-controlled and less predictable environment (Butler, 2011). Proponents of strong CLT eschew traditional teaching methods such as explicit instruction and drills, while proponents of weak CLT allow for using some traditional teaching methods as a means of working toward a strong CLT environment (Littlewood, 2014). Working autonomously is an important skill suitable for development in a task-based instruction (TBI) context (Butler, 2011). Further to the point, critical thinking is an important component of the ability to work autonomously (Lu and Wang, 2021).

Figure 2.1: ELT methodologies, Ikeda, 2012

Notwithstanding the criticisms of CLT and the adaptations that have been made to respond to the differences present in an EFL environment, the weak CLT approach is an appropriate starting point for students at a science university where English study is not a top priority and is an apt descriptor for the context of this study.

2.7.1. Computer mediated communication
In this section, I discuss CMC as a field of study in EFL and as it relates to my thesis. I first define it and note its two main areas: asynchronous and synchronous; and the kinds of
environment most common in each of those, forums and chat rooms, respectively. I discuss some of the theory regarding CMC in EFL and note the pedological affordances present in forums and chat rooms. I note how social constructivist theory explains how each environment motivates L2 learners differently and how the writing in each environment tends to have different qualities. Finally, I explain the relevance of chat rooms, normally associated with written approximations of speech, to my research on written argument. Research like that represented in this study is highly compatible with data gathered from synchronous chat room interaction, where social interaction and meaning-making is direct (Sotillo, 2000); and also compatible with data gathered from asynchronous forum threads which have other properties that make them differently suited to this study. Swain’s (1985) output hypothesis explains how CMC can be instrumental in L2 learners reaching greater sociolinguistic competence in the use of MDMs.

2.7.2. Forums and chat rooms

CMC, at the broadest view, is any form of sustained communication designed for the purpose of communication, be it text, audio, audio-video, or the transmission of text, audio and/or video via email or other file sharing means. CMC is largely categorised as synchronous and asynchronous (AbuSeileek and Qatawneh, 2013). In this study, which incorporates student writing from three sources: forums, chat rooms and an online shared MS Word document, the differences of each environment and the influence on students’ written output is explained in terms of affordances, and how this determines what communicative functions and genres student writers engage in. Somewhat to that point, Sotillo (2000) found that synchronous text interactions, such as chat, fostered fluency, as well as more informal, oral competence. Chat-based Synchronous Computer Mediated Communication (SCMC) has more in common with speech than forum-based Asynchronous Computer Mediated Communication (ACMC), which, conversely, has much in common with writing, although there are also important differences (Eslami, Mirzaei, and Dini, 2015). In addition to drawbacks there are advantages to these differences. Schmidt (1990) argues that what learners notice in input is what becomes intake for learning, regardless of whether a learner deliberately attends to a linguistic form. Although noticing is an internal and private process that cannot be observed directly research indicates incidental learning is both possible and effective when a task requires learners to focus on what is to be learned (Schmidt 1993). Similarly, Lai and Zhao (2006) found that language learners were able to notice their own errors and were more motivated to self-correction. To direct learners’ attention to form,
Swain (1985) proposed the output hypothesis, arguing that output pushes learners to process language more deeply than input. Students’ output thus arguably contributes more strongly to their inter-language (IL) development. To achieve the facilitative effects of interaction, be that face-to-face or in a forum or chat, learners need to: (1) ‘notice the linguistic characteristics of the target language input that they receive’; (2) ‘have opportunities to produce target language output’; (3) ‘notice errors in their output’; (4) ‘correct their linguistic output’; and (5) ‘engage in target language interaction whose structure can be modified as needed for comprehension’ (Chapelle, 1999, p. 109). Output may promote noticing and hypothesis-testing in that some errors which appear in learners’ written and spoken language demonstrate their hypotheses on how the target language is used (Swain 2000).

Typically, online forums, like the forums in this study, are asynchronous. Posts are often made with no specific interlocutor in mind, and often other members of the forum may simply read others’ posts with no assumed responsibility of replying. Conversely, online chats, like the chats in this study, are typically synchronous. Often, messages are posted with the expectation that there will be a prompt reply, often in real time. Forums typically have a format in which successive replies to a given post are nested, that is, indented to the right, creating multiple ragged-edged threads within the topic. In this way visitors can easily intuit the forum’s chronology, left to right as well as top to bottom and contributors can contribute to various threads, irrespective of when the previous or subsequent comment was made. This feature is usually absent in chats which are displayed in a top-to-bottom chronology. Because of this, as in oral conversations, a less linguistically advanced interlocutor may be unable to form a response or statement before the conversation has moved on to a different topic. In a forum, visitors could simply scroll up and reply to the post at the appropriate point irrespective of the ongoing flow of discourse. For this reason, forums can contain large numbers of members without some members being marginalised. Chat rooms, by contrast, must be constrained to a small number of members to avoid the conversation being dominated by the more advanced and outspoken members (Andujar and Salaberri-Ramiro, 2021).

Following Vygotsky’s (1978) observation that tools for communicating become an intrinsic part of that communication, researchers in CMC stress that forums and chat rooms are not simply a place to engage in communication; they are also a part of that communication and a factor in the students’ developing socio-linguistic competence. Thus, awareness of genre is positioned before structural linguistic skills, rather than after. Sykes (2005) argues that CMC can present or model pragmatic usages of language in an authentic way while also providing learning opportunities involving other language skills. A case in point is Abrams...
David A. Gann (C8846053)

(2008), who notes how students’ socio-pragmatic practices in chat developed as they increased their performance in using discourse markers for greetings and openings and closings of subjects, which they did not typically do in face-to-face discussions. Abram’s (2013) mention of how L2 practices were more prominent in the latter half of the term further indicates that the L2 learners gradually became acclimated to the CMC environment, as they noticed gaps between their own language use and that of their peers and then restructured their own communications as a result of heightened genre awareness. Abrams (2013) also notes that modelled L2 practices became more prominent in the second half of the term in which he conducted his study. These observations further support the idea that practitioners may begin with a weak CLT approach and transition towards a strong CLT approach as learners gain more communicative autonomy.

2.7.3. CMC’s impact on form

It has been observed that English learners conversing in asynchronous environments such as forums tend to ask questions that elicit longer answers and are more likely to post longer answers (AbuSeileek and Qatawneh, 2013). By contrast, English learners in synchronous environments, such as chat rooms, tend to communicate with short, simple, but clear expressions. However, learners in forums are also likely to use a variety of communicative categories such as asking for or offering clarification by giving examples or repetition (Ritchie and Black, 2012). These communicative categories are more often seen in language learners engaging in synchronous chat than with learners working within asynchronous environments. For this reason, synchronous environments are viewed as somewhat more conducive to fostering learner autonomy than asynchronous contexts (Sotillo, 2000).

As for the linguistic structure that typifies each of these two contexts, the posts that language learners make in chat rooms tend to be lexical-conceptual in nature (Sotillo, 2000); chats are more speech-like, while forum posts tend to be more planned, formal and often have higher lexical density. By contrast, the posts made in forums tend to be written at a predicative-argumentative level and this influences how the posts are structured (Sotillo, 2000). The forum environment is particularly suited to development of culturally appropriate language uses aligning with the genres and registers germane to debate. Thus, although forums are typically less immediate and direct than are chat rooms, as was noted above, they do have important pedagogical affordances lacking in chat.
2.7.4. Motivation in CMCs

Motivation and lowered anxiety within CMC environments is also an important issue. Ritchie and Black (2012) note that learners engaging in synchronous chat and asynchronous forums are often motivated to take greater care that the formations of their communications comport with the genre of the discourse community and general culture they are seeking to communicate within. In their exchanges, by comparing working memory (Robinson, 1995) they are therefore more likely to notice the gaps (Schmidt, 1990) between NS formations and their own, and to self-correct. For example, when discourse in a forum shifts from general discussion to debate, the change in genre is reflected in a shift in tone and register which NNS’s must notice and then adapt to or ‘align’ with (Uzum, 2010). In other words, pragmatic competence results from culturally relevant uses of language with language learners responding critically to interlocutors’ meaningful communications with likewise meaningful responses, formulated within the parameters of a given genre, more so than as a tasked response from student to part of an assignment. Pragmatic competence is typically followed by structural competence. Exchanges which the language learners value as meaningful can lead to their noticing the gap between their communications and the expression used by equally competent or more capable interlocutors. In their study, involving two intervention groups with and without explicit instruction, both of which engaged in CMC, in terms of pragmatic development, both groups outperformed the control group, which received explicit instruction, but did not participate in CMC. It was also found that the group with the explicit instruction outperformed the group with no explicit instruction (Eslami, Mirzaei, and Dini, 2015). Despite the advantages of CMC in creating more authentic learning experiences, the CMC environment cannot be expected to produce a perception of complete authenticity. Students cannot forget that much of their communications are, in fact, part of a class experience and subject to assessment. Contradictions between goals implicit in such genres of class assignment to ask classmates to provide justification for opinions concerning discussion issues and perhaps to express disagreement can cause tensions which affect the appropriate use of MDMs (Skalstad, 2005).

CMCs share many affordances although some attributes are more pronounced in one environment than the other. Generally, they facilitate language learners focus on form (Andújar and Salaberri-Ramiro, 2021). The ability to delay one’s response in a CMC reduces cognitive load and lowers learners’ affective filters so that they feel less threatened by failure while communicating in English (Andújar and Salaberri-Ramiro, 2021). The opportunity to reflect on what one intends to communicate is motivating to language learners (Bailey,
Chat room interaction can also result in focus on form (Eslami and Kung, 2016). Moreover, because synchronous communication specifically, like that found in chat rooms, involves a greater number of hurried responses, problem areas in language learners’ linguistic competence are more likely to be displayed (Andujar and Salaberri-Ramiro, 2021). These can later be corrected by the instructor or through students’ self-correction.

Other factors can also affect students’ online engagement. Belz (2002) notes the influences of various factors such as the affordances of social and institutional contexts, individual users’ personal histories and experiences with technology, and differences in availability and ease of access due to socio-economic disparity. Differences in students’ previous knowledge of forums and chat could affect student output as Belz (2002) observed.

2.7.5. Imitation, internalisation, interlanguage and social constructivism

Interlanguage, briefly defined, is a language learner’s internally constructed system, which, to the best of the learner’s understanding at any given time, approximates the language system he or she is attempting to learn (Hosseini and Sangani, 2015). It results from the L2 and the native language being combined in ways that are often not entirely accurate or appropriate (Yossatorn et al., 2022). Lantolf and Thorne (2007) explain how purposeful imitation of recognised language forms or units can move through stages of approximations with the L2 until the learner reaches competence or proficiency. This is described in Aljaafreh and Lantolf (1994) as a shift from other-regulation (via modelling, correction or other means) and self-regulation, where the language learner reaches a level of autonomy in which other-regulation is no longer needed to produce structures correctly and use them appropriately.

Hosseini and Sangani (2015) describe how purposeful imitation, through a process of linguistic hypothesis testing, develops within social settings which provide feedback as to whether the communication was successful and whether the language use was appropriate (Hosseini and Sangani, 2015). This process was first theorised and studied by Lev Vygotsky (1978) as he developed his social constructivist learning theory, writing how learning first takes place in the social world and is later internalised (Barohny, 2016). Learners try out forms and expressions in their attempts to communicate with the means and limits allow for by their Zone of Proximal Development (ZPD) (Vygotsky, 1978). OTREs model the use of various MDMs and learners are tasked to imitate and internalise these modelled uses. These concepts therefore become important as they later aid in interpreting forum and chat room
communications, subsequent written work, participant talk during the OTREs and responses during semi-structured interviews.

In a small-scale study (Gánem, 2006) of types of tasks that facilitate high quality collaboration, three types of activities were used in both paper and computerised formats. Specifically Hot Potatoes generated collaborative tasks were used in the cases of the groups completing the tasks in computerised formats. Two metrics were used: one measuring turn-taking and another measuring the degree to which participant interactions were linguistically focussed, as compared with being concerned with general comprehension matters. The findings were mixed and the groups completing the computer-based versions of the tasks did not always outperform the groups completing the paper-based versions. However, in the case of the task which most closely resembled the OTREs there was a clear bias towards the superiority of the computer-based versions of the task both in terms of turn-taking and other related interactions and in terms of language related episodes which culminated in internalisations of the targeted language points. Mohammed, Assam and Saidi (2020) investigated the use of Web 2.0 tools, including WebSequiturs. Through the use of a control and an experimental group and pre-and post-testing, they determined that the experimental group had made statistically significant gains compared to the control group in the areas of listening, speaking and relevantly here, reading and most importantly, writing. The specific skills measured are not delineated in detail, but involve the intake, i.e., internalisation, and later production of chunks and set phrases such as common expressions, collocations, and aphorisms. Therefore, their study was not unlike my own.

Lastly, Sutton (2001) provides a different way of considering interaction between group members, introducing the idea of the vicarious interactor. Sutton notes that given the complexity of human psychology there are various reasons why motivated learners may choose to remain silent or offside. She argues that more passive learners often learn effectively by ‘actively processing the interactions of others’ (Sutton, 2001, p. 232). This style of learning can be further facilitated when learners are all focussed on common content. Vicarious learning is a concept that can be applied later in this study. To this point, Boothby et al. (2014) have shown that experiences that are shared, even between vicarious interactors, are generally felt to be more intense and are likewise recalled more vividly. Hence, collaborative tasks can be experienced with greater resonance.
2.7.6. Learning transfer between ACMC and SCMC environments
The general emphasis on the bridge between language learner chat and oral competence notwithstanding, chat-based interaction can also lead to increased skills in more formal written MDM use. Wishnoff (2000) demonstrates that language learners who engaged in SCMC underwent a pragmatic interlanguage development. Language transfer describes how some rules or elements may be caused by L1 interference or by the method of L2 language instruction (Ellis, 1994). Wishnoff (2000) describes one case of language instruction which later was seen to have transferred to students’ academic writing skills in the form of increased use of hedges. Of further relevance is the dual function that many of these hedges regularly served as or within argument indicators. For example, Wishnoff (2000) studied L2 learners’ use of hedges like, framing statements like Even though . . . and modifiers (and alternatively, quantifiers) like In several cases . . ., which can function in or as premise indicators, and It seems probable that . . . and These results suggest that . . ., which also function as or in conclusion indicators. Students in the study were provided explicit instruction in the purpose and use of hedges and then tasked to use them in SCMC. Students’ subsequent use of these markers in more academic papers increased in frequency and were more skilfully employed in terms of pragmatic usage. On this basis, the use of chat room data in my study is merited, as chat-based discourse is likely to have been a factor in participants’ later MDM use in subsequent forum discussions and in their Group Written Report (GWR).

2.8. Podcasting and instruction
As I have noted, text reconstruction work was preceded by instruction delivered via podcast. The pedagogical advantages of the podcast format of instruction have been well-documented in the previous two decades. Podcasts, as a means of delivering instruction, have both merits and possible weaknesses. One relevant merit is that podcasts, when listened to out of class, allow for increased time for in-class student-student and small-group discussion (Rosell-Aguilar, 2007). Another merit is that podcasts can provide a dynamic instructional experience (Rosell-Aguilar, 2007). As students can listen to a podcast, or parts of a podcast multiple times at their convenience (Sloan, 2005) the potential for out of class listening is further maximised compared to traditional language labs with cassettes or CDs that must be listened to at that location. Lastly, when podcasts are scripted, carefully worded dialogues can be drafted at the level appropriate for the students at whom the podcast is directed (Gann and Bufton, 2012a). One weakness is that some students may not take the sufficient time and
effort to listen attentively to the podcasts (Sabet and Mahsefat, 2012) and this might affect the
students learning.

2.8.1. Text reconstruction exercises
In this section, I discuss the history behind online text reconstruction. I also cover what some
researchers in the field have said about the pedagogical affordances of text reconstruction and
specifically what they say about the pedagogical affordances of text reconstruction in the
instruction of argumentative structure.

In the first two sections, I discuss text reconstruction historically and as a disruptive
trend in the field of CALL. In the next two sections, I introduce WebSequiturs, discussing
their place in the family of text reconstruction and discussing their unique pedagogical
affordances. I then discuss how they can be repurposed for teaching MDM use. Finally, I note
the limits in the research up to present point out how this study seeks to address those gaps.

2.8.2. Background of text reconstruction
Prior to the availability of personal computers, some teachers made paper-based exercises by
hand. This was time-consuming, and, in large classes, impractical. The University of
Birmingham produced paper-based materials, and in 1981 made commercially available a
package titled Skills for Learning (Johns and Lixun, 1999). Traditional paper-based text
reconstruction exercises are typically comprised of a selection of short, scripted passages.
The text being parsed may be an exact version of what the learner has read previously or it
may be a summary. Text reconstruction exercises involve, as the name implies, the
reconstruction of short written passages performed on a computer.

Digital TREs were developed in the early 1980s at the University of Birmingham by
Tim Johns. Soon after that, John Higgins developed Storyboard, a program that similarly
required learners to reconstruct texts. Soon after, John and Muriel Higgins developed
WebSequitur, which along with a cloze exercise making tool called Rhubarb are now
available through Web-based platforms such as Creative Technologies’ TexToys (Levy,
1997). For reasons that remain unclear, however, WebSequitur is rarely mentioned. For
example, in his otherwise detailed short history of the applications which he and his
colleagues and associates developed, no mention of WebSequitur appears, unlike Rhubarb
which is mentioned (Davies, 2007). WebSequiturs share enough commonalities with the other
applications that many generalisations are justified. In their review of Hot Potatoes, Winke
and MacGregor (2001) write that the Hot Potatoes suite ‘is an excellent resource for creating
on-line, interactive language learning exercises that can be used in or out of the classroom,’ but they qualify this by adding that Hot Potatoes exercises ‘can only be considered interactive in the narrowest of senses, and amount to little more than online versions of traditional grammar activities’ (Winke and MacGregor, p. 32, 2001). They conclude that Hot Potatoes can be used in EFL, but ‘the extent to which they do so depends on the content of the exercises created by the teacher’ (Winke and MacGregor, p. 32, 2001).

2.8.3. Text-reconstruction as a trend within CALL

Historically, it can be said that the use of technology in the classroom has been dominated by approaches that see the use of computers in terms of Computer Assisted Language Learning (CALL) (Motteram, Slaouti, Onat-Stelma, 2013). Proponents of CALL have worked for decades against scepticism that came from the failure of some of the first uses of technology in the 1970s with university language learning centres consisting of large rooms with stalls, a tape player and a set of headphones. CALL can be divided roughly into three periods. First is the structuralist period, characterised by behaviouralist psychological models. This period began in the 1950s and ran into the 1970s. The second period, typified by Communicative Language Teaching (CLT), began in the 1980s and continued until the second half of the 1990s, when the third period, integrative CALL came to the fore, as technology such as Hot Potatoes advanced and computers became more accessible (Sadeghi and Soleimani, 2015). CALL has progressed far beyond the audio-lingual approach used in those language centres. Nevertheless, CALL has remained essentially a branch of applied linguistics and second language acquisition (SLA). With the recent surge in the interest and use of approaches stemming from social constructivist learning theory, however, an increasing number of teachers advocate for shifting more of the emphasis in CALL over to more educational perspectives. Proponents of this shift argue for more emic approaches that take into account the factors introduced by the contextuality of the classroom (Motteram, Slaouti, Onat-Stelma, 2013). There is a broad consensus across a range of studies which stress the need for a sociocultural framework where technology is used with a focus on learning through interaction.

Accordingly, teaching AW skills often involves guiding learners to focus their attention on certain textual features such as the components of cogent argument, e.g., premises, conclusions and the lexical items commonly used to perform these functions. By initiating ‘noticing’ acts, OTREs can help to activate implicit and explicit knowledge (Brett, 1994; Thornbury, 1997). Computers are designed to be interactive and their design and the
design of the software affects that interaction. More engaging software contributes to greater learning outcomes. Computer-based learning is more engaging than traditional classrooms as software can be geared to different levels of difficulty and offer immediate feedback. It also offers avenues for self-correction and provides positive reinforcement (Ariew and Frommer, 1987) cited in (Čechovičová, 2010). With proper scaffolding, they can also help to foster learner autonomy (Vargas and Monge, 2014). Tasking students to work on computers within a social constructivist framework displaces the top-down teacher as ‘expert’, and redirects students’ focus on the computer (Langford, 2005, p. 142).

Hot Potatoes are described as interactive and effective teaching tools, when content and users are matched appropriately (Sadeghi and Soleimani, 2015). By the same rationale, the application is only as effective as the content and design which the teacher uses (Solemani and Raeesi, 2015). Although WebSequitur is not technically part of the Hot Potatoes suite, in terms of developing students’ learner autonomy, Hot Potatoes generally enjoys positive reviews and high praise by users (Solemani and Raeesi, 2015). Bogdanov (2013) notes the language hypothesis testing aspect of the platform design. However, despite the immediacy of response-time, and the potential for divergent paths depending on what choices the user makes, the interactivity in these types of applications are ‘limited to the interaction between user and pre-scripted feedback’ and are ‘interactive in the narrowest sense’ (Sadeghi and Solemani, 2015, p. 2413). Speckens (2011) describes the environment in Hot Potatoes as based on behavioralist models, with positive and negative responses used to condition the user. Arneil and Holmes, two of the early creators of Hot Potatoes, recognised this weak point in the applications design and worked to ‘provide much more in the way of interactivity than the one-click right/wrong paradigm’ (Arneil and Holmes, 1999 p. 14). This was achieved to a degree, by allowing multiple correct answers and automated feedback which offered short explanations in the case of partially correct answers. This only applies to some applications in the Hot Potatoes suite. WebSequitur does not support multiple correct answers and does not offer explanatory feedback to the user. Teachers can edit the response to offer explanation for each incorrect selection, but this is not automated and could become time-consuming.

2.9. Semi-structured interviews
Interpretivist interview practices frame the interview less as an information extraction process and more as a process of data production in which interviewer and subject share in knowledge construction (Curato, 2012; Kvale, 2007). Interviews can be doxastic, focusing on
respondents’ beliefs (Brinkmann, 2007). At times, these interviews can take the form of conversations, containing epistemic threads which call upon the respondent to justify their beliefs (Curato, 2012). Interpretivist research can be contrasted with the work of a miner, digging for prescribed data which exists prior to and independent of the inquiry, compared to the discourse of a traveller, producing the data as he travels. The traveller-interviewer does not purport to be an expert with a perspective superior to his interlocutors, the interviewees, but gives respondents an opportunity to ‘tell their story in their own terms’ (MacCracken, 1998, p. 34). Thus, semi-structured interviews are used to reach a contextualised understanding of quantitative data, e.g., frequency and the amount of students’ selected MDM use.

2.10. Limits of available research
The literature on the subject of the use of WebSequitur use is scant and what does exist largely focuses on its efficacy in teaching general vocabulary and idiom and grammatical constructions. Few academic studies on the efficacy of the use of the Hot Potatoes applications suite have been published (Rodríguez, 2010) and, on the subject of the efficacy of WebSequiturs in language teaching, almost nothing. The co-creator of WebSequitur, John Higgins wrote two articles collaboratively which discuss WebSequiturs. Higgins and White (1999) focus predominantly on the application’s utility in guiding language learners to notice grammatical or syntactical features of a text or develop an awareness of context in narrative texts. More to the point, they note that breaking sentences at ‘constituent boundaries’ the utility of WebSequitur can be made to ‘reduce grammatical clueing’ thereby ‘forcing the learner to use discourse clues’ (Higgins and White, 1999, p. 143). In Higgins, Lawrie and White (1999), there is a marked emphasis on using WebSequitur to model the structure of paragraphs and MDMs such as connectives, as well as ‘situation-problem-response-evaluation’ scenarios (Higgins, Lawrie and White, 1999). WebSequitur, and another application called Rhubarb, were written by the same authors, as was the Hot Potatoes suite, but are licensed through Textoys, whereas Hot Potatoes is produced by Half-Baked Software, Inc. (Sadeghi and Soleimani, 2015). For that reason, WebSequitur was not included among the suite of educational applications that Hot Potatoes offers. That suite includes applications for creating cloze exercises, crosswords, matching exercises and quizzes. The WebSequitur application continues to be offered online for free, but Hot Potatoes have not promoted it, and only a few articles mentioning it, and fewer studies using it, can be found. There is, therefore, a gap in the literature which this study seeks to fill.
There is growing body of research on the use of MDMs in AW in language learners’ L2, but the extant studies differ from mine in important ways. They involve older students with high-intermediate English skills, rather than first-year students with more limited abilities (Dabaghi, Zabihi, and Rezazadeh, 2012); or are located in an ESL context, rather than an EFL one (Noble, 2010; Bax et al., 2019); or involve adolescent native English speakers, and the rhetorical mode focused on is narrative, not argument (Sanford, 2012); or the students are preteen non-native EFL students (Kisby, n.d.). Kameen (1978) provides insight into the value of collaboratively completed sentence-combining exercises and the experiential component of the activity, much of which speaks to the experience of the students in my study. However, Kameen’s exercises are paper-based rather than online, and the author is not concerned with MDMs’ role in creating coherence and cohesion. Rodríguez (2010) is concerned with using Hot Potatoes applications, including WebSequitur, to aid students in developing better paragraph writing skills. Her approach is to teach classical paragraph structure and she does not address the issue of MDMs and cohesion. Nearly all the research on text reconstruction involves or is concerned with basic vocabulary or grammar focused lessons (Čechovičová, 2010; Rodríguez, 2010), rather than lessons focusing on MDMs used to regulate and signpost logical flow in argumentative writing; or focuses on pragmatic classroom management issues (Vargas and Monge, 2014). One study, (Shintani and Aubrey, 2016), stands out, as it employs a kind of reconstruction exercise, perhaps better described as an online collaborative editing/correcting exercise. Somewhat like my study, Shintani and Aubrey (2016) focuses on the appropriate construction of conditional statements. Their study also was among a few which provided a useful classification of conditional statements and where hypothetical premises belonged within that classification.

This study addresses multiple gaps in the literature and points out a potentially important and hitherto unresearched technique for repurposing a teaching tool for better teaching practices, and helping universities reach internally set (Tokyo University of Science, n.d.; 2016) and governmentally mandated goals (Suzuki, 2015b).
2.11. Research questions
The discussion of the literature above, lead me to pose the following research questions:

1. To what extent does completing OTREs and particularly WebSequiturs influence language learners to use selected MDMs in their written arguments?

This question is aimed at understanding the role OTREs and particularly WebSequiturs can play in guiding learners to develop new writing practices, specifically, using premise and conclusion indicators, irrespective of the appropriacy of their use.

2. To what extent does completing OTREs and particularly WebSequiturs guide language learners to use MDMs appropriately in such a way that indicates skilful signalling of argumentative structures in written discourse?

While novice writers may simply interject MDMs awkwardly, without considering the context for their use; or because the MDMs feel familiar and safe; or in an attempt to suggest a logical or causal relationship where none is needed (Leedham and Cai, 2013), skilled writers of argumentative texts use MDMs considered appropriate within their discourse community, and they apply those within structures that may extend across clauses, sentences and paragraphs.

3. What are participants’ perceptions of their experiences with OTREs and particularly WebSequiturs?

It is recognised that the ability to reach meaningful answers to this question is dependent on how metacognitively sophisticated participants are; and on the trust between the researcher and participants, especially concerning the power dynamics operating within their roles as teacher and students. Nevertheless, participants’ expressions of their experience are considered a valuable data source especially where their comments were critical.

2.12. Summary
In this section, I discussed the importance of promoting the use of MDMs used as argument indicators to basic AW instruction. I identified words and phrases used in critical thinking discussions as corollaries of MDM commonly used in AW. I discussed the competing theoretical issues concerning MDMs and I explained my critical stance and rationale. I detailed the kinds of computer mediated communication and noted the particular attributes and affordances of each. I discussed the challenges of AW instruction. I also posited the writing of participants in this study within EFL and CLT. I explained my rationale for also including MDMs categorised as interactional.

I then discussed the history of text-reconstruction and its traditional uses in EFL.
pedagogy, such as teaching vocabulary and grammar. I introduced WebSequiturs and discussed the commonalities that they share with online text reconstruction in general. I also explained their unique pedagogical affordances that make them optimal for teaching argumentative structure and MDM use.

I reviewed research on the teaching of MDMs in AW and on the use of OTREs in teaching the logical flow of general language points in the written discourse. I presented research on the importance of the social constructivist dimension of group work, which, I further showed can be achieved through completing OTREs collaboratively.

The RQs in this study were developed based on gaps in the literature concerning the use of collaboratively completed OTREs aimed at the teaching of the use of MDMs commonly used in AW. Similarly, these gaps led logically to my decisions regarding methodology, which is discussed in the next chapter.
Chapter 3: Methodology: design and methods

3.1. Introduction
In this chapter, I describe the methodology used to answer the research questions (see Section 1.7). In Section 3.2, I describe my theoretical position through an interpretive approach. In Section 3.3, the rationale for the application of action research and mixed-methods in my data collection is explained. I discuss why I chose a mixed-methods approach and my use of both quantitative and qualitative data. In Section 3.5, the research context is described. This includes the location, the department to which the participants belong, the participants’ level of English proficiency, gender, and their background. In Section 3.5.2, I explain the relation between my categorisation of MDMs and Hyland’s taxonomy (2005a). This is followed by a description (see Section 3.6) of the research tools and the research design used in this study, beginning with a discussion of the initial study and how it served as a test run for the various data collection and analytical tools. I give an overview of the data sources and collection instruments and their relationship to the RQs.

In Section 3.11, I discuss the methods used in this study. I write about arrangements made in the first lessons and how the participants were selected. I give an example of how participants’ utterances during the first OTRE sessions in the pilot study led to an improvement in the scheduling of the OTRE sessions in the main study. I present the themes detected in student-student talk from transcripts of OTRE sessions.

In Sections 3.11.3 and 3.11.8, I discuss the content of the forums and chat rooms and how students interacted in each environment as they discussed the issue. In Section 3.10.9 and 3.11.10, the Group Report and the In-Class Rewrite are discussed.

In Section 3.12, I describe the semi-structured interviews and thematic analysis is also discussed in Section 3.15, and in 3.16.1. In Section 3.13, I discuss the stance I took towards ethics, what ethical issues were involved and how I protected participants’ safety and wellbeing, how I handled participants’ data, and how I communicated these precautions to the participants. Lastly, a summary is provided in Section 3.14.

3.2. Theoretical position: An interpretivist approach
As explained in Chapter 2, the participants in this study have witnessed both the advent of the Internet as we know it and the post-economic boom years, both of which have been factors for and against internationalisation. In addition, a scientific community, at a specific university, campus, and department constitutes a culture in its own right. Recognizing these
influences, data is seen as socially constructed, and moreover, unavoidably altered by the
inquiry itself (Hodkinson, 2005). Acknowledging this, I locate learning within the overlap of
social, cognitive and teaching presence (Garrison and Kanuka, 2004). I also take an
interpretivist approach (Gage, 2007) which stresses the importance of knowledge-building
between researcher and participant, and between participants. I view research itself as
deriving, not only from individuals, but also from the culture of which they are situated
temporally (Blaxter et al., 2014).

This study strives for a multi-layered understanding of the variety of ways in which
learners experience and engage with course materials. In asking the research question: ‘What
are participants’ perceptions of their experience with OTREs?’, this study uses a thick
description (Holliday, 2013), incorporating the non-static nature of the educational
experience, and the non-generalisability of the experience of individual participants (Cohen et
al., 2011). The use in Flewitt et al. (2009) and in Gao and Shum (2010) of quantified data
aimed at producing a ‘thick description’ (Pillow, 2010) of limited inferential value outside the
context, but which nevertheless effectively addresses an issue in that localised situation can be
similarly applied to explore the issues in this study. Moreover, other teachers working within
similar contexts may also benefit from this study’s findings.

Both on principle and as a matter of necessity then, this study also avoids grounded
theory (Silverman, 2014), which is based in the idea that the researcher goes into a study with
as little prior knowledge and as few preconceptions about the research context as possible
(Denscombe, 2014). My eight years teaching at TUS, during which time I have become
familiar with the general character and aptitude of the students, as well as having developed,
term by term, the use of those resources which I am studying, precludes an epistemological
framework predicated on an absence of foreknowledge. In a grounded study, the research
should be iterative, as researchers code and classify new data, while continuing to gather more
data repeatedly until the point of ‘saturation’ (Denscombe, 2014, p. 112). As a single
researcher working with participants within one school term, an approach rooted in grounded
theory would require a far greater amount of data than this small action research (defined
below) project could generate and which I am equipped to process within the 15-week term
(Brockmann, 2011). Thus, this study is based in transactional epistemology, which ‘involves
participants in the research project to validate themes’ (Marshall and Rossman, 2016, p. 48).
3.3. Research rationale

3.3.1. Introduction

In this section, I discuss the rationale by which I arrived at the use of action research, mixed-methods and triangulation during data analysis as a means of conducting this study. I then provide a source-by-source explanation for the steps by which data were generated.

3.3.2. Action research

Classrooms are dynamic and complex environments with factors affecting learning within a highly situated group of language learners. Because of its compatibility with the qualitative framework in which I am working, I have chosen to pursue this topic through action research. My approach is participatory and draws upon learners’ insights to shed light on the meaning of the data which they have helped produce. I am engaging in educational inquiry in terms of ‘research as practice’ rather than ‘research on practice’ (McAteer, 2013, p. 48).

In this study, action research was deemed a suitable method for my research goals. First, as I note in Section 3.3.2, action research is well-suited to a mixed-methods approach, which, in this case involved mixing quantitative and qualitative methods. Second, it allowed me the flexibility to respond to unexpected emergent themes that came up during the data analysis.

While action research has advantages, it has also been criticised for too close of a relationship between social values and theory; a loose connection between theory and practice; and too deep of an investment in emancipatory pedagogy, while often involving very small numbers of students/participants in highly contextualised environments (Brydon- Miller et al., 2003). As for the first objection, it is my position that as long a theory is not driven by social values, this convergence is not an issue. For example, the social values of recognizing and welcoming agency converged harmoniously in latter qualitative areas of research, but that did not affect the research process of the first quantitative part of the study.

3.3.3. Advantages of mixed-methods in action research

Action research can facilitate the integration of quantitative data into a qualitative study, as was done in this study. A common way of achieving this synthesis is a mixed-methods approach. Mixed-methods has been called a third paradigm (Denscombe, 2008), as it makes possible forms of data analysis that reveal details that neither quantitative nor qualitative analysis alone can achieve (Boeije et al., 2013). Mixed-methods approach offers pragmatic solutions to the problems raised in combining methods based in different paradigms (Flick,
In this study, I employed mixed-methods first working from quantitative data. The introduction of a simultaneous, complementary, quantitatively driven, variable-oriented component enables integrated naturalistic and statistical examination of patterns within- and across-cases and facilitates the drawing of inferences during the analytic and interpretive phases of the study’ (Evans, et al., 2012). This was very much the case in this study, as I used data expressing trends in individual participant’s MDM-use and used the same data to watch for emergent commonalities in groups of participants. Later qualitative data, such as that produced through semi-structured interviews, could be used to reach a contextualised understanding of the findings derived from the quantitative data.

Quantitative means of measuring participants’ MDM use in terms of baselines and measurable trends were employed. I then used qualitative methods to inquire into the causes of those trends and the reasons why participants had used MDMs as they had.

Acknowledging that controls cannot account for a sufficient number of variables to ensure external validity and reliability (Schofeld, 2007); and that, moreover, teaching does not enjoy a direct relation to learning (Ahmadian and Tavakoli, 2011), my intent then in this small-scale study is to work within an interpretivist paradigm, taking note of Ercikan and Roth (2006) and work from the strengths of both qualitative and quantitative data, rather than follow one and eschew the other.

The primary way in which I achieved this was through triangulation of data, an important component of mixed-methods, which is also highly compatible with action research. The cyclical or iterative nature of action research lends itself to observing and later reflecting upon complex and dynamic learning environments (Luttenberg et al., 2017) from various points of view (Ahmadian and Tavakoli, 2011). Notwithstanding some researchers’ theoretical criticisms of concurrently quantitative and qualitative approaches (Sale et al., 2002), I find the use of triangulation pragmatically suited to the strengths and constraints of this single-researcher small-scale study (Clark and Ivankova, 2016). Other ways of drawing upon both quantitative and qualitative data exist (Flick, 2014). By using these means, I produce data from multiple sources which I later analyse, searching for instances of consistency, contrast, complementarity (Hammond and Wiriyapinit, 2005); or corroboration, elaboration, complementarity, or contradiction (Brannen, 2007). As an example, in Chapter 7 of this thesis, I note that some semi-structured interview responses might have been influenced by arguably leading prompts. The credibility of the responses can be weighed according to whether other data is consistent or contrastive; corroborative or contradictory. In this study, as I later use triangulation to strengthen the claim that OTREs can play a role in
language learners use of MDMs.

In view of the small number of research participants in this study, in Chapter 2, the literature provides examples of similarly sized studies which took steps much like the steps taken in this study to insure validity in its approach, its data collection and data analysis. This research is, likewise, intended to have explanatory value (Hodkinson and Macleod, 2010). The qualities of *trustworthiness* and *credibility* are also critical. Research is trustworthy when it is persuasive because of the means by which the researcher reaches conclusions; it is credible when it is considered to faithfully present what participants have reported experiencing (Clark and Ivankova, 2016).

3.4. Research context
This study is based on data produced during a first-year undergraduate compulsory English course at Tokyo University of Science. The study was conducted in situ amongst a first-year basic English communication class of 29 students of Mechanical Engineering. The class I selected for this study was level *a* from amongst a class stream ranging from *a*-*h*, *a* being the highest in English proficiency. Within that level *a* group, my purposeful selection of participants was based on my interactions with them, as their instructor, and my assessment of their willingness and ability to communicate with me and with each other; and my expectation of their capacity for metalinguistic reflection.

3.4.1. Participants
The participants in the main study were first-year students in the Mechanical Engineering department. The podcasts and OTREs were intended to be used with first-year level-*a* students from any of the twelve departments at TUS. Level-*a* denotes that the course is intended for the students with the highest TOEIC scores (explained below) among their cohort. Because the number of students to be assigned to each of the several sections for that course and because the number of sections for each department change, depending on enrolment for a given year, no specific minimum score is prescribed for *a*-level students. However, in the ME department, for which I have been responsible for some years, the lowest TOEIC scores in an *a*-level cohort tend to be about 540. As for the entrance examination, unlike many liberal arts colleges, TUS sets no minimum level for the score on the English language portion of the exam. Instead that score is aggregated with the other scores for maths and science. A low score in the English therefore could pull down an applicant’s already otherwise low score for the entire entrance exam, but on its own could not
exclude an applicant from admission. The level-a cohort that I taught during the autumn 2019 term was assigned randomly, as it is each year, by the administration. The selection of the Mechanical Engineering department was, therefore, randomly determined. Like first-year students from any of the twelve departments at TUS, most students have written short essays, in Japanese, and in the ki-shoo-ten-ketsu organisation discussed in Chapter 2. They are unlikely to have written more than sentence-long statements in their high school English classes. As this study took place in the second semester, students were expected to have gained competence in writing at the sentence level during their first semester in order to engage in written responses to issues raised in the textbook selected by the instructor of that course. As stated in the syllabus, the first semester Freshman English course is intended to prepare students ‘for further studies at university and lay a foundation for the use of English in the future, at work and in research’ by ‘revising the basic grammar and vocabulary’ students learned at high school ‘and using this in more practical situations (Tokyo University of Science, 2019a). The emphasis in this first semester course is on speaking rather than writing.

As for the selection of the students from within this class, it should be noted that there are very few female Mechanical Engineering students at TUS, and in this class of 27 students there were only two. In the computer terminal room where the class is held there are two columns of long tables (approximately four meters in length) with a centre aisle dividing the room into a left and right side. There are three computers at each table. Therefore, students tend to sit three to each table on the left and right side of the room. There was no designated seating arrangement and class-wide groups were created by free seating until the time at which these triads (groups of three) were set for the term. In the interest of representing female students and male students as equally as possible, I chose the two groups, each with one of the two female students, at the two tables on the left and right in the last, seventh row, near the back of the classroom. There is also a single table near the back of the room directly behind the two rows at which three male students were seated. As I would be coordinating recording equipment and observing these groups visually, I determined that it would be expedient to also select this group to be the third. In this way, I selected three self-assembled groups consisting of two males and one female and a third group consisted of three males. Students were asked if they would be willing to participate in the study and the terms and conditions were explained to them Appendix 14.

To collect data about participants’ previous experience with English study, I created a short questionnaire using Google Docs and sent it to participants using university email. The
questions were either short answer or were arranged using a Likert scale of 1-5, 1 being the lowest and 5 being the highest. The questions were as follows:

1. How much experience did you have with English beyond high school classes?
2. Did you take private English lessons?
3. At what age did your English study begin?
4. Have you lived overseas, and if so, how many years?
5. Have you travelled overseas? If so, when, where and for how long?
6. How skilled of an English communicator are you?
7. How willing or motivated are you to communicate in English?

I also included in the second and third column participants’ recent TOEIC scores and the CEFR equivalents (Educational Testing Service, n.d.). (See Table 3.1)

<table>
<thead>
<tr>
<th>TOEIC</th>
<th>CEFR</th>
<th>Gender</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>745</td>
<td>B2</td>
<td>M</td>
<td>3</td>
<td>None</td>
<td>11</td>
<td>No</td>
<td>New Zealand</td>
</tr>
<tr>
<td>P2</td>
<td>595</td>
<td>B1</td>
<td>M</td>
<td>(NR)</td>
<td>No Response (NR)</td>
<td>[NR]</td>
<td>[NR]</td>
<td>[NR]</td>
</tr>
<tr>
<td>P3</td>
<td>630</td>
<td>B1</td>
<td>M</td>
<td>2</td>
<td>1 year</td>
<td>12</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>P4</td>
<td>575</td>
<td>B1</td>
<td>F</td>
<td>1</td>
<td>None</td>
<td>13</td>
<td>No</td>
<td>Vietnam, Finland, Italy 1 week each</td>
</tr>
<tr>
<td>P5</td>
<td>640</td>
<td>B1</td>
<td>M</td>
<td>2</td>
<td>None</td>
<td>13</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>P6</td>
<td>705</td>
<td>B1</td>
<td>M</td>
<td>5</td>
<td>Father is native speaker</td>
<td>3</td>
<td>3.5 years, CA, U.S.</td>
<td>3.5 years, California</td>
</tr>
<tr>
<td>P7</td>
<td>645</td>
<td>B1</td>
<td>M</td>
<td>3</td>
<td>None</td>
<td>12</td>
<td>No</td>
<td>Junior High / Guam / 3 months</td>
</tr>
<tr>
<td>P8</td>
<td>590</td>
<td>B1</td>
<td>M</td>
<td>3</td>
<td>None</td>
<td>10</td>
<td>No</td>
<td>2019 / Hong Kong / 4 days</td>
</tr>
<tr>
<td>P9</td>
<td>670</td>
<td>B1</td>
<td>F</td>
<td>2</td>
<td>1 year</td>
<td>12</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 3.1 Participants’ past English study experience, etc.

TOEIC (Test of English for International Communication) has a range of scores from 10 to 990. 10-250 ‘basic English proficiency (survival skills)’; 255-400 ‘elementary proficiency very limited communication’; 405-600 ‘elementary proficiency very limited communication’; 605-780 ‘limited working proficiency’; 785-900 ‘working proficiency plus’; 905-990 ‘international professional proficiency’ (Global Exam Blog, n.d.). TOEIC scores are assessed in two groupings, listening and reading; and speaking and writing. However, the speaking and writing portion has only been added in recent years and TUS students do not normally take it. CEFR scores equate with the minimum scores on the listening and writing portion of the TOEIC test, as follows (CEFR/TOEIC): A1/120, A2/225, B1/550, B2/785, C1/945 (Global Exam Blog, n.d.).
3.4.2. MDM categories used in this study and Hyland’s taxonomy

In this first-year second term English course, as stated in the ISE-1 syllabus, students were tasked to explain their opinions, ‘become better able to write short paragraphs’ and ‘to think logically and critically’ (Tokyo University of Science, 2019b). In terms of genre, this might be carried out as a simple student response to a question from the teacher or a question in the textbook, or in the context of a forum or chat room discussion. A Group Written Report (GWR) was also assigned in the second half of the term.

These aims played a role in determining the type of linguistic competence students were tasked to gain. In view of Basturkmen and Randow’s (2014) observation on the signposting of logical flow through MDM use, and Bax et al.’s (2019) highlighting of the role of MDMs in expressing attitude or stance toward a given statement, MDMs were selected for instruction. Lave and Wenger (1991) and (Wenger 1998) further nuanced this selection as I worked towards a selection of MDMs that would be appropriate in a variety of academic and professional contexts, while remaining within the limits of the course description. Collectively these sources pointed towards an approach which emphasised the interactive and reflexive qualities of MDM use.

Interactive metadiscourse is reflexive when the text communicates to the reader about itself. This could be a distinction between a premise and a conclusion, or it could be a reference to content which appears a page earlier or later; or it could clarification which restates a premise in a different way. A text is interactional and reflexive when the text speaks to the reader about himself or herself or takes on readers’ points of view. Often a given MDM operates in both areas.

As I have made the critical choice to follow Hyland’s (2005a) taxonomy, concerning MDMs used interactively and MDMs used interactionally, in this section, I draw parallels between the categories I have used in this study and Hyland’s categories. Although, interactional MDMs were not treated during the OTRE sessions, I have also followed Hyland’s taxonomy in reference to interactional MDMs when their use occurred. Foremost amongst the interactive MDMs covered in the OTRE sessions, and what Fisher (2001) refers to as premise indicators and conclusion indicators, which Hyland calls logical connectives. Except for MDMs used in conditional statements, nearly all premise and conclusion indicators provided to participants in this study (see Table 3.2) are also included in Hyland’s (2005a) list (see Appendix 11).

Although Hyland (1998) places hypothetical premises in the category of hedges, I am not convinced that this is a proper classification. Pragmatically, hedging is not the function of the
hypothetical premise, but rather the function is to make a predictive statement based on reasons. I suggest that the hypothetical premise is an amalgamation, or perhaps an overlap, of the hedge, with its if, and the transition, with its then, which is used as synonymous with conclusion indicators like as a result (see Appendix 11) in the category of transitions. Halliday (2014) places If-then conditional statements within his taxonomy (see Appendix 12) in the category of enhancement, which is a type of extension of a sentence’s or previous clause’s meaning, adding content, usually through a coordinating conjunction. One type of extension is enhancement. Simply stated, enhancement qualifies the existing clause in a variety of ways, including time, place, manner, cause and condition. In the class iv type of enhancement (see Appendix 12), causes (reasons and purposes) (with markers such as because and therefore), and conditionals (including If P then Q and If not P then Q statements) can be found with markers such as if, unless and therefore. In the same section of this table, in which Halliday (2014) places conditional statements are markers such as as a result, which also appears in Hyland’s taxonomy; and in order to, which equates roughly with so as to in Halliday’s class iv of enhancement. An important distinction is that Hyland (1998) is largely descriptive, and based on corpus studies. Moreover, in that study Hyland’s focus is on hedges and he explore how the markers of hedges are used in various areas of science discourse, including hypothetical statements. In my study, the focus, in the latter half of the term, is on hypothetical premises, and my approach is prescriptive, and moreover, constrained by the students’ limited English proficiency. Further to the point, in a personal email correspondence, Dr. Hyland wrote, ‘the lists are not intended to be exhaustive’, and that there would be merit in further research on the ‘different functions of transitions as this is perhaps the least studied feature of metadiscourse in the literature’ (Hyland, 2021). For these reasons, I suggest conditional statements may be more appropriately placed among transitions in Hyland’s taxonomy. It remains to be determined if this is a valid adjustment, but if it were found acceptable, it would simplify discussion of these markers, functions and taxonomies.

Taishukan’s Genius English-Japanese Dictionary with its asterisked-system for denoting whether a given word is taught in high school or university was another primary resource for determining the MDMs to be included in the OTREs. This dictionary and Sanseido’s Wisdom Dictionary are used by high school students to prepare for their university entrance exams. While the Wisdom dictionary uses a five-asterisk ranking system as compared to the four-asterisk system used by the Genius dictionary, in either case, entries with one or two asterisks denote words and phrases which students can be expected to have
studied. Publishers use the word lists derived from the TOEIC (Teaching of English for International Communication) testing organization, which also bases its lists on the COCA corpus for American usage, the BNC corpus for British usage, and the Coxhead Academic Word List (AWL) (Anon., 2021). Dictionary publishers also refer to published entrance exams from the preceding ten years (Anon., 2021).

I therefore conclude that the selection of MDMs, which I chose to present through the OTREs, is validated by its correspondence with both Hyland (2005a), the COCA and AWL lists and the asterisk systems of the two most commonly used dictionaries among Japanese students.

I also considered what categories lent themselves to dual purposes of signalling the writer’s level of certainty or attitude towards to topic while simultaneously signalling premises or conclusions through implicitly connective means. Categories such as additional reasons and levels of conviction (epistemic markers in Hyland (2005a) were added, and phrases like second, finally, in addition, moreover, and in my opinion, I think and perhaps were included. To these, I added clarification, e.g., what I mean is and creating focus, e.g., the main point is and it should not be forgotten that. Generalizations and exceptions to the rule, e.g., for the most part and but in this case were added as premise indicators. Also, consideration of other relevant facts, e.g., that may be true, but was added as a conclusion indicator. All these were later sub-grouped under the class of implicitly connective argument indicators.

Yossatorn et al. (2022) provides a categorisation of four types of conditional statements. This study is concerned with only the first two of these. The first is the factual conditional, which in this study I refer to simply as a conditional statement. The structure of a factual conditional is: If + subj.+ present tense verb, subj. + present tense verb. It typically is used to make true propositional statements or statements intended to be understood as true (Yossatorn, et al., 2022). An example is: Whenever I eat at that restaurant, I get sick.

Yossatorn et al. (2022) supports this inclusion of what they term the future predictive as the second of the four types of conditional statements. The structure of a future predictive is: If + subj. + present tense verb, subj. + will/can/might/may/should/must/have to + base form verb (Yossatorn, et al., 2022). In this study, I refer to the future predictive as a hypothetical premise. I chose to do this because of the context of science university in which this study was done. Hypothesis is a word which students at TUS respond to and identify with as part of their EFL discourse community. An example of a future predictive conditional statement, or, henceforth, hypothetical premise, is: if we do x then y will happen; or in more concrete terms:
If I eat at that restaurant, I will get sick. In this form, a predictive claim is made which can be tested. Hyland (1998) includes the specific MDMs used in those forms as presented in Table 3.2. There, Hyland’s use is altered a small amount. That is because he presents them as forms of hedging rather than as forms to be used to make strong claims. Hedging is not listed as a first- or second-year skill in TUS’ English curriculum. Halliday (2014) however, lists if... then and unless statements in his category of enhancement as a causal-conditional.

Hyland’s inclusion of these forms and their strong asterisk ranking in the Genius and Wisdom student dictionaries, as well as the student dictionaries published by Taishukan and Sanseido, suggest that these forms merit general instruction at a science university and within an EFL curriculum that stresses the need to make arguments and support claims.

In Table 3.2, a sample of the MDMs used in the OTREs in this study are presented. The first four blocks show premise and conclusion indicators. The first two are explicitly connected and the second two are implicitly connected. The fifth block is concerned with conditional statements and hypothetical premises. The first column of each block shows Hyland’s taxonomical term for the MDM category being drawn upon.

The last two main columns to the right group MDMs into casual and spoken communication and formal and written communication. The table displays tokens as spoken or written. A somewhat greater level of formality is designed into this distinction with the written tokens being somewhat more formal, though not greatly so. Speakers may speak in a way that is closer to writing than the speech of other speakers. Similarly, writers may write in a style that is closer to speech than to formal or written communication. As discussed in Chapter 2, genre can play a large role in any interlocutor’s language use along this spectrum from one communication to another. Again, to confirm reliability, three raters assessed the MDMs in Table 3.2 in terms of whether the phrases belonged in the third or fourth column. The interrater reliability is discussed in Section 4.1.6.
Table 3.2. Sample of premise and conclusion related MDMs.

<table>
<thead>
<tr>
<th>Explicitly connective premise indicators</th>
<th>Hyland (2005a)</th>
<th>casual/spoken</th>
<th>formal / written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition Markers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>since</td>
<td>because</td>
<td>for (archaic)</td>
</tr>
<tr>
<td></td>
<td>the reason why</td>
<td>the reason is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>By the same token</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explicitly connective conclusion indicators</th>
<th>Hyland (2005a)</th>
<th>casual/spoken</th>
<th>formal / written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transitions</td>
<td>so</td>
<td>then</td>
<td>therefore</td>
</tr>
<tr>
<td>Code Glosses</td>
<td>That means</td>
<td>for that reason</td>
<td>accordingly</td>
</tr>
<tr>
<td></td>
<td>thanks to x</td>
<td>given that</td>
<td>on the basis of</td>
</tr>
<tr>
<td>Hedges</td>
<td>x may be so,</td>
<td>on one hand</td>
<td>While it is true that</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implicitly connective premise indicators</th>
<th>Hyland (2005a)</th>
<th>casual/spoken</th>
<th>formal / written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing</td>
<td>first</td>
<td>to begin with</td>
<td>next</td>
</tr>
<tr>
<td></td>
<td>also</td>
<td>in fact</td>
<td>furthermore</td>
</tr>
<tr>
<td>Evidentials</td>
<td>we know that</td>
<td>it is common knowledge that</td>
<td></td>
</tr>
<tr>
<td></td>
<td>X said/told me that</td>
<td>X reported/notes/writes that</td>
<td></td>
</tr>
<tr>
<td>Code Glosses</td>
<td>like</td>
<td>for instance</td>
<td>e.g.</td>
</tr>
<tr>
<td></td>
<td>to put another way</td>
<td>i.e.</td>
<td>For example</td>
</tr>
<tr>
<td>Boosters/Hedges</td>
<td>the results show</td>
<td>the data indicates</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Implicitly connective conclusion indicators</th>
<th>Hyland (2005a)</th>
<th>casual/spoken</th>
<th>formal / written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame Markers (Label Stages)</td>
<td>to sum up</td>
<td>in conclusion</td>
<td>to summarise</td>
</tr>
<tr>
<td></td>
<td>in short</td>
<td>to wrap up</td>
<td>in summary</td>
</tr>
<tr>
<td>Code Glosses</td>
<td>that means</td>
<td>indicates that</td>
<td>suggests that</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional Statements and Hypothetical Premises</th>
<th>Hyland (1998)</th>
<th>casual/spoken</th>
<th>formal / written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedges</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conditional Statements</td>
<td>whenever; wherever; however; no matter what/where/how/etc.</td>
<td>In the case that x, y</td>
<td></td>
</tr>
<tr>
<td>Hypothetical Premises</td>
<td>If x . . . then y will z.</td>
<td>If x . . . then y will z.</td>
<td></td>
</tr>
<tr>
<td>As long as x keeps doing y then z (will)</td>
<td>For the duration that x, y will</td>
<td></td>
<td></td>
</tr>
<tr>
<td>If x doesn’t y then z is going to</td>
<td>Unless x, then y will/won’t z</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4.3. Interactive and interactional MDMs

In Hyland’s (2010) taxonomy, MDMs are divided into interactive and interactional groups. While the former is concerned with the flow of information, the latter is concerned with expressing the writer’s attitude towards the propositional statements or towards the argument itself. For the purpose of this study, relevant categories in the interactive group include
mainly transitions (Hyland, 2005a), such as because and therefore, which act as premise and conclusion indicators to mediate the transition of information from one clause in a sentence to another; or one sentence to the next. Because Hyland’s focus is different from mine, he groups implicitly connective MDMs across multiple interactional categories. MDMs that I have grouped as implicitly connective markers include phrases like summing up or to conclude, which often precede a listing of premises or a concluding statement, are classified in Hyland’s taxonomy as frame markers, as they indicate what part of a text follows the marker. The category which I have called clarification, i.e., phrases that assist the reader in understanding the main content of a text are included within Hyland’s code glosses. These include phrases like means that. Also included in code glosses are what I have categorised as examples and illustrations with the phrases like for example, such as, and in other words. Because no research is required in this first-year course, the use of evidentials such as according to Smith or Jones writes that were not considered and no instances appeared during this study’s duration. For similar reasons, endophoric markers such as mentioned earlier were not treated. Most likely, because of the small size of forum posts and chat room exchanges, endophoric markers referencing previous or subsequent sections of text passages did not appear.

Other markers in the interactional grouping, relevant categories in Hyland’s taxonomy include hedges, boosters, attitude markers and engagement markers (Hyland, 2010). Hedges indicate that the writer does not fully commit to a claim he or she is making. Conversely, boosters indicate that the writer has a high level of confidence in a given claim. My experience as a teacher at TUS led me to the informed discernment that students could more easily grasp the language function of hedges and boosters if the two categories were grouped as one category related to low and high levels of confidence. Moreover, as a researcher, I was not concerned as to whether participants were hedging or boosting their statements, but whether they employing the phrases associated with either function in order to signal premises and conclusions. I therefore used the phrase Levels of Conviction to group hedges and boosters into one more useful category. Agreement and disagreement equates with Hyland’s interactionally categorised attitude markers, which indicate the writer’s attitude towards a given claim. Hyland’s engagement markers concern the relationship the writer cultivates with the reader. Phrases like the reader will note that often precedes premises, and phrases like you can clearly see that often precede conclusions.

Acknowledgement of the dual roles of regulation of information flow and the self-expression of the writer (Hyland, 2005b) leads me to align this study with Hyland’s
taxonomy. Reflexivity exists in both interactive and interactional text passages and as I show in Chapters 5 and 6, this reflexivity was exploited by participants in ways other than the assigned use of interactive, explicitly connective, MDMs. Participants also frequently used interactional, implicitly, connective MDMs to indicate argument indicators.

3.5. Research design
In this section I discuss the research design I used to carry out this study. I detail how I gathered various kinds of data and how that data links back to my research questions. Because of the mixed-methods approach and the number of data sources, it will be helpful to view the flow chart (see Figure 3.1) below. In the top half of the chart, the reader will find, in the four horizontal paths leading left to right, three data sources: OTRE talk transcripts, forum written output and chat room written output. Each of those sources is connected by a vertical downward pointing arrow to one or two corresponding types of analysis. To these are added one more data source: the semi-structured interviews conducted after all the other data had been collected. That also follows a downward path to analysis. All four analyses feed into a triangulation where the path ends with Conclusions.
Figure 3.1: This study’s methodology in flow chart form
3.6. Initial study

Prior to embarking on the main study, I conducted a pilot study to test my research instruments. In 2017, I conducted the pilot study in my Integrated Skills in English (ISE-A) ME-a class. The -A following the ISE indicates that this is the first of three ISE courses, which follow Freshman English in the first semester, ISE- A, B and C, offered to students in their second, third and fourth semester of their first two undergraduate years. ME is used to denote the Department of Mechanical Engineering. The class consisted of 25 students. The class was designated as section-a in a ranking of eight sections streamed via the university-required English proficiency assessment scores. Section-a students, such as the ones in this study, may be described as students who are at the top of their department for that year in terms of their scores on the placement test.

Three students (male n = 2, female n = 1) participated voluntarily and matched the participants of the main study in age, L1, class level and English proficiency. The group of three itself was consistent with two of the three groups in the main study in terms of gender distribution and with all three groups in the main study in terms of English proficiency as measured by the university streaming placement exam.

The pilot had two aims. The first was to determine whether participants who had worked through the course design would further develop their frequency of MDM use and refine their appropriacy of MDM use. Because this level of instruction was not considered to speak to the particular needs of many of the mixed-level groups in sections b-h, this content was reserved for students in the a-level section.

The second aim was to pilot the use of the data collection tools and confirm that the tools operated smoothly, the schedule I had set out was manageable, and to note any unforeseen conflicts or problems. It was also an opportunity to practice the use of recording equipment and the safe storage of the recordings which were made. Ensuring that recording devices could pick up the utterances of all three students in each group and that ambient sounds such as other students, squeaking chairs and moving air from room heating would not obscure those utterances was one such concern. Another was the systematic labelling of audio files so that they could be placed in chronological order with each group’s recordings kept separate from one another were some basic practical matters for which the pilot provided an opportunity to work out. Other matters that were worked out at this time was the later thematic analysis of the student-student talk during the OTREs and semi-structured interviews and an assessment rubric of participants’ written output.
3.7. Findings of initial study

The findings of the pilot study showed a clear increase in the use of MDMs presented in the podcast and practiced in the OTREs. It also showed an improvement in the appropriacy of MDM use. Thematic analysis on the student-student talk during the OTREs indicated that participants had gained an increased metalinguistic awareness of the forms and functions of the MDMs. Thematic analysis of the semi-structured interviews indicated that the OTREs had played an important role in participants’ increased frequency of MDM use and in the improved appropriacy of that use. Interview responses also indicated that students preferred learning collaboratively through OTRE sessions over working individually and over working through paper-based textbook exercises.

Just as action research design itself is reflective (Luttenberg et al., 2017) and responds to emergent qualities that arise in specific educational contexts (Fouché and Chubb, 2017), so did my pilot study reflect upon and respond to unexpected occurrences to improve upon itself in the subsequent study (see Section 3.8.1). However, between pilot and main study, no substantive changes in the research design were needed. One minor adjustment was made in the number of OTREs per session and the length of the first two sessions. Participants’ utterances indicated fatigue. In response, during the main study, I increased the number of sessions by one and redistributed the OTREs. A more important development which came as a result of data analysis in the pilot was a further refinement of several of the communicative functions into which the MDMs were distributed. This helped me to better understand and convey to readers how identical MDMs were used to express different communications.

Based on the findings in the pilot, I was motivated to continue with a somewhat larger, but still small-scale iteration of this study, in order to further examine the efficacy of OTREs and produce more robust data to apply in response to my research questions.

3.7.1. Data sources and collection instruments

In this study I utilised various means of producing quantitative and qualitative data. Quantitative data consisted of the total output from online forums, chat rooms and a final group writing assignment. Qualitative data included transcripts from four audio recorded OTRE sessions; and transcripts from audio recorded interviews, as well as texts written as a Group Report (see Table 3.3).
Table 3.3: Four data sources

Table 3.4 shows data sources, by source of that data, the codes used to express that data, and the approximate amount of data. During interviews, participants responded mostly in English. When a participant seemed to be struggling to express their thoughts in English, I encouraged them to speak in English and attempted to confirm my correct comprehension intermittently, as they responded. Later, I reviewed the transcript and listened to the audio record with a native Japanese speaker colleague.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Produced by</th>
<th>Language</th>
<th>Approximate amount of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. OTRE talk transcripts</td>
<td>One triad of students</td>
<td>Numerical</td>
<td>Several lines of records</td>
</tr>
<tr>
<td>2. Forum written output</td>
<td>One triad of students</td>
<td>Japanese / English (code switching)</td>
<td>Roughly 950 lines of records</td>
</tr>
<tr>
<td>3. Chat Room written output</td>
<td>One triad of students</td>
<td>English</td>
<td>Variable, depending on student output</td>
</tr>
<tr>
<td>4. Semi-structured interviews</td>
<td>One triad of students</td>
<td>English</td>
<td>A page and a half of text</td>
</tr>
<tr>
<td>5. Individual student</td>
<td>Individual student</td>
<td>English</td>
<td>A paragraph to a page and a half of text</td>
</tr>
<tr>
<td>6. Assessment of the amount of DM use</td>
<td>Instructor</td>
<td>Numerical</td>
<td>Several lines of records</td>
</tr>
<tr>
<td>7. Assessment of the appropriateness of DM use</td>
<td>Instructor</td>
<td>Numerical / English</td>
<td>Several rows of digits and a cumulative total</td>
</tr>
<tr>
<td>8. Semi-structured one-on-one interviews</td>
<td>One triad of students and instructor (myself)</td>
<td>Japanese / English (code switching)</td>
<td>Approximately 60 minutes of interview transcripts</td>
</tr>
</tbody>
</table>

Table 3.4: Research tools as data sources
3.7.2. Data collection and research questions

In this section I include two tables. The first (Table 3.5) shows research questions and the data sources used to answer them.

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what extent does completing the OTREs influence language learners to use selected MDMs with greater frequency in written arguments?</td>
<td>• Audio recorded participant talk during OTRE triad work</td>
</tr>
<tr>
<td></td>
<td>• Participants written output in online forums, chat rooms and in a group project writing assignment</td>
</tr>
<tr>
<td></td>
<td>• Observation of behavior during OTRE triad work</td>
</tr>
<tr>
<td></td>
<td>• Semi-structured interviews</td>
</tr>
<tr>
<td>2. To what extent does completing the OTREs guide language learners to use selected MDMs more appropriately in written arguments in a way that demonstrates skillful signaling of argumentative structure?</td>
<td>• Interrater assessment</td>
</tr>
<tr>
<td></td>
<td>• Participants written output in online forums, chat rooms and in a group project writing assignment</td>
</tr>
<tr>
<td></td>
<td>• Semi-structured interviews</td>
</tr>
<tr>
<td>3. What are participants’ perceptions of their experiences completing OTREs</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.5: RQs and data sources

Table 3.6 presents a week-by-week schedule of the 15-week school term and within each week it shows: (1) class activities relevant to the research; (2) the category, or communicative function(s) introduced that week; the topic discussed in the forum current at that time; and (4) the research purpose. By communicative function, I mean, following Martin and Rose (2013), what it is that a speaker or writer hopes to achieve or believes will be accomplished through the use of a given MDM. Table 3.6 shows, chronologically, how forums, chat room, written reports were distributed concurrently with podcasted instructional content and textbook units were used as a basis for online discussion subsequent to OTRE sessions. It also notes according to Weeks 1-15, what activity was assigned or completed, the category of MDMs that were fronted during that time, and the topic and the purpose of each activity.
<table>
<thead>
<tr>
<th>Week</th>
<th>Research Relevant Class Activity</th>
<th>MDM Category</th>
<th>Discussion Topic</th>
<th>Research Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1 (9/13-9/19)</td>
<td>Assigned: Critically Minded Podcast (CMP) Episode 1 <em>Introductions</em>; Forum Discussion 1 <em>To Tell the Truth</em></td>
<td>None</td>
<td>To Tell the Truth</td>
<td>Establishing a baseline of MDM use</td>
</tr>
<tr>
<td>Week 2 (9/20-9/26)</td>
<td>Assigned: Critically Minded Podcast CMP Episode 2 and 3 <em>Organizing an Argument 1 and 2</em> / Dialogically Minded Podcast (DMP) Episode 1 <em>Introductions</em> and Episode 2 <em>Expressing Opinions and Agreeing and Disagreeing</em> CMP Episode 4 <em>Validity and Soundness/DMP Episode 3 Levels of Conviction</em></td>
<td>Explicitly connective premise and conclusion indicators</td>
<td>To Tell the Truth (Closed on 9/27)</td>
<td>Establishing a baseline of MDM use; using MDMs covered in the podcast</td>
</tr>
<tr>
<td>Week 3 (9/27-10/3)</td>
<td>TRE Session I Textbook work: Fat shaming CMP Episode 5 *Modus Ponens/Modus Tollens/DMP Episode 4 Creating Focus Forum Discussion 2 <em>Fat Shaming</em></td>
<td>Explicitly connective premise and conclusion indicators; implicitly connective premise and conclusion indicators: additional reasons; asking for reasons</td>
<td>Fat shaming</td>
<td>Using MDMs covered in the podcast</td>
</tr>
<tr>
<td>Week 5 (10/11-10/17)</td>
<td>TRE Session II Assigned: CMP Episode 7 <em>Side-by-Side Premise arguments and Chain Premise Arguments</em></td>
<td>Explicitly connective premise and conclusion indicators; implicitly connective premise and conclusion indicators: additional reasons; asking for reasons</td>
<td>Fat shaming</td>
<td>Using MDMs covered in class</td>
</tr>
<tr>
<td>Week 6 (10/18-10/24)</td>
<td>TRE Session 3 Assigned: DMP Episode 7 <em>Clarification and Confirmation</em> Forum Discussion 3: <em>Cups and Coins 1</em></td>
<td>Conditional Statements and Hypothetical Premises Premises that take the form of comparisons</td>
<td>Cups and Coins Part 1</td>
<td>Using MDMs covered in class</td>
</tr>
<tr>
<td>Week 7 (10/25--[no class 11/1]-11/7)</td>
<td>ChatRoom I: Fat Shaming Assigned: DMP Episode 9 <em>Comparison and Contrast; and Inclusion and Exclusion</em></td>
<td>Conditional Statements and Hypothetical Premises Comparisons as premises; and <em>Inclusion and Exclusion</em> Independent and Causally linked Premises</td>
<td>Fat shaming</td>
<td>Using MDMs covered in class</td>
</tr>
</tbody>
</table>

Table 3.6: 15-week schedule of MDM instruction
<table>
<thead>
<tr>
<th>Week</th>
<th>Research Relevant Class Activity</th>
<th>MDM Category</th>
<th>Discussion Topic</th>
<th>Research Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 8 (11/8-11/14)</td>
<td>TRE Session 4 Assigned: CMP Episode 8 Conditional Statements and Hypothetical Premises/DMP Episode 10 Examples and Illustrations Forum IV: Cups and Coins 2</td>
<td>Conditional statements and hypothetical premises Premises that take the form of example</td>
<td>Cups and Coins 1 and 2</td>
<td>Co-constructing understandings of participants’ experience with the TREs</td>
</tr>
<tr>
<td>Week 9 (11/15-9/21)</td>
<td>TRE Session 5 Assigned: CMP Episode 8.2 Conditional Statements and Hypothetical Premises</td>
<td>Conditional Statements and Hypothetical Premises</td>
<td>Cups and Coins 1 and 2</td>
<td>Co-constructing understandings of participants’ experience with the TREs</td>
</tr>
<tr>
<td>Week 10 (11/22-11/28)</td>
<td>(1) Traditional Classroom: Cups and Coins data gathering Part 1 (2) Textbook work: Blood Type Personality Profiling Assigned: (1) CMP Episode 9 and 10 Major and Minor Premises 1 and 2 (2) Forum V Blood Type Personality Profiling</td>
<td>Major and Minor Premises Inclusion and Exclusion</td>
<td>Cups and Coins 1 and 2</td>
<td>Co-constructing understandings of participants’ experience with the TREs</td>
</tr>
<tr>
<td>Week 11 (11/29-12/5)</td>
<td>(1) Traditional Classroom: Cups and Coins data gathering Part 2</td>
<td>None</td>
<td>Cups and Coins 2 Blood Type Profiling</td>
<td>Co-constructing understandings of participants’ experience with the TREs</td>
</tr>
<tr>
<td>Week 12 (12/6-12/12)</td>
<td>(1) Teacher led, shared online MS Word document generation (2) Group Work on the Group Written Report</td>
<td>Cumulative</td>
<td>Cups and Coins 2 Blood Type Profiling</td>
<td>Generating output to later be analysed</td>
</tr>
<tr>
<td>Week 13 (12/13-10/24)</td>
<td>Group Work on the Group Written Report</td>
<td>Cumulative</td>
<td>Cups and Coins 2 Blood Type Profiling</td>
<td>Generating output to later be analysed</td>
</tr>
<tr>
<td>Week 14 (12/20-1/9)</td>
<td>Chat Room II Blood Type Personality Profiling</td>
<td>Cumulative</td>
<td>Cups and Coins 2 Blood Type Profiling</td>
<td>Generating output to later be analysed</td>
</tr>
<tr>
<td>Week 15 (1/24)</td>
<td>In-Class Re-Write From Memory</td>
<td>Cumulative</td>
<td>Cups and Coins 1 and 2</td>
<td>Generating output to later be analysed</td>
</tr>
</tbody>
</table>

Table 3.6: cont.
3.7.3. Master MDM List

The Master MDM List (see Appendix 13) is a list of all the MDMs which were taught during the 15-week course. Because the course in which this study was conducted requires instruction over a range of MDMs broader than the scope of premise and conclusion indicators, the Master MDM List contains many MDMs not treated in the OTREs. The full list is included to provide a sense of how the MDMs within the scope of this study are situated within a totality of MDMs which will later be seen in examples of written output.

The contents of this list were confirmed as based in sound EFL practice discussed in Section 2.3.3. First, the items on the list were found to coincide entirely with Hyland’s (2005a) list of MDMs. They were also intended to be consistent with how certain language is used within the target discourse community (Hyland, 2017), here, first-year university level student writers. The MDMs included were intended to support students’ efforts to participate in a ‘community of practice’ (Lave and Wenger, 1991), become competent writers in a given genre (Schmied, 2011) capable of creating relevant structures (Connor, 1990). Basing the MDM selection on legitimate practices which would resonate as consistent to what students might have encountered in other related readings helped to ensure that participants’ later output would lead to valid interpretations.

This Master List was provided to students on their university webpage called Learning Environment of Tokyo University of Science (LETUS), described below. This study deals primarily with the MDMs in Set I, i.e., the categories of premises and conclusions. Agreement and Disagreement is another such category. Phrases such as *I agree that*, again, typically precede premises. The category of Additional Reasons is another example of non-logical connectives. Phrases like *secondly*, and *furthermore* do not strongly imply logical relationships in the same way *because* and *therefore* do, but rather the inclusion of more information, which may be used as a premise. Without the presence of these connectives, the onus is upon the reader to infer the logical flow. I will present examples from participants’ writing in Chapter 6.

This study also examines participants’ use of phrases used to express conditional statements and hypothetical premises. At a science university where students must write papers which discuss conditions and outcomes and which advance hypotheses, involving hypothetical premises, it is especially important for students to have a firm grasp on these forms. These include general conditional statements involving phrases such as *When/Whenever x happens then y happens (as a result)*, phrases involving hypothetical
premises, such as *If we do x . . . then y will/will not happen*, or *Unless we do x then y will/will not happen*.

![Table of Premise and Conclusion Indicators](image)

Figure 3.2: Section I of the Master MDM List: premises and conclusions

Figure 3.3 presents MDMs related to conditional statements and hypothetical premises, which are tasked for use by students during the Group Project (see Section 3.7.6) involving the Monty Hall inquiry into probability and later during their In-Class Rewrite (see Section 3.13).
As discussed in Chapter 2, the use of conditional statements and hypothetical premises is also community-based and as part of Wenger’s ‘shared repertoire’ (1998, pp. 73) of styles. The rules defining the parameters of style optimally in line expert writing within students’ target ‘community of practice’ (Lave and Wenger, 1991), where, ideally, novice writers working within a ‘network of members’ develop skills that help them become competent writers in a given genre (Schmied, 2011, p. 5). Accordingly, instruction in argumentative writing should involve teaching relevant structures (Connor, 1990).

3.7.4. Podcasts and research considerations

Based on the considerations noted in Section 2.10, I determined that the podcast format was appropriate for my students and for the delivery of the instructional content. I also considered the merits and possible weaknesses podcasts could play in my research design (See Chapter 2.8).

These scripted podcast episodes offer lectures on basic critical thinking with consistent vocal delivery (Gann and Bufton, 2012a; 2012b). The order of the series’ 12 episodes is placed in lockstep with the selected communicative functions, and are titled as follows:

Episode 1: Introductions
Episode 2: Organizing and Argument 1
Episode 3: Organizing and Argument 2 Episode 4: Validity and Soundness
Episode 5: Modus Ponens, Modus Tollens
Episode 6: Hidden Premises
Episode 7: Side-By-Side Premise Arguments and Chain Premise Arguments
Episode 8: Conditional Statements and Hypothetical Premises 1
Episode 9: Major and Minor Premises 1
The episodes with which this study is primarily concerned are Episodes 2 and 3 (see Figure 3.4) in the first half of the term, and Episode 8 in the latter half of the term. The explicit instruction on premise and conclusion indicators in Episodes 2 and 3 were reinforced in three separate OTRE Sessions 1, 2 and 3; and the instruction on conditional statements and hypothetical premises in Episode 8 (see Figure 3.5) was later reinforced in two further OTRE sessions; 4 and 5.

Figure 3.4: Critically Minded Podcast episodes on argument organisation

Figure 3.5: Critically Minded Podcast episodes on conditional statements and hypothetical premises

The Critically Minded Podcast episode sequence was determined by the order in which features of argument are presented in various books on critical thinking instruction, chiefest among these being premises and conclusions (Fisher, 2001).

These podcast episodes guarantee consistent delivery of instruction. Moreover, the instruction is carefully scripted and recorded so that the instruction is given at a speed,
volume, clarity, and level of lexical complexity that facilitates maximum accessibility and comprehension by the participants. This ensures that all participants have equal opportunity to the best instruction on MDM use that can be provided prior to the OTRE sessions. Thus, the validity and reliability of the conclusions reached when addressing RQ1 and RQ2 is supported.

3.7.5. WebSequiturs

TREs belong to a larger family of text manipulation exercises. These include deletion, insertion, substitution, and reordering (Johns and Lixun, 1999) and can be described in more detail as ‘scrambled elements (chunks, sentences, or paragraphs); sentence-boundary identification; crosswords; hangman or concentration-type word matching or supplying’ (Cobb and Stevens, 1996).

Thornbury (1997) draws a distinction between text reformation and text reconstruction. The former could take the asynchronous form seen in Shintani and Aubrey (2016) where teacher offered online corrective feedback. Here, learners ‘attend to linguistic features of input they are exposed to’ (Thornbury, 1997, p. 326). Conversely, with reconstruction, learners are presented with a correct model and then work through reconstruction of parts of a text, during which time they ‘notice the gap’ (Thornbury, 1997, p. 326), and adjust their interlanguage accordingly.

Online text reconstruction exercises retain most of the merits of paper-based ones as well as offering added benefits only possible in digital format. With online OTREs, teachers can now quickly create and edit texts; and students can maximise the amount of class time spent focussing on the study of specific language problems (Rodríguez, 2010) and receive immediate feedback (Davies et al., 2012; Uschi, 2003). Dudeney, Hockly, Sharma et al. (2008) argue that Hot Potatoes exercises should be deeply integrated into English instruction on a widespread level. Uschi (2003) asserts that teacher-made Hot Potatoes exercises can be made which are equal to those available commercially, indicating suitability for widespread adoption. Wahba, Zeinab, and Taha (2017) describe Hot Potatoes applications as capable of developing students’ metalinguistic awareness and a metacognitive awareness of their own learning process.

Known by their trade name, WebSequiturs, are a form of online text-reconstruction exercises, which involve segments of text passages presented via computer monitor in a multiple-choice format. Students may be given a short passage to study prior to the activity; or, as in the case of this study, they may be provided with a cue or related text in the field
above the multiple-choice selection area. Students are asked to reconstruct the text, from beginning to end, from a selection of sentences or sentence fragments presented at random on a computer monitor. They then click on one of the choices offered to recreate a correctly ordered coherent text passage with the same meaning as the original text.

The specific and unique style of text reconstruction provided by WebSequiturs has pedagogical affordances useful in teaching AW. WebSequiturs can be designed for the purpose of teaching the use of MDMs in AW. Learners must first recognise the indicator and know whether it belongs on the front end or back end of an argument. Because learners cannot simply scan ahead to know what follows, they are placed in a position much like that of collaborative writing with the computer. If the format were simply a ‘drop down and select’ type, then the learner would no longer be limited in that way and would no longer be in the position of seeming to co-construct the text, but merely answering questions. That inability to look forward, except through guessing and hypothesis testing (Goodman, 1995), creates a sense of temporality similar to that of writing. This sense of temporality also creates a sense of collaborative knowledge construction, as the end of the text seems unwritten until the learner and the computer co-author it. Also, ‘drag-and-drop’ or ‘drop-down’ type exercises with much of the text left blank would be too open-ended and the learner would have too many possible choices. Moreover, these types of exercises are typically set up for short one or two-word answers, and not long phrases. However, WebSequiturs share many commonalities with other kinds of OTREs. For example, they are viewed on a computer display, but can also be accessed via smartphone. They may be completed by using a mouse, trackpad, or touchscreen. Provided the display can be seen by all users they can be worked on collaboratively. Like most OTREs, feedback to the user’s decisions is provided immediately.

3.7.6. WebSequitur generation

In this study, I have used the term OTRE to refer to online text reconstruction exercises in general and have used the trademark name WebSequiturs to refer to the OTREs used in this study when I discuss these exercises’ specific qualities.

WebSequiturs are online OTREs which are generated using an application provided by Creative Technologies, which is downloaded onto a user’s computer. Once generated and posted online, these WebSequiturs can then be accessed by anyone with the link. WebSequiturs present segments of text passages in a multiple-choice format on a computer monitor.

To understand how these exercises are intended to work, it is helpful to see how they
are generated. The WebSequitur generator, which can be downloaded for free through the TexToys website, appears as shown below. A text is cut and pasted into the field and then breaks in the sentences or paragraphs are created by using the Enter key; or the text can be similarly written in any text document and breaks can be placed at that time.

When an exercise like the one below is generated, it appears as in Figure 3.6 and in its completed form as in Figure 3.7. I provided a nearly identical argument in the box above the workspace, but with the logical flow running in the direction opposite to that of the fragmented one tasked to be reconstructed. A function under the File tab provides this feature.

![Figure 3.6: A WebSequitur being constructed in the Sequitur application](image-url)
Figure 3.7: The WebSequitur from above in its completed online form

Figure 3.8: WebSequitur as it appears after completion

In Figure 3.9, in a longer passage, the breaks appear before and after the indicators for example. The student should see that researchers who have cannot run directly into proved that photographs that appear . . . due to the comma preceding proved. Likewise, flying saucers are hoaxes cannot run directly into small private airplanes without the intervening or, or, in this case double-or clause. Drawing the students’ attention to the conclusion indicator, I fragmented it between For and this reason. Unless students knew that For this reason is an indicator, they were unlikely to make the correct selection. For those who make the correct selection, the use of the phrase as a conclusion indicator is further reinforced. Continuing
through the exercise, *nor a good reason to believe* may serve as a distractor until the appropriate time comes to select it, when only two possible selections remain. The *to believe* leads into *that* to complete the conclusion and the exercise.

In Figure 3.10, the difficulty with using WebSequiturs to teach the difference between conditional statements and hypothetical premises can be explained and the method for overcoming that difficulty can be demonstrated. Because the two constructions are so similar, both in the possible breakpoints and in the MDMs commonly used, the antecedent of one hypothetical premise could easily attach to the consequent of another. For this reason, cues other than features associated strictly with conditional breakpoints were introduced. The OTRE, *Mother and Aunt Beth* (see Figure 3.10) can be seen under construction within the WebSequitur generating application.
The phrase *even though* provides contrast between Mother’s beliefs and the lack of evidence. The focus on *evidence* is then provided by the break followed by *to support*. The next *even* is then daisy-chained from the preceding *even* in *She / even says*.

The first conditional statement appears with *If we don’t agree with her, she becomes angry* with the break again before the comma to teach that If clauses are normally followed by independent clauses preceded by a comma. Then there are several breaks unrelated to conditional statements, as with the case of *because my aunt . . .*, which essentially adds new information. This is followed by the second conditional statement, *Whenever we try to talk to her about it, she says that we are against her*, with the break, again, before the comma. After that, there is contrastive *however* and an *only*, which further describes the situation by defining perceived boundaries of Mother’s behaviour. Then, there is the conclusion indicator, *So in So, what / we have decided to do [ . . . ].* The conclusion in this case is what they have decided to do in order to test their hypothesis. *So*, therefore does not come as close to the end of the passage as it normally would in a simple premise-premise-conclusion argument, but is embedded in the narrative. The phrase, *Our idea* prefaces the hypothetical premise, *if we hold a family gathering at a location that mother doesn’t think of as her territory, then she will be less likely to speak so rudely*. The paragraph ends with a conditional statement stating that *By*
doing, $x$ we will be able to do $y$. Finally, as with all the other WebSequiturs, I place the final period after the last line break so that the third to last multiple selection will still require deliberation.

In this way, the risk of losing the focus on the structure of conditional statements and hypothetical premises is avoided through the careful placement of other clause-related breaks, rather than grammatical-or vocabulary-based ones. These cues serve to provide sufficient contrast with simple premise-premise-conclusion argument indicators, and other means of regulating logical flow.

**Using WebSequiturs as an innovative MDM pedagogical tool**

WebSequiturs are OTREs which are generated using an application provided by Creative Technologies, which is downloaded onto a user’s computer. Once generated and posted online, these WebSequiturs can then be accessed by anyone with the link. WebSequiturs present segments of text passages in a multiple-choice format on a computer monitor. Students are tasked to reconstruct the text from a selection of sentences or sentence fragments presented at random in multiple-choice format on a computer display. They then select one of the displayed choices to recreate a correctly sequenced or coherent version of the original text. What makes this application unique is the immediate feedback, as, with each correct selection, the text is visibly reconstructed in the lower half of the screen.

Learners reconstruct the text sequentially from beginning to end. The reconstructed text appears, a segment at a time, below the multiple-choice area as learners work through the multiple-choice selection (see Figure 3.11). This feature and the linear quality of the learners’ movement through reconstruction is what makes WebSequitur unique. No other application offers sequential selections in multiple-choice format and while also displaying the constructed text in a completion field.
Higgins, Lawrie and White (1999) describe the Sequitur type of OTRE as having ‘great potential as a tool for the investigation of native and non-native speaker intuitions about the structure of paragraphs’ (Higgins et al., 1999, p. 346). In this study, I have used OTREs to that end.

3.7.7. Merits of collaborative TRE work

Although text reconstruction exercises can be done by students individually, it has long been noted that there are advantages to tasking students to work on these exercises in pairs or small groups. As Warschauer and Healey write, ‘text reconstruction exercises [. . .] become less frustrating and more satisfying, when students are collaborating to find the missing words’ (1998, p. 61). Learners discuss and negotiate their way to deciding which of the options they believe is most likely to be correct. Whether this is done in the target language is of little importance as long as learners are using their store of linguistic knowledge to its fullest potential (Cook, 2001). Malmqvist (2005) notes that discussion during text reconstruction exercises included many language related episodes and this motivated me to examine transcripts of student-student talk during OTRE sessions.

Speaking in the L1 should not be discouraged provided that the discourse is on topic, because restricting students from communicating in their L1 imposes constraints that prevents
them from object regulation necessary for effective learning (Brooks and Donato, 1994). Storch (2010) noted that, during form-focused text reconstruction pair work, in cases where learners discussed their choices and questioned other members’ decisions, there was more consolidated learning than cases where such negotiations did not occur.

3.8. Forum readings
Prior to AW instruction, participants read a one-page story titled “To Tell the Truth” (see Figure 3.12). The topic is one that can be expected to evoke genuine responses from students about a social issue in Japan, involving doctors often not divulging to some patients the seriousness of their illness. Subsequent to completion of each round of OTREs, participants were assigned a forum discussion. The aim of collecting this data was the production of a quantitative record of the participants’ written MDM use. This discussion task took place within the online Forum area (see Figure 3.13) on the university’s Moodle-based private webpage, Learning Environment of Tokyo University of Science (LETUS).
Mrs. Takayanagi had just turned 70. She had been in good health all her life. Recently, she had lost her appetite and was feeling extremely tired when doing her everyday chores, but she believed that it was because of age. She had been taking antacids for her stomach for a few months. After she started to feel pain in her upper abdomen she decided to visit a doctor. Her son, Teppei, took the day off work and went with her to see her physician, Dr. Edwards.

After her examination, she was told she would have to spend a weekend at the Chicago City Hospital for further examination. This made her very frightened, but her son told her that it was just a general examination and that it was a good opportunity for her, since she had not had a checkup for several years. She was placed under the care of Dr. Spencer. Several diagnostic tests were performed.

Mrs. Takayanagi’s son was waiting for Dr. Spencer outside his mother’s room. After hearing the result, Teppei asked the doctor not to tell his mother the truth.

“My mother is a very delicate person. She will not be able to cope with this terrible news. Couldn’t you just tell her that it is a peptic ulcer?”

However, Dr. Spencer was determined to tell Mrs. Takayanagi about her disease. He told Teppei that it was the patient’s right to know the truth. The doctor then went into the room, followed by Teppei.

“Mrs. Takayanagi we’ve got the test results. I’m afraid it is quite serious.”

When Mrs. Takayanagi heard the details of the diagnosis, she became so upset that nurses were called to administer sedatives by injection. Teppei was very upset with Dr. Spencer.

“I asked you not to tell my mother about her condition. Now look at what you have done. Why could you not respect my requests?”

**TASK**

**Discuss the following questions:**

1. Do you think it was right for Dr. Spencer to tell Mrs. Takayanagi the truth about her disease?
2. What factors do you think doctors need to consider when informing patients of a serious illness?

Reply to these questions by clicking on the 返信 button below. Please do not start a "new topic". Agree and disagree with the other students in this forum and give reasons for your opinions. You should give complete answers and reply (返信) to at least two other students.

Figure 3.12: Short narrative and discussion questions
Both online forums and chat rooms were utilised as an area in which students could express their opinions on assigned topics, using the MDMs and thereby produce two data sets. Forums support students in discussing topics asynchronously over an extended period of days or weeks (Moodle, n.d.a); while chat rooms support students in synchronous, i.e. real time, communication (Moodle, n.d.b). This means that as new comments are made by members in the chat room, they appear immediately after they are entered and submitted, whereas in a forum, a student cannot see new forum posts or replies until they revisit the forum or refresh their browser.

The participants’ written output gathered from the forums and chat rooms was a major source of quantitative and qualitative data as it showed changes in frequency of MDM use.
across a number of measurements, thereby addressing RQ1; and could be used to identify trends in the appropriacy of MDM use, thereby addressing RQ2.

3.9. The Monty Hall Problem

During the latter segment of the term students study the formation of conditional statements and hypothetical premises. They are tasked to join a forum discussion and write hypothetical premises concerning a question of probability both before and after undergoing two OTRE sessions aimed at teaching the formation of conditional statements and hypothetical premises. In this way, I am able to track any changes in participants’ developing skills in conditional statement and hypothetical premise formation. During a lesson, I introduce the basic idea of the class project and inform students that the project will culminate in a group writing assignment, completed at the end of the course.

This project focuses on an issue in probability that has come to be known as The Monty Hall Problem. I chose this particular topic for the ease with which it can be developed as an activity (described below) for active learning (Brokaw and Merz, 2004) and for its intrinsically counter-intuitive qualities which make it a suitable topic for critical thinking instruction at a science university (Tokyo University of Science, n.d.) (see Appendices 9-10).

Moreover, it involves an issue which, although simple in that it involves three cups and a coin, has the potential to elicit widely varying opinions regarding an issue for which hypotheses can be easily constructed, and after some discussion regarding methods of enquiry, those hypotheses can be tested and the results reported.

The problem presented involves a question of probability. A person is shown three cups, Cup 1, Cup 2 and Cup 3, told that under one of the cups is a ¥100 coin and are asked to choose one cup. Let us say that the person chooses Cup 1. Instead of revealing what is under Cup 1, the person controlling the game overturns another cup, perhaps, Cup 2, that also has nothing under it. The controller then asks the player if, for a fee of ¥30, he or she would like to change his or her original choice from Cup 1 to Cup 3. Most people will choose to remain with their original choice, and almost no one will pay the ¥30. The reason for not changing given by most persons is that the probability is clearly 50/50 and so changing cannot increase their chances.

The fact is, however, that the player should always change from their first choice. This is because, another way of saying that a player has a 1/3 chance of choosing the cup with the coin is to say that 2/3 of the time, the player will choose one of the two cups containing nothing. Therefore, after the first cup containing nothing is lifted, there is a still a 2/3 chance that the
initially chosen cup does not contain the ¥100; Hence, the player should always change cups and should be willing to pay up to the value of 1/3 of ¥100 for that privilege. When syndicated Parade columnist Marilyn vos Savant wrote in 1990, advancing the counter-intuitive idea that the player should change cups, she was opposed by mathematicians worldwide (Crockett, 2016).

In order to help students to visualise the issue, the following handout (see Figure 3.14) was distributed in class:

![Handout](image)

**PART ONE: WHAT IS THE DIFFERENCE BETWEEN LUCK AND SKILL?**

Do you believe in luck? Do you think that you are a lucky person? Do you think that some people are luckier than others? Or is luck just really a skill that some people know about and other people don’t? During the next few weeks you will conduct a carry out a scientific study of probability involving a game called Cups and Coins.

First we need to know the rules and practice playing. For now, let’s keep it simple:

There are three cups, all turned upside down. Under one of cups there is a coin. The game-master knows where the coin is. The player may choose Cup 1, Cup 2 or Cup 3.

Imagine that the game-master puts the coin under Cup 3. The player chooses Cup 1. The game-master shows what is under Cup 1—nothing. The player doesn’t win the coin.

Imagine again this example: The game-master puts the coin under Cup 3. The player chooses Cup 3. The game-master shows what is under Cup 3—the coin. The player wins the coin.

Practice the dialog on the next page until you can play the game without looking at the text. Some pairs will be asked to perform the dialog at the front of the room.

![Dialog](image)

**Figure 3.14: Cups and Coins, Part One, handout**

Students used the handout as they collected data in the class project (see Figure 3.15)
Subsequent to a forum discussion of the results, students are assigned *Critically Minded Podcast Episode 8: Conditional Statements and Hypothetical Premises*. During the following two class meetings, they work together on OTREs. They are informed about the first-choice switching variation of the Monty Hall Problem. They are then tasked to discuss the problem and form a hypothesis in preparation for the second experiment the following week. Although typically half or more of the students predict a 50% success rate, the results of this second experiment is invariably a class-wide average success rate of roughly 67%. Subsequent to the second experiment, students return to the forum that week and discuss the results and how their initial hypothetical premises have been confirmed or contradicted.

3.10. Summary: Research tools that relate to RQs 1, 2 and 3

In this section, I described several research tools, so that I will later be able to reference these without breaking the continuity of my discussion. I explained the categorisation of communicative functions that express argumentative structures as they are commonly presented in the literature cited in Chapter 2. I described the Master MDM List. I noted that the way the communicative functions and MDMs were presented might act as an additional influence on the frequency of participants’ use of MDMs (RQ1) preceding participants’ engagement with the OTREs. I explained the use of Critically Minded Podcast and how the episodes proceed in lockstep with the order of communicative functions and the Master MDM List. I noted that the way the MDMs were presented and modelled might act as an
additional source of influence for the more frequent use (RQ1) of MDMs; and guidance for participants’ more appropriate use (RQ2) of MDMs, preceding participants’ engagement with the OTREs.

I have also discussed the WebSequiturs in detail. I have explained how they are made and the rationale behind their construction in order that the reader may later appreciate their unique qualities and understand why they are specifically suited to the instruction of MDM use. I have discussed how this construction affects participants’ experience using the OTREs (RQ3).

3.11. Methods

In this section, I discuss the data collection points. Grouped by kind, and sub-grouped by other factors explained in each sub-section, these include: the OTRE sessions 1, 2, and 3 and OTRE sessions 4 and 5; Forum, Forum II and V, and Forum III and IV, Chat Rooms I and II; the Group Writing Project; the Individual In-Class Rewrite; and the Semi-Structured Interviews.

3.11.1. Initial lessons and participant selection

One point which has emerged is that assigning students to pairs invites problems, since if one student drops out of the class or is merely absent on a given day, then the student with whom that absent student has been paired is now without a partner and must then be paired up with another student or students to form a new pair or a triad. Losing a partner and then being moved into an already established pair can be demotivating. In the case of triads, i.e. groups of three, when one student is absent, the other two students can continue as a pair for that day, or for the remainder of the term, should one student drop the class. Although there may be reasons, involving group dynamics, for my part, that one pragmatic reason alone has been sufficient to adopt a practice of placing students in triads. Moreover, although in groups of three there is less social pressure on any one member to contribute, it is also arguably the case that less outgoing students can learn vicariously by listening to and observing the other two members (Sutton, 2001).

Having assigned students to triads, during the next three lessons, I noted which triads’ members were the most punctual and attentive, and observed the general demeanour of each student. Triads were chosen without any prior knowledge of their members’ English communicative skills relative to classmates. After my own previous observations and Yilmaz (2011) which likewise notes increased negotiation in mixed gender dyads, I selected the only
two female students of the class to be among the group I first asked to join the participants group. Other than an attempt to reach an even gender distribution, the selection was entirely random. I spoke with the selected triads, briefly explaining my work and the purpose my study. I also discussed each point of the consent sheet (see Appendix 14) in order to be sure that each triad member understood the scope of their participation. Table 3.6, above, provides a week-by-week overview of the key points of the course relevant to this study.

3.11.2. Online Text Reconstruction Exercise sessions

In this section, I discuss the Online Text Reconstruction Exercise (OTRE) sessions. These sessions occurred in two frames. The first of these involved the first three sessions, which focused on premise and conclusions indicators. The second of these frames involved the fourth and fifth sessions, which focused on the structures and MDMs used in conditional statements and hypothetical premises.

**OTRE session 1, 2 and 3**

After students had received out-of-class explicit instruction in the expression of premises and conclusions, delivered via Critically Minded Podcast Episodes 2 and 3 (*Organizing an Argument 1* and *Organizing an Argument 2*) (see Figure 3.4). They were then asked to complete three OTREs during the first of three OTRE sessions focusing on premise and conclusion indicators.

**Improvement in the OTRE sessions after the pilot**

In the pilot study I had noted remarks expressing fatigue, during the first two OTRE sessions, such as the one below, translated into English, which suggested that the two sessions had been too long.

Pilot3: *Ohh, I'm tired.*
Pilot1: *Yeah. How far [down the list of exercises] do we work? We have already done a lot.*
Pilot3: *What? Three more?*  
Pilot1: *Oh no!*
Pilot3: *It's already a lot!*
Pilot1: *Oh no!*
Pilot3: *My eyes are getting shaky!*

I responded to this by dividing the OTREs into three sessions. I noted that the transcripts for the student-student talk during the OTRE sessions in this study, included no such complaints
and that during all three OTRE sessions students worked together in a lively way from beginning to end.

Some time was needed for students to familiarise themselves with the interface during the first short sample exercises during the first session. Subsequently, there were several exercises of increasing length. Students were given about half the class period (45 minutes) to complete these. I had not planned on allotting that much time, but I noted that students were pausing at times to use the online dictionary to check their comprehension. I interpreted this to mean that the students were taking the task seriously, genuinely wanted to understand the text passages and were generally enjoying this kind of study.

In these OTRE sessions, as well as the two that followed the score and time required to complete the exercises was not considered indicative of a greater or lesser degree of learning. Rodríguez (2010) had likewise found that students tended to end and restart exercises in which they were scoring low. All three triads generally took approximately the same time to complete the exercises, and all were able to complete them within the allotted time. I did not observe instances of spans of inactivity, nor did I observe cases of suddenly hurried work involving random guessing. Students worked conscientiously and at a steady rate with a near equal amount of spoken input from each participant, notwithstanding the expected differences that the variations in personality and perceived competence might occasion.

In this study, after OTRE sessions were increased in number and decreased in length, there were occasional comments about being tired, but one had the feeling that they were making small talk, and not spoken from a place of real fatigue. They were voiced only at the end and with a sense of satisfaction. The follow is an example:

P2: Nanka zenkai yorii muzui. *This one is kinda harder than the previous one.*

And later:
P1: Owatta! *We’re finished!*
P2: Kanyou! *We’re finished!* and
P3: Hara hetta! *I’m hungry!*

Here, I noted that, although relieved to be finished, more satisfaction could be inferred than in the pilot study, where the same amount of WebSequitur work was done during fewer sessions so that the first two sessions were, according to participants, exceedingly long.
OTRE sessions 4 and 5

During the eighth and ninth lessons, the class was held in the computer lab. During this lesson, students worked in their triads, completing the fourth and fifth OTRE sessions. These OTREs (see Figure 3.1) focused on the language of conditional statements and hypothetical premises. These structures were chosen because of the importance of conditional statements and hypothetical premises in forming arguments and expressing opinions (as set out in the TUS English curriculum). They were also chosen because of this being a writing skill necessary to science students’ academic writing.

They were based on three passages of short and medium length involving narratives about problem solving. This kind of text reconstruction asks students to reconstruct a story involving both conditional statements and hypothetical premises. The distinguishing point that these exercises focus on is the first clause present tense, second clause present tense construction of conditional statements (e.g., *When I eat at that restaurant, I get sick*); and the first clause present tense, second clause future tense construction of hypothetical premises, e.g., *If I eat at that restaurant, I will get sick*. To develop students’ ability to recognise and later form these constructions themselves, a variety of language items were offered in the Master MDM List (see Appendix 13) discussed in the podcast and presented in the OTREs. Whereas the premise-and-conclusion exercises focus on sentence breaks at the point where logical flow was indicated by the premise or conclusion indicator, conditional statements and hypothetical premises do not switch markers when the flow of logic is reversed. For example, *If I eat at that restaurant, I will get sick* or *I will get sick if I eat at that restaurant.*
Figure 3.16: The second of the three OTREs focusing on conditional statements and hypothetical premises

**OTRE-session student-student talk**

Before presenting the aforementioned data, it is useful to learn what kind of talk students generally engaged in during OTRE sessions. I have identified some main themes (see Table 3.7) that I was able to categorise during my transcription process. In addition to grammar-based strategies, a few other themes became recognisable. They were as follows:
1. Discussion of strategy for completing the exercise:
   - Grammatical cues
   - Orthographic cues, e.g. capital letters, punctuation

2. Disagreement

3. Correction of misunderstandings

4. Explicit recognition or observance of the presence of premise indicators, conclusion indicators in the form of:
   - Premise and conclusion indicators
   - Conditional statements
   - Hypothetical premises
   - Order of metadiscourse markers within a text passage

Table 3.7 Main themes of student-student talk during OTRE sessions

3.11.3. Forums

In view of the affordances detailed in Sections 2.7.1-2.7.2, forums were regarded as an important research tool as they were one source of written data and provided a record of MDM use before the first podcast listening and the first OTRE session and subsequently after each round of OTRE sessions. These forums differed in ways that lead me to describe them in two frames. These are:

- Forums I, II, and V
  These three forums each included extended readings and several questions students were tasked to discuss over a two-week duration. Students were expected to write somewhat extensively and to return to the forum several times to comment on other students’ writing and to respond to comments and questions from other students.

- Forums III and IV and Group Report
  These two forums tasked students to write briefly and to form hypotheses predicting the outcome of Part 1 and Part 2 of the Cups and Coins enquiry into probability.

3.11.4. Forum I

Following Rodríguez (2010), I assigned one forum discussion before the first OTRE session. The purpose of this assignment was to establish a pre-OTRE baseline of participants’ use of selected MDMs against which I could later contrast MDM use after participants’ completion of successive rounds of OTREs. Students were assigned a first out-of-class online Forum discussion over a short narrative text I had written titled “To Tell the Truth” (see Section 3.8). The forum was completed and closed before the Master MDM List was made visible on the
LETUS page, before the podcast series was made available, and before the participants had received any training in MDM use via OTREs.

3.11.5. Forum II and V

Students engaged in a post-TRE out-of-class online asynchronous text-mediated discussion thread within a second LETUS Forum. This forum was different from the first, in that students had begun receiving explicit instruction in the use of premise and conclusion indicators via Critically Minded Podcast. They had also received instruction in giving additional reasons and asking for reasons through Dialogically Minded Podcast (Gann, n.d.). They had also completed the first OTRE session. Students were given two weeks to participate in the discussion there. The topic of the discussion (see Figure 3.17), fat shaming, was an issue presented in the class textbook, on basic critical media literacy and advertising with questions about how students are affected by marketing campaigns. Students were directed by a link to a March, 2018 Washington Post article titled, Why we need to take fat shaming out of fitness culture. Selected MDM use was recorded and quantised via the method and scale discussed later in Section 4.2.2.

Following this and the two intervening forum discussions on the Cups and Coins enquiry into probability, there was a final Forum Discussion 5. The subject of this was the history behind Japan’s traditional belief in blood type personality profiling and the problems that belief in this pseudoscientific idea can cause in culture and in individuals’ lives. The procedure for assigning this forum was identical to that of Forum II.

![Figure 3.17: A teacher’s post introducing the second forum](image-url)
3.11.6. Forum III discussion
Forum III and Forum IV differ from Forums I, II, IV in that their responses were much shorter, they were intended for the instruction of conditional statements and hypothetical premises, and they shared a Part 1 and Part 2 relationship.

At the end of the sixth lesson, I assigned students a forum discussion (see Figure 3.18) of the simple one-guess version of the Cups and Coins game. They were tasked to produce their first hypothesis involving a situation involving three cups, one poker chip and one guess with no opportunity to change their guess. This forum was assigned prior to listening to the Critically Minded Podcast on conditional statements and hypothetical premises and prior to the two related OTRE sessions.

![The Three Cup Question UNIT 3](image)
Figure 3.18: A teacher’s post introducing the third forum

3.11.7. Forum IV
In the eighth lesson, the Critically Minded Podcast episode on conditional statements and hypothetical premises was assigned and the first OTRE session covering conditional statements and hypothetical premises was completed by students. Forum IV, the discussion of Cups and Coins Part 2, was assigned. In this forum students were tasked to make predictions about the more complex version of the Cups and Coins game, which allows for the second guess and thereby introduces the Monty Hall Problem. In this version after selecting one of three cups, an empty cup is lifted, and the player is given the opportunity to change his or her original selection. For the purpose of the experiment, it was predetermined that players would change their first cup selection for all 30 trials. The following handout (see Figure 3.19) was distributed.
In this forum, students were tasked to use the language of hypothesising, fronted via podcast delivered instruction (Critically Minded, Episode 8) to discuss predictions of the results they expect to arrive at in their subsequent experiment.

Students were tasked to discuss how to go about forming and testing a hypothesis correctly using the selected MDMs. They were tasked to make predictions, by way of forming hypotheses, of what the rate of success would be under the conditions of Cups and Coins 2, given 30 trials.

3.11.8. Chat Rooms Introduction

Chat Rooms I and II

As with the forums, in view of the affordances detailed in Sections 2.7.1-2.75.2 chat rooms were also regarded as an important research tool as they were one source of written data and provided a record of MDM use after the third and fifth podcast listening and the first OTRE session and subsequently after each round of OTRE sessions. In contrast with forums which were completed out of class, asynchronously over two-week durations, chat rooms (see Appendix 15) were completed in class within 45 minutes (Chat I); and one hour (Chat II).
Chat Room session I

In the seventh lesson, the first text chat session was administered. This took place in the computer terminal room. Roughly eight online chat rooms were prepared, one for each triad. The seating for the students was arranged so that members of each triad were distanced from one another as with the in-class synchronous collaboration performed via shared Google Docs in Shintani and Aubrey (2016). Students were tasked to answer the same questions as during the previous forum discussion. Rather than including the entire class, as with the forum, in one chat room, each chat room included three members, thus placing a greater responsibility, and opportunity, for each member to make statements, draw conclusions, and ask their interlocutors for reasons. Students were given 45 minutes to complete their discussions. All three groups of participants were in attendance within the LETUS system chat room application (see Figure 3.20). The topic of the chat, again, was fat shaming, the same topic as Forum II. The topic was selected at the beginning of the class by a coin toss between this topic, heads being the topic of the first forum and tails being this topic. Students were encouraged to use as much of the language that they could recall from the forum while remaining spontaneous. The following figure is representative of the chat output during these sessions. The full names have been obscured for privacy.
11:00
In my opinion, we have to get know the man that we are going to give advice before do it.

11:01
Without doing that we cannot say anything.

11:02
I agree with Shotaro, but for example? the number of weight? his character?

11:02
The line we should not cross is depends personal condition.

11:02
I think if we give give advice to fat people, we should give carefully, avoiding hurting them.

11:03
I agree with Kae.

11:04
I think that it is important to know what the people think about their body shape.

11:05
In addition to Shotaro’s opinion I think it is especially important to know about the friend’s character.

11:05
I agree with opinion of Takuya

11:06
to Kenntaro/the number of weight: I think we should think about what Takuya said.

11:07
How can you know what the people think about their body shape?

11:08
In my opinion, some people may feel shamed of their bousy shape, but others may not feel so.

11:08
I absolutely sure about that, Kae.

Figure 3.20: A chat between three students
Chat Room session II

The second chat room was completed under the same circumstances as the previous one with the exception that an additional fifteen minutes were allotted and in addition to the forum discussion on blood type personality profiling, students were also asked to comment on the class itself. Some of these comments were useful in complementing the data collected from the semi-structured interviews.

3.11.9. Group report

In the fourteenth lesson, the last lesson of the year before winter break, I assigned the final Group Report, a short paper written collaboratively via one shared online document per group. The topic of the paper was the class project, the study of probability concerning the Monty Hall Problem. The paper was assigned (see Figure 3.21) as a short paper, between 500 and 750 words in length. This assignment was presented to students as an opportunity to demonstrate their most extended novice attempts at academic writing. Students responded to prompts to write specific sections of a paper: introduction of the issue, hypothesis, methodology, results, discussion and conclusion, including a section on how the import of what they learned could be applied to their subsequent scientific inquiries. The rationale for collecting this data was essentially the same as for the data gathered from the forum discussions and text chat assessments.
Figure 3.21: Task brief for Group Report

The participants of this study submitted their final report (see Figure 3.22) on January 3, during the interim before classes recommenced. In the fifteenth and final lesson, selected MDM use from the Group Report was recorded via the same method and scale as with the method and scale used for the forums and chat room sessions output.
There was also the opportunity to note how the participants’ arguments extended across wider text passages than in the forums for chat room transcripts. Because this was regarded by most students as a task equivalent in importance to a final exam, it was expected that this writing would be representative of students taking the greatest care to construct an argument according to the training in MDM use they had received during the course. The rationale for collecting this data was essentially the same as for the data gathered from the text chat assessments. In Figure 3.22, below, I have highlighted MDMs in blue. The analysis of MDM use is detailed in Section 4.2.2 where I discuss a scale that was used by three independent raters and in Chapter 6 where I discuss the findings of that assessment.
An Unexpected Development in a Simple Statistical Study

I. The topic is the probability of choosing, among three cups, a correct cup which contains a coin. We answered this question by experiment which tested our hypothesis many times.

II. What do you think and why do you think so?
In our opinion, 33% is also common sense. We should repeat this experience again and again in order to avoid assumptions. Our hypothesis is that the more times we do this experiment, the more close to 33% the probability come.

If we test 100 times, then we can choose the cup with coin about 33 times.
So, the probability is 33%.

III. Method. How we tested your hypothesis.
There are 30 pairs of cups in one group with three cups. Also one cup out of three has a coin. You cannot see the inside of the cup from the outside. Nine groups did this experiment.

Next, we decide one player and one game master. The player will decide one cup having a coin out of three. If the player can hit the cup having a coin, then the game winner is the player. Conversely, if the player chooses the cup not having a coin, then the game winner is game master. This operation was carried out 30 times in each set. Then the following result was obtained. The probability of the player winning is 29.1111%.

IV. Data. What kind of data did you collect?
Whether it is displayed as a fraction or a small number, or percentage or not. Mathematically, it shows the same thing. Therefore, the group result was calculated by percentage. And the class average also was calculated by percentage.

V. Discussion. Discuss the data.
Our hypothesis was that the probability of first experiment is 33%, and the probability of second experiment is 69%. And the results of these experiment were the first is 23%, and the second is 60%.

The results were similar to our hypothesis. It indicated that our hypothesis are confirmed. It means that we could judge whether our hypothesis are correct or not by the date which we got. So we were satisfied and felt that we gathered enough data.

We think the result would be more similar to the hypothesis, if we gathered more data. The result in general, which it is clear that the experiment has more data than our team, were that the probability of first experiment is 29%, and the probability of second experiment is 66%. they were more similar to the hypothesis. That is why more data indicates the result which is closer to our hypothesis.

We weren’t surprised by the result, because the result is similar to our hypothesis.

VI. Conclusion
The result in general were that the probability of first experiment is 29%, and the probability of second experiment is 69%. Therefore, this experiment suggests that us that our hypothesis is correct.
3.11.10. In-class rewrite
On the last class meeting in the 15th week, students were tasked to rewrite, from memory, to the best of their recollection, the paper that their group had submitted. Triads were reassigned to different seats so that they could not share information. Students were not allowed to use the Internet, except to submit the rewrite at the end of that class period, at which time the assignment submission was closed. I monitored the class from the back of the room, where I could view all computer monitors. No notes, including the Master MDM List were allowed to be viewed. The purpose of the task was to aid me in later determining if any members of the three triads had contributed little or nothing or had not even seen or read their own group’s paper.

Students who could only recall the sections of the paper that they had written themselves were encouraged to rewrite only those sections. Students who could only recall isolated sentences, sentence fragments, or MDMs that had been used, were encouraged to write. The task was given in the second half of the lesson. Students were told they could leave when they had finished the task to the best of their ability and had submitted their work electronically on the LETUS page.

3.12. Semi-structured interviews
In this section I discuss the rationale for using semi-structured interviews as a means of answering RQ1, RQ2 and RQ3. I work from the interpretivist precepts introduced in Section 2.9 in the metaphor of interviewer as traveller, who co-constructs meaning with interlocutors (Kvale, 2007). In this way, participants played an active role in forming an explanation for the data which was gathered up to the time of the interviews.

3.12.1. Rationale for interview method
In this section, I discuss the connection between theoretical position and interview methods. In planning for the interviews, the need to continue working within an interpretivist paradigm led to my selecting an approach to interviewing, which can provide for the researcher a sense of what participants are experiencing (Currie and Kelly, 2012), and later, an inductive approach to thematic analysis, which utilises the concept of coding, i.e., the categorisation of responses or topics in such a way that interpretations ‘emerge from and are grounded in the data’ (Lapadat, 2010, p. 4).

Regarding the social distance, or positionality (Marshall and Rossman, 2016, p. 145) the researcher maintains between his or herself and participants, I did not anticipate, nor did I
find, the kind of emotionally charged topics, such as in Ezzy (2010). As the teacher of the participants, I had already formed a personable, but professional relationship (Marshall and Rossman, 2016).

Based on this rationale, therefore, in these interviews, I explored salient themes detected in the OTRE group session transcripts. I searched for instances of consistency and contrast revealed by triangulation to determine to what extent and how OTRE-work relates to students’ subsequent use of selected MDMs. Interview transcripts were then analysed for the presence of other emergent themes which might further shine a light on the hidden life of what occurs between students during OTRE group work and the relationship, if any, between WebSequiturs and the development of skilled use of argumentative academic MDMs.

3.12.2. Interview questions
Interviews were conducted in January 2019 after the analysis of the data collected from the forums, chat rooms, and Group Reports. Individual semi-structured interviews were conducted with each of the participants and had a duration of approximately twenty minutes. In Table 3.8, interview questions have been connected to the research questions they intend to answer. These questions were translated into Japanese for use during the interview (see Appendix 16). A full transcript of an interview transcript example is provided in Appendix 17, and an excerpt of that interview including coding notation and comments is in Appendix 18. In the course of the transcription I relied upon the Sage guide for using interviews in qualitative research (Rubin and Rubin, 2005). I prepared a system of notation (see Appendix 19) ahead of time in keeping with Davidson’s (2009) advice against using systems of notation created after the fact. As a somewhat inexperienced interviewer, however, I found some types of unforeseen expression arose and I did my best to respond to such instances.
Table 3.8: Interview questions and RQs

3.13. Ethical considerations

In this section, I discuss the stance I took towards ethics, what ethical issues were plausibly involved and how I protect participants’ safety and wellbeing, how I handled participants’ data, and how I communicated these precautions to the participants.

The research proposal for this study was written in adherence to the following principles and was approved by the OU Human Research Ethics Committee and ranked as low risk to all involved. At each stage of the research design in this study, care was taken to adhere closely to the relevant ethical guidelines.

My research design is in conformity with the British Educational Research Association’s (BERA, 2011) Ethical Guidelines for Educational Research, including respect for persons, knowledge, democratic values, the quality of educational research, and academic freedom (BERA, 2011, p. 6) and was approved of by the senior faculty members in our English teachers’ group and the wider faculty of our liberal arts department. Integral, to my research design was Hammersley and Traianou’s (2012) five central ethical principles: doing minimal harm, fostering autonomy, maintaining privacy, offering reciprocity, and treating people equitably. Accordingly, participants were treated the same as other students, except during OTRE sessions, when small audio recorders were used for the purpose of later producing transcripts of participants’ utterances. In all other respects, participants’ experiences in the forums, chat rooms, and Group Report were identical to that of their classmates. Like all students, who went through an identical process, participants were
encouraged to use the materials (which were, themselves, geared towards promoting critical literacy and expression) to empower themselves as competent communicators of relevant opinions in discussions in and about the class.

Participants were informed in the permission form they were given that their participation was voluntary, that upon collection of the forms that if any forms were returned unsigned that another participant would be found among their classmates, that the matter would not be mentioned again, and that non-participation would have no effect on their final grade. They were also informed that they could drop out at any time. They were apprised as to what data was being collected, and how the data would be used. They were assured that their names would not be published and their personal participation in the study would never be divulged. They were informed that upon the conclusion of the study that they would be contacted through student mail and invited to read a pdf of the thesis. Lastly, they were assured that all original data collected would be deleted or otherwise destroyed upon the completion of this thesis. This information was written in easy-to-understand English and during a single group meeting at the end of the class meeting when the forms were distributed, each point was covered, the opportunity to ask for clarification was offered and each candidate’s comprehension was verbally confirmed. During the pilot, one student did in fact opt out on the day of the explanation and in the subsequent class meeting another student, who was discretely selected in a random fashion, agreed to become a participant.

The issue of ethical accountability in action research has historically been a problem, because traditional empirical research frameworks have been slow to compensate for differences with this relatively new way of conducting research. In contrast to traditional methods, in action research, the teacher-researcher works as an insider-researcher, and different concerns about the power dynamic (Layder, 2013) between researcher and participant are raised (Gibbs et al., 2017). Being cognizant of these issues, I took care at every juncture, not to objectify participants as data sources, but to maintain a reflective practice which accepts ambiguity, maintains a humane standard of compassion and integrity (Janesick, 2016) and ‘thinks emotionally’ in order to maintain ‘heightened attunement to and awareness of participants’ feelings’ (Saldaña, 2015, p. 82). Moreover, contact information was included in the form which students were told they could avail themselves of, in Japanese, should they feel that the role of participant infringed on their autonomy as a student or was detrimental to their educational experience in any way.
3.14. Summary

In this chapter, I described the methodology that I used to answer my research questions. I explained the theoretical underpinnings of my position and the rationale for generating data through the application of action research.

I discussed my mixed-methods approach and how I determined that using both quantitative and qualitative data was necessary to this study. I described several of the teaching tools, knowledge of which is necessary to understanding my discussion of the findings in Chapter 4.

I discussed the means by which I generated data and I linked each of those with one or more of the three RQs. I listed my research questions and linked each of those to one or more of the three RQs. I discussed the context in which the research was performed, and I introduced the backgrounds of the student-participants. I outlined the research design and commented on the possible influence of the materials. I explained how I recorded the data and how I used statistical analysis of frequency and variety of MDM use and range of communicative functions for quantitative data analysis; and how I used a thematic analysis approach for the qualitative data. I concluded with an acknowledgement of the ethical concerns and explained the actions I took to see to the safety of the participants and the security of the data that was generated from this study. Lastly, I have discussed how I carried out the study within the limits of the ethical guidelines of Tokyo University of Science and those provided by BERA and Open University’s OU Human Research Ethics Committee.
Chapter 4: Data analysis

In this chapter, I report on the rationale for my analysis of the data that was collected from the sources detailed earlier. In Section 4.1, I discuss the categorization of the quantitative data and the various lines along which I quantified it. I also discuss the measurements which I made, the rationale for their use, the extent to which they were used, and any assumptions which were made about each and their limitations. I present the scale (Section 4.2) that I developed for appropriacy of MDM use and how I involved two knowledgeable and experienced colleagues in independent informed assessment using that scale. In Section 4.3, I explain the scale for assessing the construction of hypothetical premises. In Section 4.5, I discuss qualitative analysis involved in thematic analysis of OTRE session transcripts, later written output, and the semi-structured interview transcripts. There is a summary of the chapter in Section 4.6.

4.1. Quantitative analysis of MDM use

Statistical analysis with regard to (1) the frequency of use (RQ1), and (2) the appropriacy of use (RQ2) for the MDMs specific to premises and conclusions, and conditional statements and hypothetical premises used in argumentative writing (AW) was performed on quantitative data collected using the following four collection tools:

- Five forum discussions
- Two chat room sessions
- One final group report
- One in-class In-Class Rewrite from Memory task

4.1.2. Forums as a data collection tool

The number of participants’ forum activity is shown in Table 4.1. The details, such as word count, etc. are provided with commentary in Chapters 5 and 6.
Table 4.1: Number of posts for each participant during each forum

<table>
<thead>
<tr>
<th>Participant (P): (F)orum: Number of Posts in each Forum</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: F1: 3 / F2: 1 / F3: 1 / F4: 1 / F5: 1</td>
</tr>
<tr>
<td>P2: F1: 1 / F2: 1 / F3: 1 / F4: 1 / F5: 1</td>
</tr>
<tr>
<td>P3: F1: 3 / F2: 2 / F3: 1 / F4: 1 / F5: 1</td>
</tr>
<tr>
<td>P4: F1: 3 / F2: 2 / F3: 1 / F4: 1 / F5: 3</td>
</tr>
<tr>
<td>P5: F1: 2 / F2: 0 / F3: 1 / F4: 1 / F5: 0</td>
</tr>
<tr>
<td>P6: F1: 1 / F2: 2 / F3: 1 / F4: 2 / F5: 1</td>
</tr>
<tr>
<td>P7: F1: 1 / F2: 0 / F3: 0 / F4: 0 / F5: 1</td>
</tr>
<tr>
<td>P8: F1: 4 / F2: 3 / F3: 1 / F4: 1 / F5: 1</td>
</tr>
<tr>
<td>P9: F1: 1 / F2: 2 / F3: 1 / F4: 1 / F5: 1</td>
</tr>
</tbody>
</table>

Table 4.1: Number of posts for each participant during each forum

4.1.3. Chat rooms as a data collection tool summary

Participants’ basic activity is shown in Table 4.2. The details, such as word count, etc. are provided in detail with commentary in Chapters 5 and 6. In short, it can be seen here that between participants, the number of turns varied. However, one commonality is the increase in turns in the second chat. This may be due in part to the additional 15 minutes allotted in the second chat session, but other factors may have been involved and this is explored in Chapters 5 and 6, where the details such as word count, etc. are provided and explained in detail.

Table 4.2: Chat Rooms’ durations and number of turns for each participant

| Participant (P): Chat I (45 minutes) Number of Turns / Chat II (1 hour) Number of Turns |
|--------------------------------------------------|---------------------------------|
| P1: 13/35                                         | P2: 22/65                       | P3: 14/42                       |
| P4: 12/37                                         | P5: 14/45                       | P6: 32/69                       |
| P7: 15/23                                         | P8: 16/23                       | P9: 15/47                       |

Table 4.2: Chat Rooms’ durations and number of turns for each participant

4.1.4. Group Written Report as a data collection tool

A final Group Written Report (henceforth GWR) was assigned for each triad, with the task rubric shown in Figure 3.21. No minimum or maximum word count was set as, for the purposes research, it seemed best to suggest a length of about a page to a page-and-a-half, and not to over-prescribe the content, but to allow the influence of the instruction, especially that of the OTREs to play out in the participants’ output. The three GWRs appeared in lengths that allowed each group to complete the rubric and were roughly a page to a page-and-a-half in length. The text was subject to the same quantitative analysis as were the participants’ written output in the forums and chat sessions.
4.1.5. In-class Rewrite from Memory assignment as a data collection

I noted that since the GWR was a group effort involving the three participants, attributing any general trends in MDM use found in that report to any one member would be problematic. In the case that a member of a group did not contribute equally (or at all) to the composition, trends in MDM use would be erroneously attributed to that member; and this would affect my interpretation of the data.

An individual in-class “Rewrite from Memory” assignment was given on the last day of class. Participants were not told about the task beforehand and were given up to one hour to complete the rewrite. It was considered that a rewrite with similar MDM selection and distribution would suggest that participants, not only had closely read the submitted report, but had also contributed to it. Conversely, participants who had not contributed to the (GWR) would be unlikely to be familiar enough with the submitted draft to produce a rewritten report that approximated any section of the group effort.

This procedure made it possible for the strands of the individual contributions to be separated during a later analysis, thereby making it possible to compare MDM use in forums and chat rooms with the work on the (GWR) for which they had been individually responsible. This could then help to answer RQ1 regarding how the OTREs had influence the frequency of MDMs use and to answer RQ2 concerning whether the OTRE work had guided participants to use the MDMs appropriately, i.e., in a way consistent with their ‘community of practice’ (Lave and Wenger, 1991) (see Section 2.3).

4.1.6. Confirming MDM genre

An issue arose about whether some MDMs should be taught as being more commonly used in written or spoken form, or equally in both. In order to investigate this matter, three raters assessed the entire MDM List in terms of whether each word or phrase was considered to be nearly always used in only spoken discourse or written discourse or was often used in both.

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Both</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>152</td>
<td>45</td>
<td>81</td>
</tr>
<tr>
<td>Rater 2</td>
<td>149</td>
<td>45</td>
<td>84</td>
</tr>
<tr>
<td>Rater 3</td>
<td>152</td>
<td>48</td>
<td>78</td>
</tr>
</tbody>
</table>

Table 4.3: Raters’ distribution of MDM List items as spoken, written or both
These results above (see Table 4.3) were analysed using the IBM SPSS (Statistical Package for Social Sciences, version 27) application for statistical analysis. They show that 45% of the 278 items were considered by raters to be used in both writing and speech. The Master MDM List did include both predominantly spoken and written MDMs. However, because the OTRE reconstruction texts were written to be understood as written texts and not as written forms of speech, nearly all of the words and phrases fronted in the OTREs were ones that fell within the category of MDMs nearly exclusively written; or MDMs which are frequently written but also frequently spoken, but not within the category of MDMs which are almost exclusively spoken. Cohen’s Kappa was used because of its appropriateness in cases where a large amount of data was involved and its effectiveness in validating the percentage of agreement between two raters (McHugh, 2012). The results for Cohen’s Kappa are displayed below, group by the three categories: spoken, spoken and written, and written; and within each category the kappa for each rater (IR):

<table>
<thead>
<tr>
<th>Category</th>
<th>IR 1 and IR 2</th>
<th>IR 1 and IR 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoken</td>
<td>93.5%</td>
<td>81.1%</td>
</tr>
<tr>
<td>Both spoken and written</td>
<td>86.7%</td>
<td>57.4%</td>
</tr>
<tr>
<td>Written</td>
<td>94%</td>
<td>83.3%</td>
</tr>
</tbody>
</table>

Table 4.4: Cohen’s Kappa for raters’ categorisation of MDMs as spoken or written

Overall, the average of interrater reliability was 82.6 which was within an acceptable level of agreement. This indicates that the MDMs used in the OTREs, were largely correctly categorised.

An inspection of the total written output of three randomly selected participants showed that between 53% and 56% of the items from the MDM List which were used fell within the categories of Both and Written with 58% of those being from the Written category. While the forums and chat rooms were conducted under the understanding that text-mediated forum and chat communication exists in a grey area between writing and speech, it was an issue that was commented upon afterwards and subsequent iterations of this study should take this distinction into greater account.
4.1.7. Types of quantitative data collected
In this study, seven kinds of quantitative data relating to MDM use were collected and are as follows:

- Total Word Count
- MDM Count
- MDM Word Count
- Word-length of MDM Phrases
- Percentage of Total Word Count Used in MDMs
- Unique uses of MDMs (as contrasted with repeated uses of one)
- Number of Communicative Functions used

The term \textit{MDM} in this study refers exclusively to MDMs used to signal premises and conclusions. The \textit{Total Word Count} was recorded, on a one-word-to-one-unit basis, so trends of the amount of output could later be established. To understand this count and the trend represented over the course of the term, that figure needed to be considered in light of other quantitative aspects. Changes in measurements like \textit{Percentage of Total Word Count Used in MDMs} could not be understood well without knowing the total number of words. For example, if a participant used three MDMs in a forum post totalling twelve words and another participant used fifteen MDMs in a forum post of 100 words, then the percentages alone could be misleading. Knowing the total word count helped create a frame of reference in which the MDM count could be understood.

\textit{Word-length of MDM Phrases} was also a related measurement. For a complete understanding of how MDM use was analysed quantitatively this needed to be considered. For example, if within a twenty-seven-word post, two MDMs, one four-word and one five-word were used, then this would result in a measurement of 33\% of the total word count, even though only two MDMs were used. Conversely, the use of five single-word MDMs would be counted as only roughly 15\% percent of the total.

While longer phrases between three and five words in length are not always more syntactically complex than single-word MDMs and do not always require greater skill to use, their use by participants merits noting. In the forums, where participants had access to the Master MDM List, the use of longer and unfamiliar MDMs, where a single-word well-known one would have sufficed, indicates additional effort to refer to the document. This suggests a possibly greater level of that engagement, motivation and metalinguistic awareness at the time of that MDM use. In the chat rooms, because participants were not allowed to refer to the
Master MDM List, this even more strongly suggests engagement, motivation and metalinguistic awareness, because longer and more unfamiliar phrases are more likely to have required more effort to recall than single-word MDMs. Thus, the recording and presentation of MDM word length was merited as it both nuanced the other aforementioned figures and suggested varying levels of engagement, motivation and metalinguistic awareness.

In addition to the data explained above, I also recorded the following two general data:

- Unique MDMs
- Premise/Conclusion indicator-relevant communicative functions

Unique use of MDMs was tracked in order to differentiate repeated or habitual use of a usually single MDM, e.g., but or since, from more varied and arguably exploratory, purposeful and metalinguistically aware uses of several MDMs, e.g., however or although, within each writing task. For the purposes of this study, relevant communicative functions are those functions in which explicitly connective or implicitly connective MDMs are used in such a way as to signal premises or conclusions in the context of AW. This includes functions such as clarification or creating focus where signalling argument indicators may not have been the primary communication goal. Variance of communicative functions was also tracked as a way of quantitatively assessing the well-roundedness of the argumentation applied in student written output in forums, chat rooms, and the Group Report. I use the phrase well-rounded to describe text passages in which participants approached an issue through multiple communicative functions. Some participants wrote rather discursively and off-topic; or they repeated points using alternative MDMs to the same purpose as in a previous sentence. Variation in relevant communicative functions, in contrast, was associated with a more well-rounded approach to the topic being discussed.

Data that I generated involved two major themes in this study: conditional statements and hypothetical premises. Uses of MDMs related to these two themes were tracked throughout the term, but were considered most relevant in Forums III and IV, the Group Report and the In-Class Rewrite, all of which involved the Cups and Coins class project and the formation of predictions of outcomes.

- Implicitly Connective and Explicitly connective Premise and Conclusion Indicators
- Implicitly Connective and Explicitly connective Conditional Statements and Hypothetical Premises
As noted earlier, I tracked the use of MDMs used to signal premises and conclusions as well as conditional statements and hypothetical premises. Although students were trained through the OTREs to use explicitly connective indicators, e.g., because and so and the if-then construction of hypothetical premises, they often used implicitly connective indicators such as the main point is, or as discussed later, novel approaches to constructing hypothetical premises. The OTREs had presented mainly explicitly connective indicators. Therefore, differentiating between use of explicitly connective and implicitly connective MDM use provided an indication of how much of a role the OTREs played in influencing and guiding participants’ MDM selection. Cross-referencing this data with the data from Unique MDM Use nuanced the interpretation of this data related to premise and conclusion indicators. Because only three forms of explicitly connective hypothetical premises were presented, such cross-referencing was somewhat less indicative of the OTRE’s influence.

4.1.8. Categories of argument indicators

Two separate groups of argument indicators used in this study consisted of (1) Premise and Conclusion Indicators and (2) Conditional Statements and Hypothetical Premises. The titles, descriptions and examples for the Premise and Conclusion category are listed in Table 4.4:
Explicitly Connective Premise Indicators (EC-PCI)

Indicator: Explicitly Connective Premise Indicators (EC-PI)
Description: Phrases that denote causative or logical links between beliefs and reasons
Example: because and since

Explicitly Connective Conclusion Indicators (EC-CI)
Description: Phrases that express common logical links between reasons and beliefs.
Example: So, therefore and for that reason

Implicitly Connective Premise Indicators (IC-PCI)
Indicator: Levels of Conviction
Description: Refers to phrases that indicate writer certainty that a conclusion is true
Example: e.g. weak, e.g., perhaps; medium, e.g. probably and I think; and strong,
Indicator: Agreement and Disagreement
Description: Refers to phrases expressing agreement with a premise or conclusion
Example: (For reason x) I agree that smoking should not be allowed in public buildings.
Indicator: Reservations
Description: Refers the writer is uncertain with the same conclusion as their readers
Example: What you claim may be true, but . . . (contradicting premise)
Indicator: Creating Focus
Description: Refers to phrases that writer uses to express important points.
Example: We must keep in mind that . . .

Table 4.5: Premise and Conclusion Indicators

The titles, descriptions and examples for the Conditional Statements and Hypothetical Premises category are listed in Table 4.5. Originally this study intended to count only explicitly connective premise indicators such as since and because; and explicitly connective conclusion indicators such as so and therefore. However, during the analysis it became apparent that some participants were using other phrases, with varying degrees of appropriacy, to indicate premises and conclusions. Other communicative functions, such as Creating Focus, Interjections and Clarification and Confirmation; often accrete around the central structure of an argument (Intaraprawat and Steffensen, 1995). MDMs having a primary function of conveying interactional content were often employed by participants to signpost textual features, e.g., premises and conclusions. These interactions were categorised as implicitly connective premise indicators and conclusion indicators (IC-PIs and CIs).
Table 4.6: Conditional statements and hypothetical premises

<table>
<thead>
<tr>
<th>Implicitly Connective Conditional Statements and Hypothetical Premises</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicator:</strong> Implicitly Connective Conditional Statements (IC-CS)</td>
</tr>
<tr>
<td><strong>Description:</strong> Present tense phrases from which coincidence can be inferred</td>
</tr>
<tr>
<td><strong>Example:</strong> <em>On Tuesdays, the price is reduced by 15%.</em></td>
</tr>
<tr>
<td><strong>Indicator:</strong> Implicitly Connective Hypothetical Premises (IC-HP)</td>
</tr>
<tr>
<td><strong>Description:</strong> Future tense phrases from which causality can be inferred</td>
</tr>
<tr>
<td><strong>Example:</strong> <em>If I don’t change cups, the probability is 1/6.</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explicitly Connective Conditional Statements and Hypothetical Premises</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicator:</strong> Explicitly Connective Conditional Statements (EC-CS)</td>
</tr>
<tr>
<td><strong>Description:</strong> Coincidence between the first and second clauses is clearly suggested</td>
</tr>
<tr>
<td><strong>Example:</strong> When I remove the air from the jar, the candle flame dies.</td>
</tr>
<tr>
<td><strong>Indicator:</strong> Explicitly Connective Conditional Statements (EC-CS)</td>
</tr>
<tr>
<td><strong>Description:</strong> I call this a <em>conditional imperative</em>, containing phrases like <em>have to</em> or <em>should</em>; a statement that in order to achieve <em>x</em>, you must first do <em>y.</em></td>
</tr>
<tr>
<td><strong>Example:</strong> If you want to get there by 3:00 p.m., then you need take the express.</td>
</tr>
<tr>
<td><strong>Indicator:</strong> Explicitly Connective Hypothetical Premises (EC-HP)</td>
</tr>
<tr>
<td><strong>Description:</strong> I call this a <em>conditional potentiality</em>. A given condition allows certain outcomes.</td>
</tr>
<tr>
<td><strong>Example:</strong> <em>When we study alone, we cannot get other students’ responses.</em></td>
</tr>
<tr>
<td><strong>Indicator:</strong> Explicitly Connective Hypothetical Premises (EC-HP)</td>
</tr>
<tr>
<td><strong>Description:</strong> There is a clear claim of causality between the condition expressed in the present tense and the result expressed in the future tense.</td>
</tr>
<tr>
<td><strong>Example:</strong> If I remove the air from the jar, then the candle flame will die.</td>
</tr>
</tbody>
</table>

### 4.1.9. Calculation of MDM use

To calculate the frequency of MDM use, an Excel file (see Figure 4.1) was used for each participant. By inputting the quantitative data, across the horizontal axis, whilst listing the nine data sources, top to bottom along the vertical axis, a file was created to generate the aforementioned bar charts presented in Chapters 5 and 6.
Figure 4.1: Raw data for one participant’s MDM use

The category titles and descriptions are listed as follows in Table 4.6:

<table>
<thead>
<tr>
<th>Category</th>
<th>MDM count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>This is the number of MDMs used, irrespective of the number of words used in a given MDM.</td>
</tr>
<tr>
<td>Category</td>
<td>MDM Word Count</td>
</tr>
<tr>
<td>Description</td>
<td>This is the number of words dedicated to MDM use in each writing activity.</td>
</tr>
<tr>
<td>Category</td>
<td>Total number of words</td>
</tr>
<tr>
<td>Description</td>
<td>This is the total number of words used in each writing activity.</td>
</tr>
<tr>
<td>Category</td>
<td>Percentage</td>
</tr>
<tr>
<td>Description</td>
<td>This is the percentage of words dedicated to MDM use in each writing activity.</td>
</tr>
<tr>
<td>Category</td>
<td>Unique MDMs</td>
</tr>
<tr>
<td>Description</td>
<td>This is the number of different MDMs used irrespective of how many times each was used. This could then be used to describe how many different MDMs were used within isolated categories and I could track any increase in the variety of MDMs used.</td>
</tr>
<tr>
<td>Category</td>
<td>Number of categories</td>
</tr>
<tr>
<td>Description</td>
<td>This is the number of categories of dialogic discourse in which participants engaged. I could track any increase in the variety of categories covered in participants’ writing.</td>
</tr>
<tr>
<td>Category</td>
<td>Number of one-and-two-word MDMs, three-word MDMs, and four-or-more-word MDMs</td>
</tr>
<tr>
<td>Description</td>
<td>As participants work towards competency in using a greater range of MDMs, they encounter progressively longer MDMs on the list. For example, they may begin with using the ubiquitous <em>I think</em>, and progress through <em>In my opinion</em>, and <em>It is my belief that</em>.</td>
</tr>
<tr>
<td>Category</td>
<td>Appropriateness of MDM use</td>
</tr>
<tr>
<td>Description</td>
<td>Using the aforementioned scale of appropriateness (see 4.3.), I input the highest value, between 1 and 4 achieved on a given written activity. This figure was arrived at by averaging my own evaluation with the evaluations of two native English speaker colleagues in the English teachers group at TUS, who had been given a printout of the explanation and had been briefed about its use.</td>
</tr>
</tbody>
</table>

Table 4.7: MDM Category titles and descriptions
In order to analyze the participants’ written output, Roman numerals were used as identifiers. By running search and replace operations, Roman numerals were pasted into text passages before each MDM. Search operations for different Roman numerals could then provide an accurate count for the frequency of MDMs used in each of the different categories and sub-categories.

These identifiers were as follows:

I. i Implicit premise indicators (*transitions, code glosses*)
I. ii Implicit conclusion indicators (*transitions*)
I. iii Additional reasons (*sequencing, evidentials, code glosses*)
II Levels of conviction (*hedges, boosters*)
III Agreement and disagreement (*attitude markers*)
IV Creating focus (*boosters*)
V *Not relevant*
VI Validity
VII-IX *Not relevant*
X Comparison and contrast
XI Implicit conditional statements (*hedges, according to Hyland (1998]*)
XII Implicit hypothetical statements (*hedges, according to Hyland (1998]*)
LCXI.i (LC: logically connective) Explicit conditional statements (*transitions, code glosses*)
LCXI.ii (LC: logically connective) Explicit hypothetical statements (*transitions, code glosses*)
XII Major and minor premises (*hedges, boosters*)
XIII Examples and Illustrations (code glosses)

Note: I had initially used the terms *non-logically connective* and *logically connective* to refer to argument indicators for which I subsequently used the terms *implicitly connective* and *explicitly connective*. The reason for this is that Hyland (2005a) uses the term *logical connective* to refer to MDMs within his category of *transitions*, while I am referencing MDMs functioning as explicitly connective indicators, which are listed in several of Hyland’s categories. To avoid a perception of contradiction, two new terms were used to add clarity, since both explicit and implicit connections between premises and conclusions are logical. The prefix NLC and LC appears in Table 4.5, which displays how MDMs were originally tagged.
4.1.10. Charts displaying MDM use

Charts displayed in this were designed according to the following rationale. For the purposes of analysis, I divided the student written texts into four groups:

- Forums I, II, and V
- Forums III and IV
- Chat Rooms I and II
- Group Report and In-Class Rewrite

The first set of forums were grouped together on the basis that they were assigned as responses to textbook text passage readings of critical discussion topics. The type of written output involved extended statements which built up arguments. It also included exchanges between students. The second grouping involved responses to the task I had set in relation to the Cups and Coins class project. Here students were to tell their opinion of what they expected the outcome of the first and second (Forum IV) experiments. Because of the way I had set the task perhaps, students did not exchange ideas outside their oral group discussions, but only used the forum space to post their first (Forum IV) and second (Forum IV) attempts at constructing hypothetical premises. Because of these difference in output, I have grouped the forums in this way. The third grouping, in contrast to the forums, involved in-class synchronous small group chat room discussions, also of textbook fronted topics (previously discussed in forums). It also included the individual in-class rewrite of the Group Report. It was included in this group because of the synchronous in-class context and the prohibition of access to the Master MDM List or other notes.

I created charts for each of the above groupings, displaying counts for the following:

- IC and LC PIs and CIs (implicitly connective and explicitly connective premise indicators and conclusion indicators)
- IC and LC Conditional Statements (CSs) and Hypothetical Premises (HPs)
- Word Counts including the following data points:
  - Total Word Count
  - MDM Count
  - MDM Word Count
  - Percentage of Total Word Count Used in MDMs
  - Word-length of MDM phrases
- Appropriacy of MDM use (see Chapter 6)
4.2. Qualitative analysis
This section discusses how qualitative data was drawn from several sources including the transcripts of participants’ talk during OTRE sessions and responses made during semi-structured interviews. This section also discusses how qualitative aspects of participants’ language output gathered from forums and chat rooms as well as in their Group Report and In-Class Rewrite assignments were assessed. While the previously discussed quantitative data collection was aimed at measuring frequency of different aspects of MDM use for the purpose of addressing RQ1, qualitative data was gathered to answer RQ2 by measuring the appropriacy of MDM use. Other qualitative data was gathered into thematic analysis to answer RQ3 and to further nuance answers to RQs 1 and 2.

4.2.1. Qualitative analysis of MDM use
In this section, I discuss how I established a scale for assessing the appropriacy of MDM use, the influence on this study of background literature including similar studies, and practical considerations based on the factors present in a small-scale study. I explain how I used data generated by raters to reach an assessment of the pragmatic quality, or appropriacy, of participant’s MDM use. Next, I write on the thematic analysis of student-student talk which occurred during OTRE sessions. I list the elements for which I searched and other post-priori elements which I discovered, inductively, as emergent themes during the analytic process. I discuss the raters’ use of that list to arrive at a valid description of the themes present there. Finally, I relay participants’ responses to semi-structured interview cues and discuss raters’ role in validating that thematic analysis. The resulting interpretation supports an answer to RQ2 as it suggests the degree to which OTREs guided participants to use MDMs more appropriately.

4.2.2. Establishing a scale of appropriacy of MDM use
MDM use was also evaluated qualitatively to assess appropriacy. In this way, I could apply my data analysis toward answering RQ2, that is, whether OTREs were useful in guiding participants to use MDMs more appropriately. Past observations on my part have indicated that students sometimes use selected MDMs in ways that do not connect pragmatically with the ongoing discourse. To respond to RQ2, it was necessary to develop a scale of appropriacy. From years of teaching writing to EFL students, I had reached some informed and reasoned understanding of the stratification of quality in student writing. I also had the experience in assessing student writing to make an informed estimate about the degree to which raters were likely to be capable of drawing fine lines between appropriate use of
MDMs. Because the question of appropriate use is somewhat subjective, I considered that a higher number of divisions in my scale would create undue fatigue in my raters which might adversely affect their assessment. Moreover, because of the low number of raters in this small-scale study, their assessments were likely to vary beyond the degree to which such a highly stratified scale could be regarded as useful.

Previous similar studies suggested varied scale designs. Cheng and Steffensen (1996) involved 40 raters and were able to use a scale with 12 divisions between A+ and F, Sanford (2012) involved three raters and four-division scales. Noble (2010) likewise involved a scale with four divisions: High Distinction, Distinction, Credit and Pass. On this basis, I devised a scale with the titles of Unskilled, Average, Skilled and Exceptional with the respective numerical value of 1 to 4 to create clear distinctions between divisions (see Table 4.7) and advised raters that in cases where there was some ambiguity in the participant-writer’s intent, or a mixture of merit and demerit, a .5 could be added or subtracted, but generally, a single digit assessment would suffice. After the assessments were completed, following Cheng and Steffensen (1996), these numeric assessments from the three raters were combined and averaged for each text. Noble’s (2010) criteria involved clear argumentation, logical structure, cohesion, evidence and grammar. My criteria involved the first three of these. An appropriately used MDM would make logical sense, contribute to the clarity of the argument in which it was employed and help to create textual cohesion (Hyland and, 1976).

In these ways, background literature was drawn upon to develop an approach to collection of quantitative and qualitative data and later during semi-structured interviews participants and I could form more nuanced questions regarding the situationality (Sanford, 2012) of any specific case.
Table 4.8: MDM Appropriacy Scale

4.2.3. Independent assessment of appropriacy of MDM use

The assessment for appropriacy of MDMs use underwent a four-step process. I had previously analysed the participants’ MDM use in terms of frequency and had marked each by pasting in a Roman numeral to identify communicative function. The raters could then note the location of each MDM which required assessment. See Figure 4.2.
Figure 4.2: An example of an independent reader’s assessment

A total of three native English speakers, one British and two American, one of those being myself, served as raters who read and rated the participants’ forum, chat room, Group Report and In-Class Rewrite output. The independent raters looked at essentially 100% of the portions of the total output which involved MDM use in the forums, chat rooms, and rated the MDM use by using the scale shown in Table 4.7.

Ratings for each of the nine output sources were averaged separately. These totals were used as the source that automatically generated the charts presented in Chapters 6. Interrater reliability was calculated using the SPSS (Statistical Package for Social Sciences, version 27) for Rater 1 (myself) and Rater 2; and for Rater 1 (myself) and Rater 3. Cronbach’s Alpha was chosen as it is frequently used for Likert type rating systems.

In cases of disagreement in the assessment, raters were consulted to re-evaluate an MDM use in view of factors they might not have considered. For example, if the MDM use merited a high score (considering it was a word that would not have been treated with instruction at the high school level; or, in the case of a word or phrase to which they had given a low score) evaluators were asked to consider how that MDM had linked preceding and subsequent MDMs to create a system of meaning that spanned a paragraph; so that even though the MDM itself was not uncommon, it nevertheless merited a higher score than it had received.
4.2.4. Thematic analysis
Thematic analysis was performed on both the transcripts of the student-student talk during the OTRE sessions and on the interview responses. To distinguish between data referring to the former and the latter, henceforth, I use the term *utterance* to refer to spoken data collected from the OTRE session transcripts and I use the term *response* to refer to spoken data collected from the semi-structured interviews. The approach can be described first as thematic, as this form of qualitative analysis can be used for data sets involving small numbers of participants and can be applied to a variety of types of data (Clarke and Braun, 2017). I followed the seven steps outline in Kuckartz (2014) first, noting utterances (from OTRE sessions) or responses (from semi-structured interviews) that I viewed as important. Second, from these I developed a preliminary list of topics and categories that resonated with the research questions. Third, I coded them, identifying basic units for the larger blocks of meaning (Clarke and Braun, 2017), using the coding method discussed in Section 4.2.7. Fourth, I searched out other similar utterances from OTRE sessions and grouped those according to those categories. I later performed the same operation on the responses from semi-structured interviews. Fifth, in some cases, I identified sub-categories, according to the research questions. Sixth, I recoded the data and further prepared it for writing up. Seventh, I carried out analysis and wrote the analysis presented in this thesis (Kuckartz, 2014). Thus, both deductive methods, based in the theoretical issues and inductive methods, that were in the data itself were used in what Swain (2018) refers to as a hybrid approach to thematic analysis. Following Kuckartz (2014), units of measurement included full sentence and complete thoughts. Any pertinent reflections on my part were also recorded along with the data. As Kuckartz (2014) notes, coding by one person is not optimum but is often necessary, for example, in a dissertation where, as with my own circumstance, where resources and outside assistance are limited. Coders need to be apprised of the purpose and meaning of what they are being asked to do. However, extensive theoretical knowledge on their part is not considered to be necessary Kuckartz (2014).

4.2.5. Thematic analysis of transcripts of student-student talk during OTRE sessions
Transcripts were analysed through thematic analysis (Guest et al., 2012). The data from the transcripts were generated from student-student talk from the three triads during five OTRE sessions that were assigned in class throughout the first two-thirds of the term. *A priori* codes (Swain, 2018) were those which were presumed would be present based on a reasoned consideration of the context. Also, the number of themes that emerged during this analysis
are grouped as *a posteriori* (Swain, 2018). The codes used for both groups are as follows:

- Proposing an answer (to the multiple-choice array)
- Discussing the text
- Discussing OTRE procedure
- Reading the text aloud
- Strategy
- Noticing of MDMs

A posteriori codes were as follows:

- Disagreement / Doubt
- Fatigue
- Commiseration/Encouragement
- Concern over choice/confirmation (seeking and giving)
- Expectation/Disappointment
- Satisfaction/Excitement
- Interest in maintaining the score (displayed on the OTRE)
- Mentions of remaining time/work

Using high quality headphones, the audio recordings of the OTRE sessions were reviewed and transcribed by a trained native Japanese speaker. I then reviewed the transcripts (see Figure 4.3) to standardise the spelling and other points of presentation. In a few instances, parts of English pronunciation were not clear and needed to be transcribed by myself. In these cases, both the transcriptions and their recordings were reviewed simultaneously. However, when crosstalk or murmuring was present, a second native Japanese speaker was asked to catch what had been uttered. At times, the murmuring proved too difficult to be deciphered.
The transcripts were analyzed by myself at which time I verified the presence of the a priori themes. I noted the presence of additional post-priori themes (see Figure 4.4). Based on those findings, I produced the interrating tool below (see Figure 4.5) as a spreadsheet.
<table>
<thead>
<tr>
<th>Discussion of strategy for completing the exercise</th>
<th>Grammar-based / mention of grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Orthographic/punctuation-based, e.g. upper-case letters, comma, etc.</td>
</tr>
<tr>
<td>Disagreement</td>
<td>Correction of misunderstandings</td>
</tr>
<tr>
<td>Explicit recognition or observance of metalinguistic content</td>
<td>Premise MARKER / conclusion MARKER</td>
</tr>
<tr>
<td></td>
<td>Premise / conclusion</td>
</tr>
<tr>
<td></td>
<td>Conditional Statement / C.S. marker</td>
</tr>
<tr>
<td></td>
<td>Hypothetical Premise / H.P. marker</td>
</tr>
<tr>
<td></td>
<td>Order of markers</td>
</tr>
<tr>
<td>Fatigue / Mention of remaining time</td>
<td></td>
</tr>
<tr>
<td>Encouragement to continue</td>
<td></td>
</tr>
<tr>
<td>Emotional engagement:</td>
<td></td>
</tr>
<tr>
<td>excitement over correct answer</td>
<td></td>
</tr>
<tr>
<td>/ disappointment over incorrect answer / concern over choice</td>
<td></td>
</tr>
<tr>
<td>Interest in score percentage</td>
<td></td>
</tr>
<tr>
<td>Reading the text aloud / Discussing meaning of the text</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.4: List of themes distributed to each rater

The first column represented macro-themes, and, in some cases, micro-themes were displayed in the second column. I then extend the spreadsheet to the right marking the top vertical row to act as an identifier for the row number corresponding to the transcript (see Figure 4.5).
The three raters, one British and the other American, performed independent thematic analysis on the transcript. Before doing so, I conferred with each rater individually and trained them on the use of the spreadsheet. In short, raters were instructed to familiarise themselves with the relevant MDMs and the themes. They were told to denote the presence of a given theme in the appropriate row and column with a numeral 1. They were advised that single utterances could be considered as representing multiple themes. Therefore, single columns could contain multiple 1s. A column which tallied the number of 1s for each row was later inserted at the end of each session. A total of seven OTRE session transcripts were included so that over 10% of the complete transcript was analyzed. A portion of 10-12% the total available transcript was drawn upon following Campbell, et al. (2013); Cotton and Wilson (2006); Weston, et al. (2001); Hodson (1999), who used 10% of transcripts in their qualitative analyses. While I analyzed all the transcripts, each of the two other raters analyzed transcripts from different OTRE sessions. I later compared each of their ratings to my own. Afterwards, I met with each rater separately and discussed instances of disagreement. In some cases, the reason for disagreement was because either I or the other rater had missed an
utterance conveying a theme. Another reason was that the utterance, being in Japanese, was not understood by one rater. Still another reason was that the number 1 had been mistakenly placed in a different column to that which was intended. Nearly all cases of disagreements were reconciled. In cases where a rater held an opinion about an intended meaning of an utterance, the disagreement was left as is. One example of this occurred when there was a difference in perceived nuance of an utterance between the British rater and myself. Following Sanford (2012) the percentage of agreement was calculated by dividing the number of themes in which the raters had arrived at an agreement by the total number of themes.

4.2.6. Semi-structured interviews
Thematic analysis of interview transcripts was used to answer the research questions of the study. Compared to other approaches, like discourse analysis and grounded theory, thematic analysis could be used without deep knowledge and long training (Braun and Clarke, 2006). Also, it is a malleable approach and could be applied to meet the requirements of the content (Braun and Clarke, 2006).

In my preparation of questions, my carrying out of the interviews, and my subsequent analysis, I applied an interpretive approach and to allow participants their voice and agency in that interpretation, while avoiding an uncritical acceptance of my presentation of their responses as a direct transmission of their inner mental life (Braun and Clarke, 2006). Thus, my approach was ungrounded, exploratory and content-driven, rather than hypothesis-driven (Guest et al., 2012). The interview questions were aimed at discovery of themes related to my research questions, rather than producing ‘rich description of the data overall’ (Braun and Clarke, 2006, p. 12). The themes were viewed as emerging through the identification of patterns, as I analysed and reflected upon what had become salient and returned to the content for further analysis (Braun and Clarke, 2006).

Unlike the quantitative approach that I applied to the participants’ written output, my intent here was to explain that quantitative data through the qualitative data. I searched for the expression of generalities (Guest et al., 2012) across the data set (Braun and Clarke, 2006), i.e., all nine participants’ responses. However, I sometimes shifted the focus to one participant and triangulated specific statements, or what I interpreted as general predispositions, with other data sets, such as MDM frequencies or OTRE session transcripts. For RQs 1 and 2, this was performed by asking participants probing questions concerning if and how they were influenced and guided to use MDMs more frequently, with greater variety, and at higher
levels of appropriacy. In addition, thematic analysis was used to answer RQ3. Participants’ comments illuminated what might have otherwise remained obscure aspects of their perceptions of their experiences in the forums and chat rooms, Group Report and In-Class Rewrite.

4.2.7. Themes presented in the interview
Through the coding process, the main themes detected from the transcript were categorised. In addition to grammar-based strategies, other themes were noted. They are presented in Figure 4.6. The organization shown in Figure 4.6 was reproduced on a spreadsheet with the uppermost row numbered from 1 into the hundreds based on the length of the interview. Each column corresponded to a line from the interview transcript. As raters coded the transcript, they entered the numeral 1 into the corresponding cell when they detected one or more themes. Each interview session progressed from left to the right and ended with a column in which the total number of instances were tallied.

I analysed all the transcript and prepared a copy which included the portion of the transcripts which related to the interview cues. One other rater, a native-English speaking colleague at TUS, analysed the entirety of the prepared document using the coding tool which I had prepared. The rater was trained in the use of the coding tool (see Figure 4.6) and the meaning and significance of the themes. Both I and the other rater independently coded the transcript placing numeral 1s in the appropriate rows and columns. Afterwards, I met the rater and discussed instances of disagreement. Cases of disagreements which included mistakenly placed numeral 1s and understandings of Japanese responses, were reconciled. Again, following Sanford (2012), the percentage of agreement was calculated by dividing the number of themes agreed by the raters by the total number of themes. The results of the interrating are presented in 5.6.2.
Figure 4.6: Tool used for thematic analysis of interview responses

4.2.8. Establishing a scale of appropriacy of MDM use

Subsequent to the aforementioned independent assessment and during closer inspection of participants’ attempts at multi-clause-structured hypothetical premises, a theme emerged in which different approaches on the part of the participants became salient. The noticing of these differences led to their categorization. I later came to view these categories hierarchically in terms of relative similarity to the form modelled in OTRE Sessions 4 and 5.

Participants had been given instruction on the distinction between conditional statements and hypothetical premises via podcast during the final two OTRE sessions dedicated specifically to teaching the following directive: hypothetical premises are to be written in a single-sentence two-clause *If-Then* structure where the *If* clause is in the present
tense and the then clause is in the future tense, generated through phrases such as will+verb or until. Note that hypothetical premises are a kind of conditional statement, but that not all conditional statements are hypothetical (Peters, 2014). For example, Every time I eat at that restaurant, I get sick is a conditional statement, whereas, If I eat at that restaurant, I will get sick is a hypothetical premise which can be used to test the veracity of the first claim.

During the OTRE sessions students were tasked to identify and note conditional statements and hypothetical premises and to distinguish between the two. Subsequently, we revisited the reconstructed texts in a teacher-led class discussion in which I stressed the importance of science students developing this writing skill so that they could later state their hypotheses in their academic and professional writing in line with the writing practices of their discourse communities. I informed students that this construction was to be used in a Group Report on a class project involving forming hypotheses, collecting data to test those hypotheses, and reporting on the results.

Later, during discussions with two colleagues, the following scale (see Table 4.8) was developed, based on perceived levels of differences from the prescribed level-4 model. In this scale a 1 denotes a general failure to recognise or understand the task. Level 2 and the sub-levels within Level 2 denote: a word or phrase such as When being used in the first clause, which is more appropriate to general conditional statements rather than specifically predictive hypotheses; and/or the following in second clause: a present tense (rather than a future tense) construction; future tense suggested through the use of words and phrases that reference outcomes such as probability becomes or prediction is; or inappropriate modal constructions, such as would rather than the prescribed will.

<table>
<thead>
<tr>
<th>Clause 1</th>
<th>Clause 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Present tense: If . . . then Future tense will be</td>
</tr>
<tr>
<td>3</td>
<td>Present tense: If . . . then Future tense (incomplete construction)</td>
</tr>
<tr>
<td>2.75</td>
<td>Present tense: When Modal, e.g. could, would, can, will (but incomplete)</td>
</tr>
<tr>
<td>2.5</td>
<td>Present tense: When Suggested future, e.g. the probability becomes; or the percentage gets</td>
</tr>
<tr>
<td>2</td>
<td>Present tense: When Present tense is, does, incomplete construction</td>
</tr>
<tr>
<td>1</td>
<td>Non-task oriented action: e.g. focuses on method or reports results, rather than advancing a hypothesis</td>
</tr>
</tbody>
</table>

Table 4.9: Scale for assessing the appropriacy of hypothetical premise constructions

Using this scale provided a means of producing a stratified assessment which could later be nuanced by drawing upon OTRE transcripts and interviews.
4.3. Summary

This chapter discussed how both the quantitative and qualitative data were analysed. First, it discussed the way in which data was categorised. Second, it outlined how a variety of aspects and shifts in MDM use were defined and tracked listed. Third, it discussed the rationale for making those measurements and for the units of measurement chosen as well as explained some assumptions about the data and the limitations of what could be inferred. Lastly, this chapter explained how various data points were calculated was explained.

I discussed how qualitative data was generated from transcripts of student-student talk during OTRE sessions and how thematic analysis was performed by raters to verify the presence of a number of themes. I also explained how qualitative data analysis was performed, again, by raters, on participants’ writing in forums, chat rooms, a Group Report and an In-Class Rewrite to assess the level of appropriacy of MDM use. Lastly, I discussed the use of one-on-one semi-structured interviews in which I further responded to both RQ1 and RQ2 and also responded to RQ3 as the participants’ comments helped me to understand the experience of completing and learning from the OTREs, from the participants’ perspectives.

In the following two chapters I discuss the how various aspects of participants’ MDM use were analysed quantitatively. I present the quantitative data derived from that analysis. I then present qualitative data, and discuss levels of appropriacy of MDM use, student-student talk during OTRE sessions, forum and chat room talk, group and individual written reports, and finally, one-on-one semi-structured interviews to further nuance the understanding of the findings.
Chapter 5: MDM (token) use by frequency

5.1. Introduction
This chapter reports on the data which responds to RQ1 of this study: To what extent does completing OTREs and particularly WebSequiturs influence language learners to use selected MDMs in their written arguments? The most noteworthy findings relate to previous studies on students’ uses of MDMs. I noted, in Chapter 1, the research that has been done in MDM use in argumentative writing. I also noted the differences between those studies and my own and pointed out the gaps which this study goes some way in filling. There are few cases that very closely resemble the context in which this one took place, with important differences including native English speaker status; whether the study took place in an EFL or ESL context; the educational level, as in Khedri, Heng and Ebrahimi (2013); the level of language competence, as with Dabaghi, Zabihi, and Rezazadeh (2012); the ages of the students, as in Intaraprawat and Steffensen (1995); and the writing genre, as in Sanford (2012), to name several. That said, each of these studies and my own share an interest in metadiscourse markers. The findings of Chapter 5 serve to suggest merit in further investigation into the proposition that OTREs could be implemented by language teachers in various contexts in order to more effectively teach students the value of increasing their use of MDMs.

In this chapter I discuss the various themes that emerged as I analysed both quantitative and qualitative data from a variety of sources. Following an interpretivist model, quantitative data was augmented with qualitative data to draw out contextually meaningful trends in MDM use. To display the quantitative data and convey trends to the reader, I use tables and bar graphs to display language use trends. I spend some time in explaining each set of graphs, beginning from the simplest concepts to establish the validity and reliability of the analysis, to answer the RQs.

Because I am now discussing MDMs in terms of the quantitative analysis, henceforth, I use the term token when discussing the MDMs in the context of this study. I continue to use the term MDM when referring to the use of metadiscourse markers in general.

5.2. Data Sources and frames
In the following sections, I present data showing the written output of the nine participants in this study, in terms of the number of words they used in nine writing activities to identify trends in the frequency of token use. I grouped these nine activities into four frames (see Table 5.1). The first of these frames is three out-of-class online forums. These are Forums I, II
and V, the subject of which was social issues presented in the selected course textbook. The second of these frames includes Forums III and IV, which concerned probability and required formation of hypothetical premises both before and after podcasted instruction (via Critically Minded Podcast) and subsequent WebSequitur activities. The subject of these hypotheses was the Cups and Coins Enquiry into Probability. The third frame includes two timed in-class chat room discussions. The subject of these chats was identical to the issues discussed in the previous forum. The fourth frame involves an out-of-class Group Written Report (GWR) on the same subject, completed via online MS Word documents shared amongst each triad; and a similar individual In-Class Rewrite (ICR) which consisted of participants writing as much of their final GWR as each participant could recall; using the same tokens as those used in their GWR and Forums III and IV (Frame 2).

<table>
<thead>
<tr>
<th>Frame</th>
<th>Data source</th>
<th>Interaction</th>
<th>Subject/activity</th>
<th>Language focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Forums I, II and V</td>
<td>Online (asynchronous)</td>
<td>Social issues</td>
<td>Premises and conclusions</td>
</tr>
<tr>
<td>2</td>
<td>Forums III and IV</td>
<td>Online (asynchronous)</td>
<td>Probability</td>
<td>Hypothetical premises</td>
</tr>
<tr>
<td>3</td>
<td>Chat rooms I and II</td>
<td>In-class (synchronous)</td>
<td>Social issues</td>
<td>Premises and conclusions</td>
</tr>
<tr>
<td>4</td>
<td>GWP and ICR</td>
<td>Online (asynchronous) and In-class (synchronous)</td>
<td>Probability</td>
<td>Hypothetical premises</td>
</tr>
</tbody>
</table>

Table 5.1: Frames and data sources

In this way, the nine writing tasks were grouped into four frames. Frame 1 was out-of-class, asynchronous and invited large group discussion. Frame 2 was out-of-class and asynchronous, but essentially required completing a task to produce hypotheses. Frame 3 was timed and synchronous and involved a conversational tone between fixed interlocutors in a small group. Frame 4 involved a traditional Group Written Report (GWR) and a later Rewrite from Memory task, both of which were one-way communication between writers and conceptualised readers.

In each frame, I first provide an average of the nine participants’ token use. In most cases this method of initial presentation is borne out by most members’ individual performances. Otherwise, the bulk of the data would be too unwieldy to present. When the average is not representative of the whole, and where, instead, there are groupings of different
trends, I also provide an additional graphic display in the form of a bar graph of one representative example for each grouping.

5.1.1. Quantitative data
In this section, I discuss the data that I gathered from a variety of quantitative data sets. First, I present the data. Second, I discuss the general view of the average of the nine participants’ quantity of written output and frequency of token use and provide a visual representation in bar graph form. Then I determine any cases of groupings of trends which are consistent with or contrastive (Denscombe, 2014) with the average. Excepting the first sets of data reporting total word counts, all data has been standardised per 100 word and presented with other including statistical figures, e.g., mean, standard deviation, range, in table format.

Indicators are abbreviated as follows:
- Implicitly Connective Premise Indicators (IC-PI)
- Implicitly Connective Conclusion Indicators (IC-CI)
- Explicitly Connective Premise Indicators (EC-PI)
- Explicitly Connective Conclusion Indicators (EC-CI)
- Implicitly Connective Conditional Statements (IC-CS)
- Implicitly Connective Hypothetical Premises (IC-HP)
- Explicitly Connective Conditional Statements (EC-CS)
- Explicitly Connective Hypothetical Premises (EC-HP)

Sources of the written output are abbreviated as follows:
- P: Participant (e.g. Participant 1 is P1)
- F: Forum
- CR: Chat Room (and alternately C: Chat)
- GWR: Group Written Report
- ICR: In-Class Rewrite

Where relevant, I intersperse these graphs with observations on student-student talk during the OTRE sessions, forums, chat rooms, and GWR and ICR to support the quantitative data with qualitative data.
5.2. Total word output by participants
In this section, I discuss participants’ written output in terms of total word count for each of the nine writing tasks in order to form and maintain a perspective on each participant’s token use. It is important to note that factors such as language proficiency, motivation, social skills and interest in the forum topic may have led some participants to produce different amounts of total written output in a given forum or chat room. The purpose of this section is to establish that: (1) participants understood the purpose of the OTREs; and (2) participants were, for the most part, evenly matched in terms of willingness or eagerness to participate in the forums, chat rooms, and final report.

5.2.1. Participants’ reported understanding of the purpose of the OTREs
In order to better understand the data which follows, I asked participants to state, in their own words what they thought the purpose of the OTRE sessions were. Some of their answers were as follows:

- P1: To cooperate with my group members and to get the skills to discuss.  
  Gann: That’s very interesting. Could you comment on the content of the language? What kinds of things were being targeted?  
  P1: To get a skill to tell one’s opinion more persuasive.

- P2: To use languages we learned at home, and teichaku—  
  Gann: Teichaku?  
  P2: Teichaku . . . To take in.

- P4: The purpose is thinking the argument of language and we can think about the social problems for example . . . [and continued citing the forum topics and group project.]

These participants’ responses demonstrate that the general purpose of the activities was sufficiently comprehended. Participants understood that they were not being assessed during the OTRE sessions, but were undergoing a restructuring of their awareness and understanding of the language units being presented.

I asked them about the purpose of my using the software application.

- P5: It’s to make easy to . . . like imitating or manesuru (to imitate)  
  Gann: Imitating what? It’s very interesting.  
  P5: Before the class, we usually read the example sentences and in the class, we imitate the sentences we have read. [Trans.] Solve the problem by imitating
the sentence you read before class. Applying the shape . . . So, consciously imitating the shape means consciously using the shape, so it can be used more practically. I think you can learn through practical usage.

- P9: . . . after this class, we can’t talk about complex argument without indicators. We did WebSequiturs with two other students, so the purpose to do WebSequiturs were, I think, with friends we consult with partners, so only one can’t answer the question but with two other students we can answer the questions.

These participants’ responses, which are representative of the nine, demonstrate that they understood both the significance of the procedure for completing the OTREs and the importance of the social dimension of the activities’ collaborative nature.

I also asked participants for an example of kinds of communication they used when doing OTREs in their triads.

- P9: [Trans] I have about 4 choices, maybe 2? [referring to the WebSequitur] Even though I thought [something], the other two people said [another thing]. I thought about this, after saying the reason properly, and because there was a reason, I thought that it might be so. You can do it. Something like that . . .

Gann: Can you tell me how well you remember specific language points of the exercises?

P9: I said a moment ago to talk about complex problems or for example politics problems or cultures. I said before this class gesture was important, communicate, but deep conversations . . . we can’t without indicators.

These responses establish that participants not only understood the importance of the social dimension of the activities’ collaborative nature, but that they had been able to act on that understanding through meaningful and supportive communication with their group members.

5.2.2. Forums I, II and V

In this section, I present the data from the first frame of participants’ written output sources. The word count for Forum I establishes a baseline for quantity of output prior to the first OTRE session. Table 5.2, below, shows that the average word count was around 140 words in Forum I. The average word count generated by Participants 1-9, in Forums II decreased somewhat after the first OTRE and remained at that level later in Forum V. Perhaps, greater attentional resources were expended after the goal of using tokens had been explained,
resulting in fewer words being written.

Participants fell into three groups: those who wrote a low number of words (fewer than 55 words) (P2, P5 and P8); those who wrote a middle range number of words (between 56 and 180 words) (P1, P3, P4, P7 and P9); and P6 who wrote more than 400 words (see Table 5.2). I chose these distinctions based on the groupings of participants’ written output which I observed. Although participants’ written output in the first forum varied in terms of word count: in the second forum; they tended to converge at around 100 words within a range of about 150 words between the high and low word counts. In the second forum, there was a general decrease in the word count.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Forum I</th>
<th>Forum II</th>
<th>Forum V</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>172</td>
<td>154</td>
<td>33</td>
</tr>
<tr>
<td>P2</td>
<td>51</td>
<td>46</td>
<td>55</td>
</tr>
<tr>
<td>P3</td>
<td>151</td>
<td>73</td>
<td>47</td>
</tr>
<tr>
<td>P4</td>
<td>185</td>
<td>94</td>
<td>34</td>
</tr>
<tr>
<td>P5</td>
<td>120</td>
<td>Ø</td>
<td>59</td>
</tr>
<tr>
<td>P6</td>
<td>405</td>
<td>254</td>
<td>76</td>
</tr>
<tr>
<td>P7</td>
<td>161</td>
<td>152</td>
<td>31</td>
</tr>
<tr>
<td>P8</td>
<td>70</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>P9</td>
<td>125</td>
<td>92</td>
<td>69</td>
</tr>
<tr>
<td>Average</td>
<td>160</td>
<td>108</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 5.2: Number of words by each participant as an overall measure of Forum engagement

Individually, there was a general increase with six of the nine participants increasing their output from between 50-75 words. Similarly, in the fifth forum, on average, participants maintained their output, although due to P5’s and P8’s zero-participation in Forum V, the average was pulled down a small amount. The other participants, excepting P1 and P8, all increased their output by 50-75 words. Overall, notwithstanding P1’s 75-word decrease from Forum I to Forum V; P4’s hundred-word decrease between Forums I and II, the overall picture we see in terms of written output is fairly consistent in each forum. The increase of average token count in Forum V, which was about blood type personality profiling in Japan, may have been the result of interest the topic generated. P6’s comparatively large wordcount was comprised largely of references to personal experience talk or references to real life examples to set up a point that he planned to make. In Forum II, for example P6 wrote:

I personally feel that any word that an overweight person might take the wrong way could be considered fat shaming. Even though we might think that it would be helpful or we are being nice to an overweight person, they could take it the wrong way with what we say or don’t say from negative past experiences. Our body language could also be construed as fat shaming. One thing we should be careful of besides our words is
how we look at someone who is overweight. For instance, when we stare at them, they may think we are looking down at them. Hence, just a small casual action like the above could offend an overweight person. For these reasons, I believe that we should be careful on how we treat overweight people. I believe we should do our best to treat all people the same whether they are normal body weight or overweight. It’s my opinion that we should just be kind to everyone to avoid fat shaming. This is the most important point.

The structure of a sustained line of thinking is present in the block of text (above). Also in Forum V, P6 wrote:

For instance, fortune-telling may not be valid. I sometimes see them in the morning news or in the shrine but I don’t know how they are being decided. Even though it may not be valid, I think [. . .]

More than other participants, P6 adopted a warmer more sociable tone. Also, it was not always the case that P6 wrote in blocks of the size presented above. He also participated by posting medium sized text blocks—still larger than those posted by other participants—and then interacted dialogically with other participants in multiple shorter responses. Other participants were more constrained or economical in their use of words. As P6 was an outlier in this regard, however, these passages and a description on his output in the forums and chat rooms is important in understanding the differences in ratio of token Count to Total Wordcount.

5.2.3. Forums III and IV

In this section, I introduce data concerning the second frame, also forum generated, but directed to the writing task. Table 5.3 (below) displays the nine participants’ average word count in this second frame in Forums III and IV, and shows that participants’ output was more uniform than in Forums I and II, and V. This was likely owing to the perceived simplicity of the issue presented and the fewer words required to complete the task. Eight participants either maintained or, by small amounts, increased their output between the two forums; and five participants increased their output by a low number of words, as with P1 and P7; and a medium number of words as with P2, P3 and P4 and a high number of words, as with P6.
In this section, I present data (see Table 5.4) regarding the quantity of written output participants generated in the third frame (during the first and second chat room discussions). As with participants’ output in the forums and the Group Report, their averaged output in the chat rooms is fairly uniform. P1, P2, P4, P5, P6, and P7 all wrote roughly 100 words, and in Chat Room II, P1, P2, P4, P6, and P7 all increased their output by between 25-75 words.

The following conveys trends found in participants’ individual written output. The contrast between participants’ low wordcount in the forums and their greater output in the chat rooms was an issue in the pilot study. This was an important question because it related directly to the role WebSequiturs might have played in learning.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Forum III</th>
<th>Forum IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>55</td>
<td>122</td>
</tr>
<tr>
<td>P2</td>
<td>36</td>
<td>111</td>
</tr>
<tr>
<td>P3</td>
<td>35</td>
<td>130</td>
</tr>
<tr>
<td>P4</td>
<td>95</td>
<td>138</td>
</tr>
<tr>
<td>P5</td>
<td>55</td>
<td>Ø</td>
</tr>
<tr>
<td>P6</td>
<td>159</td>
<td>328</td>
</tr>
<tr>
<td>P7</td>
<td>55</td>
<td>111</td>
</tr>
<tr>
<td>P8</td>
<td>Ø</td>
<td>91</td>
</tr>
<tr>
<td>P9</td>
<td>74</td>
<td>149</td>
</tr>
<tr>
<td>Average</td>
<td>71</td>
<td>148</td>
</tr>
</tbody>
</table>

Table 5.3: Number of words written by each participant in forums

<table>
<thead>
<tr>
<th>Participant</th>
<th>Chat Room I</th>
<th>Chat Room II</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>77</td>
<td>213</td>
</tr>
<tr>
<td>P2</td>
<td>172</td>
<td>294</td>
</tr>
<tr>
<td>P3</td>
<td>85</td>
<td>44</td>
</tr>
<tr>
<td>P4</td>
<td>185</td>
<td>220</td>
</tr>
<tr>
<td>P5</td>
<td>101</td>
<td>313</td>
</tr>
<tr>
<td>P6</td>
<td>376</td>
<td>812</td>
</tr>
<tr>
<td>P7</td>
<td>77</td>
<td>210</td>
</tr>
<tr>
<td>P8</td>
<td>88</td>
<td>114</td>
</tr>
<tr>
<td>P9</td>
<td>174</td>
<td>223</td>
</tr>
<tr>
<td>Average</td>
<td>148</td>
<td>271</td>
</tr>
</tbody>
</table>

Table 5.4: Number of words written by each participant in Chat Rooms I and II

Word counts and comments from the participants indicate that the immediacy of synchronous chat was an attractive factor in motivating participants to be active in the forums. Without that immediacy, some students felt they were working alone and this could be demotivating. P7 spoke about his non-participation in the forums, contrasted with his above average chat room engagement:
Gann: Here is your first forum one comment, very short and in the second forum you did not participate and in the third forum you did not participate. So, your input is almost nothing. Now in the chatroom I see that you have spoken more and I think that what you say is sometimes very thoughtful, very interesting.

P7: Ahhh yes.

Gann: Yes, sometimes what you’re saying is thoughtful, but can you tell me why was your participation in the forum and your participation in the chat room so different?

P7: To tell the truth, chat room is group talking but the forum is only me. Only my idea. And so . . . if it is group work, it will bother the other two people if you don't say [anything], so I'm trying to do my best.

Gann: I see so you feel a kind of giri [obligation] to participate

P7: Uhh yes.

Gann: Giri ka douki [motivation] dotchi desuka? [Which was it, obligation or motivation?]

P7: Douki. Douki. Yes, but [in the] forum [giri], [Translation] I thought that I would not bother the others in the forum and others, so I didn’t do much.

Despite frequent reminders and encouragement to use the forum as a testing ground for token use, it is evident that it was social pressure and a sense of responsibility to others, which proved to be the determining factor. The fact that the chat was done in-class and the forum was assigned as homework may also have some explanatory power. Students are not losing free time by participating in in-class chat, but during time spent out of class, writing in the forums is seen as time lost from other activities. It may also be that the participants do not have experience discussing topics in multiple thread forums and therefore did not readily engage in them as a matter of course. This recalls Belz’s (2002) observation that different degrees of exposure to forum environments can be a factor in how students respond to being tasked to converse there. Students tend to reply to the teacher and not to each other. Years of training in the public school system’s top down instructional style can make the practice of removing the teacher from the center difficult to change. The above points, which emerged during the semi-structured interviews, indicate that the fluctuations in token use had more to do with the time and location of the communication than with the effectiveness of the WebSequiturs.
To recap this section, participants were happy to engage in text communication during the class period. However, outside the classroom, English study may not have a high priority. Most participants responded to the teachers’ forum questions once and considered their work done; but they often did not see the value of responding to each other’s comments. All participants indicated that they enjoyed the OTRE-work and did not tire of its use over the course of the term.

5.2.5. Group Written Report and In-Class Rewrite

In this section, I report on the fourth frame, which included the Group Written Report (GWR) and the closely related In-Class Rewrite (ICR) which followed it.

The average written output for the GWR was just over 400 words and the about half that for the ICR (see Table 5.5). Individually, in the GWR, word count ranged from 416 to 535 words. In the ICR, output was consistent across participants, except for P9 where the falloff was considerable. In the other cases, excepting P5 (who did not attend class on the day of the ICR task), the output was not low enough to suggest that the GWRs had not been a group effort. Higher output could not be linked to forum participation as the case of P7 demonstrates. While P7’s output in the ICR was the second highest of all nine participants, his forum participation was the second lowest. In Section 5.2.4, he explains that the asynchronous platform and the perception of being alone in the virtual space had been demotivating. Nevertheless, he performed well in terms of word count in the GWR and the ICR. He also was able to engage actively in the chat rooms. This suggests that the OTREs may have effectively equipped P7 with the linguistic tools for engaging in discussion of the assigned issues. In later sections, I will examine P7’s use of tokens in his argumentative writing in order to further explore this issue and address RQ1 and RQ2.

In summary to this section, the total output of the participants demonstrates a sustained and consistent response without which this study might be compromised. Within all the output sources, the quantity of each individual’s output was sufficient to a degree that comparisons about token-related data can gathered and later contributed to the whole of the argument of this thesis.
In this section, I discuss token use in Forum I prior to the first OTRE session and following that; the written data generated during the weeks when the OTRE sessions periodically occurred. I initially discuss token use from Forum I, which was assigned before the first OTRE session. I then discuss token use in Forums II and V which were assigned after the WebSequitur sessions focusing on premise and conclusion indicators had been completed. I explain how Forums I, II and V were tasked as extended discussions of textbook fronted issues. Following this, I report on token use in Forums III and IV, which were assigned as shorter discussions and tasked students to form conditional statements and hypothetical premises. After this, I discuss Chat Room I, Chat Room II. Then I discuss the Group Written Report and the In-Class Rewrite.

This section follows the same general method of presentation as the previous section, as I examine changes in token use. Little can be directly construed from the student-student talk during the OTRE sessions to answer the RQs directly because they did not often speak in metalinguistic terms. Hence, we must first look at changes in the frequency of token use in the forums, chat rooms, Group Reports In-Class Rewrites and then draw upon the student-student talk when applicable. In addition, semi-structured interviews also are an important source for nuancing the quantitative data. These data speak to RQ1 as it is concerned with token use irrespective of the skill with which the tokens were employed. In each assignment I have tracked four data points:

- total word count
- number of tokens (token frequency normalised per 100 words)
- number of words dedicated to token use
- percentage of words dedicated to token use

In this section, I perform a thematic quantitative analysis of nearly all sections that follow,
unless otherwise stated, the values reported have been standardised to account for token frequency per 100 words. While normalisation is necessary in order to provide a comprehensible view of the nine participants’ frequency of token use, doing so also negates various individual differences, such as total word count and other more qualitative attributes. I attempt to draw these out from the transcripts of the OTRE sessions and interviews.

5.3.1. Forum I
Forum I, which was concerned with patients’ rights and doctors’ responsibilities (see Section 3.7), was assigned and later closed before the first OTRE session. The data from this forum therefore provides a baseline for each participants’ token use. In Table 5.6, the average number of tokens used per 100 words and the total word count in each forum is displayed, and it can be seen that the figures for the token count are fairly consistent.

Table 5.6: Average token count of nine participants

<table>
<thead>
<tr>
<th>Output source</th>
<th>MDM (token) Count per 100 words/Total Word Count for Forums I, II, V (Averaged for P1-P9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F I</td>
</tr>
<tr>
<td></td>
<td>11.1/160.0</td>
</tr>
</tbody>
</table>

5.3.2. Frame 1: Forums I, II and V
In the following table (Table 5.7) the data for Forums I, II, and V are presented together. Statistical figures such as mean average, standard deviation and range are recorded. In the three tables (Tables 5.8-5.10) below that, the data concerning token counts, frequency and total word count are presented separately for each of the three forums. Note that the symbol, ∅, denotes cases where participants did not participate, while a 0 indicates that the participant did participate, but did not use tokens.
Table 5.7: Token use in Forums I, II and V, by P1-P9

As can be seen by these values, there was a general consistent and sustained effort on the part of all participants. In the following tables (Tables 5.7-5.9), the token counts can be viewed in light of the total word count in each forum.

Values such as those seen in P1 and P2 show that there could be different starting points, in terms of token frequency and a variety of responses to the tasks both before and after the first and subsequent OTRE sessions, but that token use became and remained an important part of each participants’ written output.

In P3 the word count between Forum I and II decreased and then rose in Forum V. At the same time token count increased between Forum I and II and Forum II and V.

When we take in all the elements of P2, P3, P4, P6, and P9 we observe trends consistent with the overall average. There is general increase in written output which makes
the general increase in the percentage of words dedicated to tokens noteworthy. It addresses RQ1, suggesting that WebSequiturs played a positive role in influencing participants to increase and sustain their use of tokens over longer passages.

5.3.3. Frame 2: Forums III and IV
In this section, I repeat the analysis performed in the preceding section. In this second frame, there was an emphasis on the formation of hypothetical premises that could predict the outcome of the Cups and Coins question. There was less discourse between participants (and other students) and more direct discussion by participants of the issues and the prediction of the outcome. In Table 5.11, the average number of tokens used per 100 words and the total word count in each forum is displayed, and these two values are consistent with each other.

<table>
<thead>
<tr>
<th>Output source</th>
<th>F III</th>
<th>F IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.2</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td>47.0</td>
<td>54.8</td>
</tr>
</tbody>
</table>

Table 5.11: Average token use for nine participants, Frame 2

In the following table (Table 5.12) the data for Forums III and IV are presented together. Statistical figures such as mean average, standard deviation and range are recorded. Following that are two tables (Table 5.13 and Table 5.14) containing data concerning token counts, frequency and total word count that are presented, separately for both forums.

<table>
<thead>
<tr>
<th>MDM Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forwards III, IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token / Token Word Counts: Forums III/IV</th>
<th>Total MDMs</th>
<th>Freq./ 100 words</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>Total Word Count</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4/4</td>
<td>8</td>
<td>9.1</td>
<td>4.0</td>
<td>0</td>
<td>8</td>
<td>88</td>
<td>44.0</td>
<td>15.6</td>
<td>22</td>
</tr>
<tr>
<td>P2</td>
<td>4/4</td>
<td>6</td>
<td>6.6</td>
<td>3.0</td>
<td>1.4</td>
<td>2</td>
<td>91</td>
<td>45.5</td>
<td>13.4</td>
<td>19</td>
</tr>
<tr>
<td>P3</td>
<td>3/5</td>
<td>6</td>
<td>7.3</td>
<td>3.0</td>
<td>0</td>
<td>0</td>
<td>82</td>
<td>41.0</td>
<td>8.5</td>
<td>12</td>
</tr>
<tr>
<td>P4</td>
<td>5/5</td>
<td>13</td>
<td>10.1</td>
<td>6.5</td>
<td>2.1</td>
<td>3</td>
<td>129</td>
<td>64.5</td>
<td>43.1</td>
<td>61</td>
</tr>
<tr>
<td>P5</td>
<td>7/11</td>
<td>13</td>
<td>11.4</td>
<td>6.5</td>
<td>0.7</td>
<td>1</td>
<td>114</td>
<td>57.0</td>
<td>2.8</td>
<td>4</td>
</tr>
<tr>
<td>P6</td>
<td>4/4</td>
<td>18</td>
<td>7.6</td>
<td>9.0</td>
<td>7.1</td>
<td>10</td>
<td>237</td>
<td>118.5</td>
<td>57.3</td>
<td>81</td>
</tr>
<tr>
<td>P7</td>
<td>3/4</td>
<td>14/33</td>
<td>8</td>
<td>4.0</td>
<td>1.4</td>
<td>2</td>
<td>86</td>
<td>43.0</td>
<td>17.0</td>
<td>24</td>
</tr>
<tr>
<td>P8</td>
<td>0/0</td>
<td>0/0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P9</td>
<td>10/10</td>
<td>19</td>
<td>13.3</td>
<td>9.5</td>
<td>0.7</td>
<td>1</td>
<td>143</td>
<td>71.5</td>
<td>3.5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5.12: Token use in Forums III and IV, by P1-P9, Frame 2
Viewing Table 5.12 for the average token count of the nine participants, it can be seen that, compared to the previous set of values showing token use for Forums I, II and V, the word counts in Forums III and IV were much lower. For this reason, values such as frequency of token use per 100 words were much more reactive to comparatively smaller changes in frequency of token use.

Participants’ output in Forum III and Forum IV (in the second frame) was more uniform than in Forums I, II and V (the first frame). This was no doubt owing to the perceived simplicity of the issue presented and the fewer words required to complete the task. Also, there may have been a perception that these two forums were less about discussion and more about hypothesis formation.

InForum III, most participants showed a decrease in token usage and in participation altogether. The subjects of Forum III and Forum IV involved predicting results of the class inquiry into probability, and hence, afforded opportunity for use of selected phrases and structures. However, although Forum III was done before the OTRE work in conditional statements and hypothetical premises, and Forum IV was done after the OTRE work, the number of tokens used by P1 and P4, and to a slighter degree P9, decreased in Forum IV. P2, P3, P7 and P8 increased their frequency of token use. P5 and P6 maintained their frequency with no noteworthy change.

In summary, seven of the nine participants either maintained or increased their token count between Forums III and IV, and five participants increased their token word count (i.e., the total number of words used in tokens), while two maintained the same word count. The

---

**Table 5.13: F III, P1-P9 token counts, frequency, total word count**

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts:</th>
<th>Freq. / 100 words</th>
<th>Total Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4</td>
<td>12.1</td>
<td>33.0</td>
</tr>
<tr>
<td>P2</td>
<td>4</td>
<td>7.3</td>
<td>55.0</td>
</tr>
<tr>
<td>P3</td>
<td>3</td>
<td>6.4</td>
<td>47.0</td>
</tr>
<tr>
<td>P4</td>
<td>5</td>
<td>14.7</td>
<td>34.0</td>
</tr>
<tr>
<td>P5</td>
<td>7</td>
<td>11.9</td>
<td>59.0</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>5.1</td>
<td>78.0</td>
</tr>
<tr>
<td>P7</td>
<td>3</td>
<td>9.7</td>
<td>31.0</td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>P9</td>
<td>10</td>
<td>14.5</td>
<td>69.0</td>
</tr>
</tbody>
</table>

**Table 5.14: F IV, P1-P9 token counts, frequency, total word count**

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts:</th>
<th>Freq. / 100 words</th>
<th>Total Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>4</td>
<td>7.3</td>
<td>55.0</td>
</tr>
<tr>
<td>P2</td>
<td>5</td>
<td>13.9</td>
<td>36.0</td>
</tr>
<tr>
<td>P3</td>
<td>3</td>
<td>8.6</td>
<td>35.0</td>
</tr>
<tr>
<td>P4</td>
<td>8</td>
<td>8.4</td>
<td>95.0</td>
</tr>
<tr>
<td>P5</td>
<td>6</td>
<td>10.9</td>
<td>55.0</td>
</tr>
<tr>
<td>P6</td>
<td>14</td>
<td>8.8</td>
<td>159.0</td>
</tr>
<tr>
<td>P7</td>
<td>5</td>
<td>9.1</td>
<td>55.0</td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td>Ø</td>
<td>Ø</td>
</tr>
<tr>
<td>P9</td>
<td>9</td>
<td>12.2</td>
<td>74.0</td>
</tr>
</tbody>
</table>
exceptions were P5 whose values decreased and P8 who did not participate in either forum. These values for token use in Forums III and IV further address RQ1. They are supportive of the proposition that OTREs can influence language learners to use tokens with greater frequency.

5.3.4. Chat rooms I and II
In this section, we look again at token use in Chat Rooms I and II. The synchronous and in-class nature of these writing tasks is more relevant in this section on token use than in the previous section on total word count. While the participants in the forums could refer to the Master MDM List provided on their LETUS page, during the chat room discussions, participants were allowed no materials except an image of the discussion topics displayed on the middle of three monitors provided at each desk for every two students. Thus, the output in the chat rooms provide an indication of how well students retained, within their working memory the tokens that they had previously seen used during the OTRE sessions.

In Table 5.15, the average number of tokens used per 100 words and the total word count in each forum is displayed, and this value increases considerably in CR II.

<table>
<thead>
<tr>
<th>Output source</th>
<th>CR I</th>
<th>CR II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.7/149.7</td>
<td>12.3/290.2</td>
</tr>
</tbody>
</table>

Table 5.15: Average token use for nine participants, Frame 2

In the following table (see Table 5.16), the data for Chat Room I and Chat Room II are presented together. Statistical figures such as mean average, standard deviation and range are recorded. In the two tables which follow (Table 5.17 and Table 5.18), data concerning token counts, frequency and total word count are presented, separately for both chat rooms.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts: Chat Rooms I/II</th>
<th>Total MDMS</th>
<th>Freq./100 words</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>Total Word Count</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>5 13</td>
<td>17.0</td>
<td>5.9</td>
<td>8.5</td>
<td>6.4</td>
<td>9.0</td>
<td>290.0</td>
<td>145.0</td>
<td>96.2</td>
<td>136.0</td>
</tr>
<tr>
<td>P2</td>
<td>10 25</td>
<td>34.0</td>
<td>7.3</td>
<td>17.0</td>
<td>9.9</td>
<td>14.0</td>
<td>466.0</td>
<td>233.0</td>
<td>86.3</td>
<td>122.0</td>
</tr>
<tr>
<td>P3</td>
<td>14 23</td>
<td>37.0</td>
<td>28.7</td>
<td>18.5</td>
<td>6.4</td>
<td>9.0</td>
<td>129.0</td>
<td>64.5</td>
<td>29.0</td>
<td>41.0</td>
</tr>
<tr>
<td>P4</td>
<td>11 20</td>
<td>31.0</td>
<td>7.7</td>
<td>15.5</td>
<td>6.4</td>
<td>9.0</td>
<td>405.0</td>
<td>202.5</td>
<td>24.7</td>
<td>35.0</td>
</tr>
<tr>
<td>P5</td>
<td>11 14</td>
<td>25.0</td>
<td>6.0</td>
<td>12.5</td>
<td>2.1</td>
<td>3.0</td>
<td>414.0</td>
<td>207.0</td>
<td>149.9</td>
<td>212.0</td>
</tr>
<tr>
<td>P6</td>
<td>17 30</td>
<td>47.0</td>
<td>4.0</td>
<td>23.5</td>
<td>9.2</td>
<td>13.0</td>
<td>1188.0</td>
<td>594.0</td>
<td>308.3</td>
<td>436.0</td>
</tr>
<tr>
<td>P7</td>
<td>4 13</td>
<td>17.0</td>
<td>5.9</td>
<td>8.5</td>
<td>6.4</td>
<td>9.0</td>
<td>287.0</td>
<td>143.5</td>
<td>94.0</td>
<td>133.0</td>
</tr>
<tr>
<td>P8</td>
<td>8 15</td>
<td>23.0</td>
<td>11.4</td>
<td>11.5</td>
<td>4.9</td>
<td>7.0</td>
<td>202.0</td>
<td>101.0</td>
<td>18.4</td>
<td>26.0</td>
</tr>
<tr>
<td>P9</td>
<td>17 17</td>
<td>34.0</td>
<td>8.6</td>
<td>17.0</td>
<td>0.0</td>
<td>0.0</td>
<td>397.0</td>
<td>198.5</td>
<td>34.6</td>
<td>49.0</td>
</tr>
</tbody>
</table>

Table 5.16: Token use in Chat Rooms I and II, by P1-P9, Frame 3
Below, the data for token counts are presented for participants, when the results serve to illustrate a general trend or points of interest.

The average of all nine participants’ output is shown in Table 5.16. The values displayed for the output of P1, P2, P4, P6, P7, P8 and P9 are consistent with the overall average either maintaining or increasing token use frequency. P3’s and P5’s token use does not resemble the overall average, as they show decreases, although quantities in both are greater than any of the other seven.

The Master MDM List was prohibited from use during the 45-minute to one-hour long chat rooms, so that the retention of the OTREs could be inferred from participants’ written output. Token use was less frequent during the forums than in the chat room sessions.

The token use in Chat Room I (see Table 5.17) was much lower than the output in Chat Room II (see Table 5.18) where token use increased greatly. Perhaps, in the first chat, some time was spent learning to use the chat interface. In the overall average of the nine participants’ token use, the count increased a small amount and the token word count increased due to the use of longer phrased tokens. This was reflected in the output of P1, P2, P4, P5, P7 and P8 (see Appendices 20-21). Although a large increase in token word count followed the increase in the general total word count, the token count (the number of discrete tokens) did not greatly increase. This indicated an increased use of longer phrased tokens and therefore a broadening variety of tokens provided on the Master MDM List and presented in the OTREs, which was investigated further by examining participants’ writing (see Section

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts:</th>
<th>Freq. /100 words</th>
<th>Total Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>5</td>
<td>6.5</td>
<td>77.0</td>
</tr>
<tr>
<td>P2</td>
<td>10</td>
<td>5.8</td>
<td>172.0</td>
</tr>
<tr>
<td>P3</td>
<td>14</td>
<td>16.5</td>
<td>85.0</td>
</tr>
<tr>
<td>P4</td>
<td>11</td>
<td>5.9</td>
<td>185.0</td>
</tr>
<tr>
<td>P5</td>
<td>11</td>
<td>10.9</td>
<td>101.0</td>
</tr>
<tr>
<td>P6</td>
<td>17</td>
<td>4.5</td>
<td>376.0</td>
</tr>
<tr>
<td>P7</td>
<td>4</td>
<td>5.2</td>
<td>77.0</td>
</tr>
<tr>
<td>P8</td>
<td>8</td>
<td>9.1</td>
<td>88.0</td>
</tr>
<tr>
<td>P9</td>
<td>8</td>
<td>4.6</td>
<td>174.0</td>
</tr>
</tbody>
</table>

Table 5.17: CR I, P1-P9 token counts, frequency, total word count

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts:</th>
<th>Freq. /100 words</th>
<th>Total Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>13</td>
<td>6.1</td>
<td>213.0</td>
</tr>
<tr>
<td>P2</td>
<td>25</td>
<td>8.5</td>
<td>294.0</td>
</tr>
<tr>
<td>P3</td>
<td>23</td>
<td>52.3</td>
<td>44.0</td>
</tr>
<tr>
<td>P4</td>
<td>20</td>
<td>9.1</td>
<td>220.0</td>
</tr>
<tr>
<td>P5</td>
<td>14</td>
<td>4.5</td>
<td>313.0</td>
</tr>
<tr>
<td>P6</td>
<td>30</td>
<td>3.7</td>
<td>812.0</td>
</tr>
<tr>
<td>P7</td>
<td>13</td>
<td>6.2</td>
<td>210.0</td>
</tr>
<tr>
<td>P8</td>
<td>15</td>
<td>13.2</td>
<td>114.0</td>
</tr>
<tr>
<td>P9</td>
<td>17</td>
<td>7.6</td>
<td>223.0</td>
</tr>
</tbody>
</table>

Table 5.18: CR II, P1-P9 token counts, frequency, total word count
5.4.

Summing up this section, the written output of seven of the nine participants, to varying degrees, was consistent with the overall average. P3’s output differed in that he wrote fewer words in Chat Room II than had he had in Chat Room I. P6’s total word count increased similarly to the overall average, but his token word count decreased by a large amount; conversely his token count increased only a small amount. This was the only occasion in which the percentage of total words used as tokens by a participant decreased by a large amount. These data indicate that tokens were retained in memory and that participants were able to access and retrieve them in time to employ them in synchronous discussions and further aids in addressing RQ1.

5.3.5. Written Report and Individual In-Class Rewrite

In this section, I repeat the analysis performed in the preceding section, but in this fourth frame with the asynchronous collaborative written output of the GWR and the in-class, timed individual In-Class Rewrite. Participants’ written output is examined first in terms of their group writing and then in terms of their individual performance as they attempted to recall and employ the tokens and general language used in the group effort. Tasking students to do this identified any non-contributors to the group report so that I could avoid incorrectly attributing token use to participants who played little or no role in the writing. This task also supported me in making inferences later about how well tokens were retained following the complete series of OTRE sessions.

In Table 5.19, the average number of tokens used per 100 words and the total word count in each forum is displayed, and this figure increases considerably in CR II.

<table>
<thead>
<tr>
<th>Output source</th>
<th>MDM (token) Count per 100 words/Total Word Count for GWR and ICR (Averaged for P1-P9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWP I</td>
<td>5.6/417.3</td>
</tr>
<tr>
<td>ICR II</td>
<td>10.0/215.8</td>
</tr>
</tbody>
</table>

Table 5:19: Average token use for nine participants, Frame 2

In the following table (Tables 5.20) the data for the Group Written Report (GWR) and the individually written In-Class Rewrite are presented together. Statistical figures such as mean average, standard deviation and range are recorded.
Table 5.20: Token use in the GWP and ICR, by P1-P9, Frame 4

In the two tables which follow (Tables 5.17 and 5.18), data concerning token counts, frequency and total word count are presented, separately for both output sources. Below, where results serve to highlight trends seen across groupings of participants the data for token counts are presented via bar graphs.

Table 5.21: GWR, P1-P9 token counts, frequency, total word count

Table 5.22: ICR, P1-P9 token counts, frequency, total word count

While no minimum word count was imposed, participants were tasked with writing a paper which included an introduction to the problem about probability being studied, a section presenting the hypothetical premise, a description of the methodology, a presentation of the data, a discussion of the data, and a conclusion. The frequency of words used as tokens decreased in the GWR partially owing to the fact that a greater proportion of words were used in the sections of the report that required less argument and more description and explanation.
of the process by which the students carried out the task. The participants’ writing often focused as much on expressing surprise at the results of the Monty Hall experiment as on constructing an argument.

Output for the GWR ranged from 416 to 535 words. As for each triad’s performance on the GWR, they achieved token counts and token word counts within the following ranges: Group 1: 21 tokens using 31 of the 535 total words; Group 2: 22 tokens using 39 of the 416 total words; and Group 3: 32 tokens using 56 of 426 words.

In the ICR (see Table 5.22), excepting P6, who increased the number of tokens used from that used in his group’s GWR, from 22 to 32, all other participants’ token use decreased as did their token word count. Due to the concurrent decrease in total word count, and arguably, the participants’ purposeful token use, the frequency of token use per 100 words increased in all nine cases. The three largest increases in frequency of token use occurred with the three participants in Group 3 where it roughly doubled their frequencies to those found in the GWR. These results strongly suggest that following the completion of the OTREs and writing tasks, all participants made appreciable gains in their use of tokens.

The value for the average frequency of token use in the ICR is best understood by considering the three comparatively high frequencies of token use from Group 3. P1 and P3 and to a somewhat lesser extent, P4, were judged to have expressed noteworthy increases. Other participants did not show similar increases. However, no participants decreased their frequency of use from the frequencies they achieved in either Forum V or Chat Room II. However, the number of tokens used did decrease in all nine participants. Interview responses suggest that the absence of the Master MDM list which was present in Forum V and the lack of synchronous social interaction which was present in Chat Room II is likely related to those decreases.

5.3.6. Quantitative Summary

The data collected in this section that overall, the OTRE sessions were followed by an increase in token use. Moreover, given the interpretive nature of this study, inconsistent results were viewed as the likely influence of many other contextual factors as is further suggested by participants’ responses during the semi-structured interviews. While it is instructive to consider the average total word count, the token count, and the token word count over the course of the nine assignments, a high token count may only record the repeated use of two or three identical markers. Cases of widening gaps between token count and token word count per 100 words in Forum V and the maintaining of that gap even as the
token count increased in Chat Room II; and the GWR suggests that participants not only began using more tokens, but they also began using a broader variety of token phrases. This broader use is apparent, because the first set of WebSequiturs used a greater number of one-word tokens. Later, more tokens, some of them longer, were introduced and the token count / token word count ratio indicated a broadened selection of tokens. For example, over the course of the term, as in Ma and Wang (2016), participants who, initially used only I think began to also use in my opinion, and it is my belief that. It should be made clear here that longer tokens are not considered to be better, but their use, as noted in the higher token word count, does demonstrate that a broader variety of tokens were being used. This adds further support to the argument that OTREs are an effective way to influence language learners to make more frequent use of tokens, thereby contributing to answering RQ1. On the other hand, this does not indicate unique uses of tokens. It does not preclude cases in which participants use two or three longer phrases repetitively. In the next section, I will discuss data related to the unique token use.

5.4. Unique MDM (token) use

In this section, I discuss increases in unique token use. By unique, I mean single uses of a given token after which a different token expressing the same meaning is used.

I arrived at the number of unique token uses by using the MS Word document containing the collection of forum and chat room posts and the GWP and ICR texts and using the search function set for open parentheses which preceded each Roman numeral (see Section 4.2.3) which in turn denoted each token. Using printed copies of the Master MDM List (see Appendix 13), I checked each forum post for each participant separately, writing a checkmark next to each multiple use of any given token. I counted the number of unchecked tokens and used that value to represent the number of unique uses.

The increase in unique token use adds weight to the argument responding to RQ1. It demonstrates that participants not only increased their frequency of token use, but that they also increased the variety of tokens used. Figure 5.1 shows that for the output sources in Frame 1 (FI, FII and FV), Frame 3 (CI and CII) and Frame 4 (GWR and ICR), the number of unique token uses increased. In the case of Frame 2 (FIII and FIV), which focused on hypothesis formation, the amount of unique token use showed a slight decrease in use.
Figure 5.1 Average of nine participants’ unique MDM (token) uses

In brief, the figures shown in Figure 5.1 indicate that participants internalised (Lantolf and Thorne, 2007) the tokens they had practiced using during the OTRE sessions. Over the course of the term, participants’ range of tokens increased and began to draw upon the expanding range of relevant communicative functions, which in Section 4.1.7, I defined, for the purposes of this study, as types of communication that are utilised when explicitly or implicitly connective MDMs are used to signal premises or conclusions in the context of a written argument. Participants shifted from repetitive uses of words within a given communicative function, e.g., so, and began also using indicators such as therefore, and for that reason.

Participants’ use of unique tokens roughly followed the same trajectory as that of the general token count. Figure 5.2 shows that this trajectory is generally born out in each participant’s writing. It is important to note that Figure 5.2 is non-normalised, as it compares the total number of unique tokens with the total number of token uses in each output source. In each frame, except for the second frame (Forums III and IV), which remained nearly the same, participants’ use of unique tokens increased as token use increased. The gap between the token use and unique token use remained consistent, and in some cases converged, indicating that a participant was using nearly all unique tokens. Just as the increase in the token count over the course of the term supports the efficacy of WebSequiturs in influencing the participants to make greater use of tokens, the increase in the use of unique tokens further
supports that conclusion. This speaks to the question raised in the previous section regarding whether, in addition to using longer phrases, participants also used a greater variety of tokens, or used only a limited selection of longer tokens repeatedly.

Figure 5.2: Average of nine participants token use and unique token use

5.4.1. MDM (token) use across communicative functions

In the previous section I presented data showing that participants had not only used a greater number of argument-related tokens, but also a greater variety of tokens. In this section, I show that participants also increased their token use across a progressively greater range of the communicative functions related to explicit and implicit premise and conclusion indicators. Figures 5.3 - 5.6 show that, on average, there was a greater number of unique tokens used and that they were drawn from a broader range of communicative functions. The average upward trend began at Chat Room I and generally increased (see Figure 5.5). As with the data points discussed earlier, the increase continued until the Group Written Report and the subsequent In-Class Rewrite when the levels fell (see Figure 5.6). As mentioned, this is likely because of inaccessibility to the Master MDM List and a lack of social queuing from interlocutors. During the interviews, participants commented on the importance of the social dimension of written discussion and collaborative writing (see Section 6.7).
Figure 5.3: Average for Unique tokens and Communicative Functions in Forums I, II and V

Figure 5.4: Average for Unique tokens and Communicative Functions in Forums III, IV
5.5. Explicitly connective premises and conclusion indicators

In the previous sections of this chapter, I have shown evidence of broadening token use within and across communicative functions, although I have treated them in a general way. As noted earlier, in this study I have constrained the scope of MDMs to: (1) premise and conclusion indicators and (2) conditional statements and hypothetical premises. In this section I will discuss changes in participants’ use of two classes of premise and conclusion indicators, that is, implicitly connective and explicitly connective premise and conclusion indicators.
In Section 4.1.6, I explained that implicitly connective premise and conclusion indicators could appear within communicative functions including Levels of Conviction, Agreement and Disagreement, Reservations, Creating Focus, and Examples. Critically Minded Podcast, discussed in Section 3.6.4 focuses on the explicitly connective type and, as I wrote earlier, at the outset of this study I was only interested in tracking participants’ use of tokens categorised as explicitly connective premise and conclusion indicators. I had intended to collect quantitative data, i.e., changes in the number and frequency of premise indicators such as because, since and the reasons are, etc. and conclusion indicators such as so, consequently, resulting in, and for that reason. However, an emergent theme arose during my research into the background literature, as Halliday (2009) showed that many MDMs can carry out multiple functions simultaneously. It also emerged during my preliminary qualitative analysis, as I noted, that participants often used implicitly connective indicators where an explicitly connective indicator would have provided more cohesion to their arguments.

In the following sections, I will present participants’ use of implicitly connective and explicitly connective premise and conclusion indicators in the forums, chat rooms, the GWR and ICR.

In order to present a visual representation of the data, but minimize the number of graphs, in some cases, I present only one participant’s graph among a grouping of similarly performing participants. I reference those participants as, for example, as a frequency of use observed in P3, P4, P6, P7, P8 are presented as similar to P9’s output, which is then presented in Figure 5.23. In this way, I minimise the number of graphs and present general trends in token use.

5.5.1. Implicitly connective and explicitly connective MDMs (tokens) in Forums I, II and V

I now turn to the data pertaining specifically to premise and conclusion indicators, which I divide into two types: implicitly and explicitly connective. I intend to show how premise indicators, implicit (Imp) and explicit (Exp) increased after the completion of the preceding OTREs and continued to do so until the introduction of tasks involving conditional statements, which is covered in later sections.

In this section, I repeat the analysis performed in the preceding section. In the following table (Table 5.23) the data for Forums I, II and V are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the
figures for the normalised number of uses of implicit (Imp) and explicit (Exp) premise and conclusion indicators.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts: Forums</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>MDM Freq. / 100 words (normalised) F I</th>
<th>MDM Freq. / 100 words (normalised) F II</th>
<th>MDM Freq. / 100 words (normalised) F V</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>14 8 9</td>
<td>10.3</td>
<td>1.0 1.7</td>
<td>6.0 3.0</td>
<td>8.1 0.0</td>
<td>5.19 0.0</td>
<td>7.38 2.5</td>
</tr>
<tr>
<td>P2</td>
<td>3 2 10</td>
<td>3.0 2.0 2.6</td>
<td>4.0 5.0</td>
<td>5.9 0.0</td>
<td>2.17 2.2</td>
<td>4.50 4.5</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>12 5 14</td>
<td>8.7 1.7 5.8</td>
<td>1.5 10.0 3.0</td>
<td>7.9 0.0</td>
<td>2.74 4.1</td>
<td>9.23 1.5</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>12 9 15</td>
<td>9.3 2.7 2.1</td>
<td>0.6 4.0</td>
<td>2.0 5.4 1.08</td>
<td>7.45 0.0</td>
<td>7.97 2.9</td>
<td></td>
</tr>
<tr>
<td>P5</td>
<td>17 0 0</td>
<td>3.0 0.3 5.2</td>
<td>0.6 9.0 1.0</td>
<td>7.5 0.83</td>
<td>0 0</td>
<td>0 0</td>
<td>0</td>
</tr>
<tr>
<td>P6</td>
<td>30 16 21</td>
<td>18.3 4.0 5.8</td>
<td>2.6 10.0 5.0</td>
<td>6.2 1.23</td>
<td>5.91 0.4</td>
<td>4.57 1.8</td>
<td></td>
</tr>
<tr>
<td>P7</td>
<td>12 12 14</td>
<td>11.0 1.7 1.0</td>
<td>1.5 2.0</td>
<td>3.0 7.5</td>
<td>0.0 6.58</td>
<td>1.3</td>
<td>9.91 2.7</td>
</tr>
<tr>
<td>P8</td>
<td>10 0 9</td>
<td>5.3 1.0 4.7</td>
<td>1.0 9.0</td>
<td>2.0 12.9</td>
<td>1.43</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P9</td>
<td>14 7 13</td>
<td>8.3 3.0 3.1</td>
<td>1.7 6.0</td>
<td>3.0 7.2</td>
<td>4.0 5.43</td>
<td>2.2</td>
<td>7.38 1.3</td>
</tr>
</tbody>
</table>

Table 5.23: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 1

The results shown in the average of the nine participants’ use of explicitly and implicitly connective indicators (see Figure 5.7) reiterate the general increased use of argument-related tokens and an increase in explicitly connective premise and conclusion indicators.

A theme emerged in which the increase in implicitly connective indicators is mirrored by a decrease in explicitly connective indicators. Likewise, a decrease or plateauing of implicitly connective indicators was matched by a corresponding increase in explicitly connective indicators. It appears as though the two classes of tokens are behaving as competitors within a context of finite resources (i.e., the text passage). For this reason, the general frequency of token use is not always additive but can be more like a tug-of-war between textual and interpersonal tokens. Leedham and Cai (2013) noted a similar shift as Chinese students’ use of certain linking adverbials decreased as other tokens were used to create cohesion.

This shift from implicitly to explicitly connective indicators can be seen in P2, P3 and P7 (see Figures 5.8-5.10) from Forum I to Forum II. The converse of this relationship can be observed from Forum II and Forum V in P3 (see Figure 5.11) and P9 (see Figure 5.12) where an increase in implicitly connective indicators was matched by a decrease in explicitly connective indicators. Irrespective of competition between these two classes of tokens, explicitly connective indicators made increases by a small amount as with P1 and larger.
amounts as with P3, P4, P6, P7, P8 and P9 (see Table 5.23).

Figure 5.7: Average of P1-P9’s’ frequency of Implicitly and Explicitly Connective token use for Forums I, II and V

Figure 5.8: P2’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.9: P3’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V

Figure 5.10: P7’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
Figure 5.11: P6’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V

Figure 5.12: P9’s frequency of Implicitly and Explicitly Connective token use for Forums I, II and V
To summarise, an expanding range of communicative functions were used by participants. They also continued to employ explicitly connective indicators by degrees, suggesting that OTREs had played a role in influencing participants to use them with greater frequency, thereby further strengthening the response to RQ1 and adding further weight to the argument that OTREs can be an effective tool for influencing language learners to make more frequent use of a variety of tokens.

5.5.2. Implicitly and explicitly connective premises and conclusion indicators in Forums III, IV

In the next two sections, I discuss participants’ written output Forums III and IV in terms of tokens used by participants following OTREs modelling of the premises and conclusions and then in terms of tokens used in construction of conditional statements (including hypothetical premises).

In the following table (Table 5.24) the data for Forums III and IV are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the figures for the normalised number of uses of implicit (Imp) and explicit (Exp) premise and conclusion indicators.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts: Forums</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>MDM Freq. / 100 words (normalised)</th>
<th>MDM Freq. / 100 words (normalised)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FIII</td>
<td>FIV</td>
<td>Imp/Exp</td>
<td>Imp/Exp</td>
<td>F III</td>
<td>Imp/Exp</td>
</tr>
<tr>
<td>P1</td>
<td>3</td>
<td>2</td>
<td>2.0</td>
<td>0.5</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>P2</td>
<td>2</td>
<td>2</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>P3</td>
<td>3</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.0</td>
<td>1.0</td>
<td>2.8</td>
<td>0.0</td>
</tr>
<tr>
<td>P4</td>
<td>1</td>
<td>5</td>
<td>2.0</td>
<td>1.0</td>
<td>2.8</td>
<td>0.0</td>
</tr>
<tr>
<td>P5</td>
<td>7</td>
<td>6</td>
<td>1.5</td>
<td>1.5</td>
<td>2.1</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>7</td>
<td>3.5</td>
<td>2.0</td>
<td>0.7</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.4</td>
<td>1.0</td>
<td>2.5</td>
<td>1.9</td>
</tr>
<tr>
<td>P7</td>
<td>3</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>P8</td>
<td>0</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>P9</td>
<td>6</td>
<td>5</td>
<td>4.0</td>
<td>1.5</td>
<td>1.4</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 5.24: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 2
The participants’ averaged written output (see Figure 5.13) during Forum III and Forum IV illustrates other ways one type of token use can increase at the expense of another. In nearly all cases, participants’ use of explicitly connective premise and conclusion indicators in these three writing tasks fell well below the number of implicitly connective ones. The reason for the decrease can be explained by the focus on formation of hypothetical premises in those forums. A decrease in explicitly connective premise indicators can be seen in Figure 5.13 as students wrote of the context in which the Monty Hall Problem was explored in a more prosaic style, reserving their focus on explicitly logical expression for the hypothetical premise.

Figure 5.13 Average of P1-P9s’ frequency of Implicitly and Explicitly Connective token use for Forums III and IV

In this instance the average token use of the nine participants does not always accurately represent the individual cases. The participants’ use fell into three main pattern types. First, P1 (see Figure 5.14) and similarly P3 and P7 were cases where participants’ use of explicitly connective premise indicators was present in Forum III and then absent in the Forum IV.
Figure 5.14: P1’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV

P6 (see Figure 5.15), and similarly P9; show a more common pattern in which implicit premise indicators decreased in Forum IV from their use in Forum III while the use of explicit argument indicators increased.

Figure 5.15: P6’s frequency of token use in conditional statements and hypothetical premises in Forums III and IV
P2 used an equal number of both types of indicators in Forum III and in Forum IV increased the number of uses of both.

![Bar chart showing the frequency of implicitly and explicitly connective premises and conclusion indicators for Forums III and IV.]

Figure 5.16: P2’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV

Other participants’ use of argument indicators appear more chaotic. For example, P4 (see Figure 5.17) used no implicitly connective indicators in Forum III but then used them more than most other participants in Forum IV. At the same time, P4’s use of explicitly connective indicators dropped by two-thirds, showing that it was not always the case that OTRE sessions were immediately followed by greater frequency in the use of the modelled token uses.
Figure 5.17: P4’s frequency of Implicitly and Explicitly Connective token use for Forums III and IV.

P5’s (see Figure 5.18) activity is also worth noting, as explicitly connective premise and conclusion indicator use decreased by half while implicitly connective premise and conclusion indicator use ceased altogether. These trends are, arguably, explained in the next section as resulting from more focus being directed by the participants toward hypothesis construction.
5.5.3. Implicitly and explicitly connective hypothetical premises and conclusion indicators in Forums III and IV

In this section, I introduce the theme of participants’ use of hypothetical premises and conclusions. I am beginning the discussion of conditional statements in this second frame because, although they may have appeared incidentally in the first frame, it was in this second frame where their use was first tasked, a pre-OTRE baseline was established, and post-OTRE token use was measured. I intend to show here, and in subsequent sections, how the use of premise indicators gave way to an increased emphasis on use of conditional statements.

In the following table (Table 5.25) the data for Forums III and IV are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the figures for the normalised number of uses of implicit (Imp) and explicit (Exp) premise and conclusion indicators.
Table 5.2: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 2

As noted earlier, Forum III and Forum IV were different from Forums I, II and V and did not involve emotive social issues. The tasks involved an extended enquiry into probability, using the Monty Hall Problem (see Section 3.8). The topic necessitated participants engaging in a more analytical writing genre. Participants used fewer interpersonal tokens and focussed on using the structures which they had practiced in the OTREs (see Figure 5.59). While Forum III was completed before OTRE Sessions IV and V, both of which modelled the formation of hypothetical premises, participants did not have prior training in the formation of hypothetical premises during Forum III. Forum IV was completed after participants had worked through OTRE Sessions IV and V. During their time engaging in Forum IV, participants had the training to prepare them for forming hypothetical premises. As the task set in these two forums involved the formation of hypotheses, much of the total word count dedicated to token use was taken up by use of conditional statements and hypothetical premises. Participants tended to make a shift from the premise and conclusion indicators, which they had studied at the beginning of the term, to hypothetical premises, which they had practiced during the previous lesson during OTRE Sessions IV and V.
Figure 5.19: Average of P1-P9s’ frequency of token use in conditional statements and hypothetical premises in the Forums III and IV

The individual uses of token in this frame (Forums I and II) closely resembled the general trend suggested in Figure 5.19. In P2 (See Figure 5.20), and similarly, P6 and P7, there was an increase between the pre-OTRE and post-OTRE writing tasks from no use of conditional statements and hypothetical premises at all to use at a level consistent with that of others in the Forums III and IV.

Figure 5.20: P2’s frequency of token use in conditional statements and hypothetical premises
An interesting contrast was P3 (see Figure 5.21) whose general token use was similar to these cases, except for the lack of implicitly connective indicators and the quantity of explicitly connective indicators (twice that found in nearly all the others).

![Figure 5.21: P3’s frequency of token use in conditional statements and hypothetical premises in the Forums III and IV](image)

In the case of P6 and P7, comparing their use of tokens with that of the previous section (Figure 5.22) indicates a reallocation of resources from premise and conclusion indicators to conditional statements and hypothetical premise indicators in the post-OTRE Forum IV. This is suggested by decrease in premise and conclusion use between Forum III and IV in the upper table of Figure 5.22; and the increase in the use of conditional statements in between Forum III and IV in the lower table of Figure 5.22. This suggests that subsequent to the OTREs modelling conditional statements and hypothetical premises, these participants were influenced to shift from premise and conclusion indicators to conditional statement and hypothetical premise indicators.
Table 5.2: P6’s and P7’s token use for premises and conclusion indicators

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts: Forums</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>MDM Freq. / 100 words (normalised) F III</th>
<th>MDM Freq. / 100 words (normalised) F IV</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>P6</td>
<td>4 7</td>
<td>3.5</td>
<td>2.0</td>
<td>0.7 1.4</td>
<td>2.5 1.9</td>
<td>1.0 2.0</td>
</tr>
<tr>
<td>P7</td>
<td>3 1</td>
<td>1.5</td>
<td>0.5</td>
<td>0.7 0.7</td>
<td>1.8 0.0</td>
<td>1.0 1.0</td>
</tr>
</tbody>
</table>

Table 5.26: P6’s and P7’s token use for premises and conclusion indicators

The kind of shift shown in Figure 5.22 was not always the case. In P4 (see Figure 5.22) and P9, there was large (P4) and moderate (P9) use of explicitly connective conditional statement and hypothetical premise indicators in Forum III, which then decreased in the post-OTRE Forum IV to levels similar to those seen in P2, P6, and P7. Perhaps these latter frequencies represent values closer to the norm of what is required to express the conditions and state the hypothesis in the Monty Hall situation.
In summary, although there were some individual differences in uses of tokens, a general trend was observed in which indicators associated with conditional statements and hypothetical premises increased in the post-OTRE Forum IV. It was also the case that, in Forum IV, although explicitly connective indicators did not always outnumber implicitly connective indicators, in five out of the eight cases in which participants were active (excepting P8 who did not contribute) their use increased. This indicates that following the OTREs which focussed on the constructions of conditional statements and hypothetical premises, participants had learned these constructions. It also suggests that they recognised the importance of stating those thoughts explicitly.

5.5.4. Implicitly and explicitly connective premise and conclusion indicators in Chat Rooms I and II

In this section, I present findings from the data gathered from the participants’ output in the two chat rooms. As with the previous sections, I will first examine the data involving the explicitly and implicitly connected premise and conclusion indicators. Then I will examine the implicitly and explicitly connective conditional statements and hypothetical premises.

In the following table (Table 5.28) the data for Chat Rooms I and II are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the figures for the normalised number of uses of implicit (Imp) and explicit (Exp)
premise and conclusion indicators.

| Implicitly / Explicitly Connective Premise and Conclusion Indicators |
|---|---|---|---|---|---|
| Chat Rooms I, II |
| Participants | Token Counts: Chat Rooms | Mean | Standard Deviation | MDM Freq. / 100 words (normalised) CI | MDM Freq. / 100 words (normalised) CII | Range |
| CI | CI | Imp/Exp | Imp/Exp | Imp/Exp | Imp/Exp | Implicit/Explicit |
| P1 | 4 | 11 | 11.0 | 0.0 | 4.2 | 0.0 | 5.2 | 0.0 | 6.0 | 0.0 |
| P2 | 8 | 23 | 2.0 | 0.5 | 1.4 | 0.7 | 4.7 | 0.0 | 7.1 | 0.7 | 2.0 | 1.0 |
| P3 | 11 | 23 | 7.0 | 1.5 | 7.1 | 2.1 | 12.9 | 0.0 | 47.7 | 4.5 | 10.0 | 3.0 |
| P4 | 8 | 15 | 8.5 | 2.0 | 2.1 | 0.0 | 3.8 | 0.5 | 6.4 | 0.5 | 3.0 | 0.0 |
| P5 | 11 | 14 | 4.5 | 0.5 | 6.4 | 0.7 | 7.9 | 3.0 | 2.6 | 1.0 | 9.0 | 1.0 |
| P6 | 13 | 26 | 20.0 | 3.0 | 7.1 | 2.8 | 2.9 | 0.5 | 2.6 | 0.6 | 10.0 | 4.0 |
| P7 | 4 | 12 | 11.0 | 1.0 | 1.4 | 1.4 | 5.2 | 0.0 | 5.7 | 0.0 | 2.0 | 2.0 |
| P8 | 8 | 14 | 4.5 | 0.5 | 6.4 | 0.7 | 8.0 | 1.1 | 10.5 | 1.8 | 9.0 | 1.0 |
| P9 | 14 | 13 | 7.0 | 3.5 | 2.8 | 2.1 | 6.3 | 1.7 | 5.4 | 0.4 | 4.0 | 3.0 |

Table 5.28: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 3

The average of the nine participants’ use of implicit premise and conclusion indicators (see Figure 5.24) shows a general increase. In Chat Room I, the students learned how to use the LETUS chat application as well as develop a sense of online identity (Heikoop, 2013) in this new environment. Also, the second chat, per students’ request, ran an extra 15 minutes longer than the first. These factors may explain why participants’ use of premise and conclusion indicators increased. However, while the frequency of use of implicitly connective indicators increased, use of explicitly connective indicators remained about the same in each task. Because implicitly connective indicators often fulfil both the function of argument indication as well as carrying emotive content, such a distribution might be expected in a chat where more interpersonal communication would be socially appropriate. Expressing levels of conviction, creating focus, and clarification for example, often also signal premises and conclusions. Participants may have felt that implicitly connective tokens felt more appropriate in a synchronous discussion context.
Participants’ performances are generally represented by the average shown in Figure 5.23. Two exceptions to this were P1 who maintained the same frequency of implicitly connective indicators (consistent with other participants in C I, but did not increase in C II, and used no explicitly connective indicators in both C I and C II; and P5 whose frequency of both decreased.

In summary, on average, participants’ use of tokens is consistent with the timeline of their exposure to and engagement with the OTREs and the hypothesis that OTREs can influence language learners to use tokens with greater frequency.

5.5.5. Implicitly and explicitly connective conditional statements and hypothetical premise indicators in Chat Rooms I and II

In this section, I continue presenting the analysis performed in the chat rooms. In the following table (Table 5.29) the data for the use of conditional statements and hypothetical premises in Chat Rooms I and II are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the figures for the normalised number of uses of implicit (Imp) and explicit (Exp) premise and conclusion indicators.
Table 5.29: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 3

The graphic representation of the average frequencies of conditional statement and hypothetical premise indicator use (see Figure 24) suggests little or no increase. An examination of the individual cases, however, gives a different view; seven of the nine participants increased their use with decreases only in P3 and P6. The frequency of token use of P1, and similarly P5, suggests that OTREs played a role in influencing students to increase their use of both implicitly and explicitly connective conditional statements and hypothetical premises. The frequency of token use of P2, and P5, P7 and P8, likewise, shows an increase in their use from none to 0.3 for explicitly connective and 0.5, 0.5 and 0.9 for implicitly connective indicators, respectively. The frequency of token use of P4 (see Figure 5.25) displays a case in which the participant was already using both implicitly and explicitly connective indicators and increased that use.

In both chat room discussions, participants sometimes posed hypothetical situations around the topics being discussed. For example, in Chat Room I, they discussed fat shaming and in Chat Room II students discussed blood type personality profiling. They were also offered the opportunity to add comments about the enquiry into probability, but most did not have time or opted not to do so, instead having an extended general discussion of the textbook topic. Therefore, the number of uses of both explicitly and implicitly connective conditional statements and hypothetical premises in both chat rooms were few and from Chat Room I to Chat Room II increased somewhat, but not greatly.
5.5.6. Implicitly and explicitly connective premises and conclusion indicators in the GWR and ICR

Like Forums III and IV, the Group Written Reports (GWR) and the In-Class Rewrite (ICR) involved extended enquiry into probability. Both were different from the forums in genre. While Forums III and IV were forward-looking and involved making predictive statements,
the GWR and ICR were reports on what hypotheses participants had formed earlier and what results had been found after having tested those hypotheses. Moreover, these assignments were framed as reports, and not as discussions.

In the following table (Table 5.30) the data for the GWR and ICR are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the figures for the normalised number of uses of implicit (Imp) and explicit (Exp) premise and conclusion indicators.

<table>
<thead>
<tr>
<th>Implicitly / Explicitly Connective Premise and Conclusion Indicators</th>
<th>GWR, ICR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Token Counts: Written Reports</td>
</tr>
<tr>
<td>P1</td>
<td>14</td>
</tr>
<tr>
<td>P2</td>
<td>14</td>
</tr>
<tr>
<td>P3</td>
<td>14</td>
</tr>
<tr>
<td>P4</td>
<td>20</td>
</tr>
<tr>
<td>P5</td>
<td>21</td>
</tr>
<tr>
<td>P6</td>
<td>20</td>
</tr>
<tr>
<td>P7</td>
<td>28</td>
</tr>
<tr>
<td>P8</td>
<td>28</td>
</tr>
<tr>
<td>P9</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 5.30: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 4

In the GWR, Group 1 used no explicit premise and conclusion indicators (see Table 5.30). Compared to their performance in the Forums III and IV, Groups 2 and 3 each maintained their use of both explicitly and implicitly connective argument indicators somewhat, although at about 75% their earlier frequency.

Comparing the frequencies of token use between the GWR and the ICR, with the exception of P2, who used no explicit premise or conclusion indicators, all other participants (except for P5, who was absent on the day of the ICR) maintained their use of explicit premise and conclusion indicators. Participants’ use of implicitly connective premise indicators varied. P1, P2, P3, P4, like P6 (see Figure 5.26), maintained about the same frequency of token use in their ICR as their group during the GWR. In contrast, while P7’s (see Figure 5.27) use of both implicitly and explicitly connective indicators decreased greatly, P8’s (see Figure 5.28) and P9’s increased greatly. That seven of the nine participants were...
able to either maintain or increase their frequency of token use without the advantage of having access to the Master MDM List, indicates that the indicators presented in the OTREs were retained in working memory (Robinson, 1995) and were accessible to participants during the ICR. That the other participants were also able to use some of these, rather than only a few or none at all, also indicates that the OTREs were effective in influencing participants to make frequent use of tokens.

Figure 5.26: P6’s frequency of Implicitly and Explicitly Connective token use for Chat Rooms I and II
Figure 5.27: P7’s frequency of Implicitly and Explicitly Connective token use for GWP and ICR

Figure 5.28: P8’s frequency of Implicitly and Explicitly Connective token use for the GWP and ICR
5.5.7. Implicitly Conditional statements in the GWR and ICR

In this section, I discuss the final written output of the participants. I point out the trends seen in the use of tokens expressing conditional statements which appeared in the GWR and the ICR. Because I am discussing the token use frequency and not the appropriacy of that use, I present the frequency of conditional statements, including hypothetical premises, collectively. In Chapter 6, I discuss appropriacy of use of conditional statements and hypothetical premises separately. I also show how participants responded to the task by attempting to discuss logical or causative connections between sets of conditions.

In the following table (Table 5.31) the data for Forums III and IV are presented together. Statistical figures such as mean average, standard deviation and range are recorded along with the figures for the normalised number of uses of implicit (Imp) and explicit (Exp) premise and conclusion indictors.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Token Counts: Written Reports</th>
<th>GWR, ICR</th>
<th>GWR, ICR</th>
<th>GWR, ICR</th>
<th>GWR, ICR</th>
<th>GWR, ICR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Freq./100 words (normalised)</td>
<td>Freq./100 words (normalised)</td>
<td>Range</td>
</tr>
<tr>
<td>P1</td>
<td>1</td>
<td>4.0</td>
<td>1.5</td>
<td>1.4</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>P2</td>
<td>1</td>
<td>3.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>4.0</td>
<td>1.5</td>
<td>1.4</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>P4</td>
<td>1</td>
<td>2.5</td>
<td>0.5</td>
<td>2.1</td>
<td>0.7</td>
<td>0.6</td>
</tr>
<tr>
<td>P5</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>P6</td>
<td>1</td>
<td>3.5</td>
<td>2.5</td>
<td>3.5</td>
<td>2.1</td>
<td>0.2</td>
</tr>
<tr>
<td>P7</td>
<td>1</td>
<td>2.0</td>
<td>1.5</td>
<td>1.4</td>
<td>0.7</td>
<td>0.2</td>
</tr>
<tr>
<td>P8</td>
<td>1</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>P9</td>
<td>1</td>
<td>1.5</td>
<td>1.5</td>
<td>0.7</td>
<td>0.7</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Table 5.31: Token counts, frequency, mean, standard deviation, range, total word count for P1-P9, Frame 4

In the GWR, each of the three groups used tokens to express the conditional statements necessary to complete the task. In the ICR, in which participants were tasked to write, from memory, as much as they could recall from their GWR, the eight participants who were present on the day of the ICR all employed a greater number of tokens in the construction of conditional statements, including hypothetical premises (see Table 5.31) than they had in their GWRs. In the cases of P1 (see Figure 5.30), P3, P6, P7, the number of uses of conditional statements and hypothetical premises increased noticeably. In no cases was
there a decrease.

5.29: Average of P1-P9s’ frequency of token use in conditional statement indicators and hypothetical premises in the GWP and ICR

Based on the results in both the preceding section and this section, I conclude that none of the eight students present on the day of the ICR were non-participatory during the composition of the GWR. All played a role in applying, or watching the application of, what had been learned.
in the OTRE sessions and practiced in the forums and chat rooms. The data in this section once again strongly suggests that OTREs and the related work that followed were effective in influencing participants to use conditional statements in cases where those constructions were suitable.

5.6. Participants’ perception of the efficacy of OTREs

In Section 5.5.7, I concluded that the data indicated that OTREs were an influence in participants’ increased use of tokens. This is a conclusion towards which the data has been shown to point, in each frame, from data source to data source. Even so, from the quantitative data alone, it is difficult to determine if and to what degree the OTREs, as distinguished from podcasted lecture content, the Master MDM List, and the example of other more advanced students in the forums and chat rooms, played a role in the increased use of tokens. In the following sections I will present qualitative data from several sources that indicate, in answer to RQ1, that OTREs were an important, and in some cases, primary influence in students’ increased use of argument indicators.

5.6.1. Participants’ understanding of the OTREs purpose

Participants were asked about their understanding of the purpose of the OTREs and the OTRE sessions. Some participants seemed unable to answer comprehensibly or their answers were simply, to learn English. Among the more engaged exchanges were the following. P3 said:

I think the purpose of this WebSequiturs is to discuss with the group and find one answer because it will be easier to solve the problem.

P7 replied, ‘I think especially group work, sharing. Sharing my ideas with group members.’

P1 replied similarly:

To cooperate with my group members and to get the skills to discuss.

Gann: That’s very interesting. Could you comment on the content of the language?

What kinds of things were being targeted?

P1: To get a skill to tell one’s opinion more persuasive.

Gann: What specific language points were the focus of the exercises?

P1: When I do the OTRES, we thought [about] the connection of the sentences. So, we can have a skill to read the article more correctly.

P4 reported that she had seen the focus in practical terms that she had been able to put it immediate practice:
I learned the key words. Last Monday, I have to give a presentation [in a different English class] so, I remembered the *kasetsu jikenshou* (hypothesis verification) conclusion so I could make a presentation with a series of it.

These responses indicate that while participants were aware of the linguistic purpose of the OTREs, they also valued the collaborative aspect of the experience as an opportunity to gain social skills needed to form a shared understanding of the text. The theme of MDMs used as connectives (Hyland, 2005a), connections between sentences (or clauses) called *setsuzokugo* in Japanese, also arose in the semi-structured interviews.

5.6.2. Findings of thematic analysis of OTRE transcripts

These results, categorised in Figure 5.75, show that while there were some instances of discussion of grammar-based and punctuation-based cues. However, there were over twice as many instances of participants discussing argument indicators. The results also indicate that during the completion of the exercises, students were engaged socially, and were not inhibited from disagreeing or correcting what they perceived as other participants’ mistakes. The data also shows that they also collaborated to encourage one another and were emotionally invested in the activity. Fatigue, which had been noticed in the pilot was all but absent in the shortened sessions. While there was some interest in maintaining a high score, it does not appear to have been a high priority. Rather students were more concerned with exploring the meaning of the texts as they often read through the passages together and discussed how best to proceed.

As for themes which respond to RQ1, participants spoke frequently of premise indicators and conclusion indicators, either noting their presence in the sample text or multiple-choice array; or mentioning their absence or speaking of them in expectation of their appearing in a subsequent multiple-choice array. While raters noted 54 such instances, they noted less than one-third that number, 17, utterances about grammar and four mentions of orthographic cues such as capitalisation or punctuation. This distribution of comments strongly suggests that participants were actively influenced to become explicitly aware of the presence and function of argument-related tokens during the OTRE sessions.
Interrater reliability of thematic analysis of OTRE session transcripts and interview transcripts

Two raters, full-time native speaker colleagues at TUS assisted in the thematic analysis of over 10% of the student-student talk during seven of the OTRE sessions. Following Sanford, (2012) in order to consider reliability of the number of themes detected in the data set, percent interrater agreement was calculated by dividing the number of interrater agreements by the total number of detections for the fourteen themes in each of seven sessions. The results of the check for interrater reliability are noted at the bottom of Figure 5.75. The percentage of interrater agreement were as follows: Session 4: 92%; Session 5: 92%; Session 9: 92%; Session 10: 78%; Session 13: 86%; Session 14: 86% and Session 15: 86%. There was an average percentage of agreement of 87.4%.

<table>
<thead>
<tr>
<th>OTRE Session</th>
<th>Session 4</th>
<th>Session 5</th>
<th>Session 9</th>
<th>Session 10</th>
<th>Session 13</th>
<th>Session 14</th>
<th>Session 15</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrater (IR)</td>
<td>R #1</td>
<td>R #2</td>
<td>R #1</td>
<td>R #2</td>
<td>R #1</td>
<td>R #2</td>
<td>R #1</td>
<td>R #2</td>
</tr>
<tr>
<td>Discussion of strategy for completing the exercise</td>
<td>Grammar-based / mention of grammar</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Orthographic/punctuation-based, e.g. upper-case</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Disagreement</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Correction of misunderstandings</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Explicit recognition or observance of metalinguistic content</td>
<td>Premise MARKER / conclusion MARKER</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>conclusion</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Conditional</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Statement / C.S.</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Hypothetical</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Fatigue / Mention of remaining time</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Encouragement to continue</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Emotional engagement: excitement over correct answer / disappointment over incorrect answer /</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Interest in score percentage</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Reading the text aloud / Discussing meaning</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Interrater Agreement: 0.92 0.92 0.92 0.78 0.86 0.86 0.86

Figure 5.75 Interrating tool used in thematic analysis of OTRE transcripts of participant talk

Because there is no section in this chapter dedicated specifically to semi-structured interview responses, but instead, those responses are used to support or nuance the presentation of OTRE utterances, I present the results for the interrater reliability here. The interview transcripts were analyzed thematically and coded. The coded transcript was then rated by two native English speakers. The interrater reliability was then calculated by percentage of agreement to arrive at a figure of 93.0%.

Below are some of the common themes that were noted in the OTRE transcripts.
Discussion of procedure for beginning an exercise

Participants began forming their first collaborative relationships by working out the procedure for completing the exercises. It is noteworthy that they communicated with one another, rather than working through it with few utterances.

P3: Doko tekusuto. Kore? The text, WHERE?, i.e. Where is the text (that the instructions at the top of the webpage refers to)?
P1: Kokono tekusuto. (pointing) This text. [Pointing observed in video recording.]
Also, there was discussion of procedure for finishing an exercise:
P5: Matte, yarakashita? I did it?
P3: [unintelligible.]
P4: [laughs]
At the end of one of the first activities, P5 asks group members for confirmation that the exercise has been completed.
P5: Yarakashita? Did I do it wrong?
P4: Insatsu shita kara ne. After printing it out, okay?.
P5: Ah, I didn’t choose last sentence. Insatsu! Print!

Negotiation of answers

It is also noteworthy that, as participants worked through the many steps of each exercise, they continually communicated with one another.
P7: Chotto matte. Just a moment.
P8: ‘D’ ka? Is it ‘D’?
P8: ‘DVD’ kara. From ‘DVD’.
P9: [unintelligible] ‘D’.
P9: Ja, ‘D’ de iiinjainai? Alright. D is good, right?
P7: Un. Hora! Yeah. See! I was right!

These exchanges, suggest that the participants found the activity engaging and intellectually stimulating, adding to the argument for the relevance of these exercises in later increases in the frequency of token use.
Disagreement

Expression of disagreement was considered as a sign of active discourse as participants felt the motivation to express differing opinions. There were not many such cases, but one is presented below:

P6: Sore o jibuntachi ga omotteiru, kedo jijitsu jyou wa chigau mitaina hanasi ja naino? We think it is right but actually we are wrong.
P5: Ja, iika. Well, okay.

They select “A”.

P1: Ja, “B” da ne. Okay, then it’s B.

Such exchanges demonstrate that participants were invested in their opinions, but also reached reasoned consens, collaboratively.

Correction of misunderstandings

Participants correcting each other was also observed.

P7: ‘Twelve’ dayo ne? Tashika. It’s ‘twelve’ isn’t it? Certainly it is.
P9: Twelve…ah, Twelve.
P7: Dakara chigau deshoo? So, it’s not the correct choice then, is it?
P7: Eh? Doko? Huh? Where?
P9: ‘About twelve percent . . . ’ [Pointing observed in video recording.]
P7: Ah, ‘percent’ no hanashika. Kore. Oh, is it talk about percentage? It’s this one.

Here, P7 failed to note the connection between the preceding text fragment and the multiple-choice selection, 12 percent........ where two decades in the sample text was changed to twenty years in the reconstruction. Twelve and twenty are often mistaken for each other by Japanese students. I had also reversed the appearance of the two terms from two decades and twelve percent in the text provided above the exercise to twelve percent and twenty years in the multiple-choice selections. Hence, P7’s reaction to being informed that twelve corresponds to percent and not to years. While it may be said that this does not apply directly to the application’s utility in promoting MDM usage, it does suggest that the social nature of the engagement with the text, as well as the way the application directs language learners to focus
on the text in a way they otherwise might not have.

These are examples of typical ways participants work collaboratively and negotiate meaning. They often carefully reasoned through decisions, although they also often progressed in a seemingly haphazard way. Perhaps, from a researcher’s point of view, I am not privy to subtleties that the participants’ hold in common. The clipped discourse and truncated utterances, which are common throughout the transcripts, are likely due to the shared stimuli on the computer monitor and the participants’ shared background knowledge. This could explain the frequent lack of explicit verbal cues and articulated explanations.

Sometimes, explicit verbal cues were observed, as participants actively employed the points from the lesson to complete an OTRE. In the following examples, participants demonstrated that they were anticipating a premise indicator:

P2: Kondo puremisu da yo. *A premise will follow this.*
In another instance:
P4: Kono hen kana. *Around here, maybe?*
P5: Un. *Yeah.*
P4: Dakara kokokara ippantekina hanashi ni natteru kara… *Therefore, from here, the text is basically general talk…* Kore wa chigau darou. *Maybe this one is not right.*
P5: Ketsuron dewa nai. *This is not a conclusion* [among the four multiple-choices in the WebSequitur].
P4: ‘Because’…”

The last utterance here suggests that P4 notices an indicator and understands the word *because* as a premise indicator. WebSequiturs, in this instance, are likely to have caused students to explicitly focus on the tokens presented and to discuss them from a developing metalinguistic perspective. The frequent questioning, implied in the audio recording by raised inflection suggests tentative hypothesis-testing and a culturally normal maintenance of rapport with other participants. In answer to RQs 1 and 2, the social aspect of the OTRE shapes the decision making process and makes that process more salient via vocalisation.

5.6.3. Prior knowledge and instructions in the use of MDMs (tokens) in AW
In order to better understand participants’ pre- and post-OTRE use of tokens, it is helpful to consider their responses when asked about their previous experience with tokens in high school or in some cases, other contexts. Their reactions were varied in ways that show the diversity of language learning backgrounds:

*Was anything you studied during these exercises*
P1: I know some of the words but others I didn’t know so I think I have a good experience.
Gann: Could you tell me, what were some of the words you didn’t know?
P1: Hence . . . Actually.

P2 responded, saying, ‘Before this study I did not know academic argument words . . . so I can take it [in].’ P2 expressed this idea of internalizing words and phrases again later when he used the word, *teichaku*, meaning to take in meaning and to literally fix it permanently in one’s mind. P3 continued in this vein:

[Trans.] I learned about discourse markers [. . .] and if you use discourse markers, this part of the sentence, and what role MDMs play in it. It is possible to come here or go there, and so, to call it a map, a destination, a writer, or to guide what the writer wants to convey is effective or important.

This was an interesting comment as P3’s metaphor is one of a traveller’s tool and calls to mind similar Vygotskian concepts discussed in Section 2.7.2 and Section 5.6.7, where P2 spoke of tokens as weapons and tools.

P5, P6, and P8 all indicated a passing familiarity with some tokens but a lack of knowledge of their appropriate use, which is the main theme of Chapter 6.

- P5: I know some of these words, but I don’t know how to use correctly, so these words almost new to me. [. . .] [Trans] Before taking this class, I didn't write or read much academic texts, so I thought it was important, but after all, what exactly is the language of the discussion? I didn't know, so it was a vague feeling, but I feel like I've grasped it a little after taking classes.
- P6: Well, some of them I already know of course because my father is American and we will talk in English at home. But still there were quite a few words I didn’t know and [it was a] good experience learning.
- P8: Yeah, I didn’t know. [It] was new.
  Gann: Can you give me some examples?
  P8: [Trans] I didn't really understand what the language of discussion was, so everything that came in was new.

P9, by contrast, indicated that she had had some prior instruction in this area:

- P9: I knew a little critical thinking or indicator words for example, before, *because, so . . . setsuzokugo* (connectives; in Hyland (2005a) *transitional*
markers. (see Appendix 1)).

Gann: Where did you learn those words?
P9: When I was in high school.

I understood this as saying that P5 and P9 had come to appreciate that other than topic-specific vocabulary, metadiscourse markers are needed to control the logical flow of the text. I also wished to inquire into what extent the participants felt the WebSequiturs were related to this new awareness. This answer was generally representative of the others as we see in the next question when P2 and P3 each tell how the class changed their perspective on writing.
P5 responded:

P5: Before the class, not so much. I don’t know how these words is important. I didn’t read so much academic sentences. So, after the class, I know these words is important.
P3 answered similarly and elaborated:

P3: [Trans.] I think in the dissertation, the language of the argument is important. Since any person should draw the same conclusion from the [writer’s] point of view, if the interpretation of the information differs depending on the person, then the . . . [writer has failed to] accurately convey the results of the research. The value of research also is reduced.

In other words, tokens are understood by P3 not only to control the clarity of the text, but also how that text is filtered within the writer’s intended discourse community.

Gann: What were some of the specific language points that we studied?
P6: Hmm, one is definitely premises. Can I mention like . . . [he digresses momentarily] the importance of discussion—but that’s not the answer . . .

Gann: So, we looked at simple premises and we looked at conclusions.
P6: Oh, that sort of thing? Like hypotheses.

P6 thus demonstrated a firm familiarity with the major types of tokens that had been modelled through the OTREs. Like P3, P6 similarly commented on how the use of tokens raises the intellectual quality of discourse:

Even though I know I probably won’t use when I talk. When I took this course, it sounds like it is pretty important. If those words [that are used in] premises were not included vs those words [that] are inside, those premises are included, those words are more cooler and more like better discussion is like going on, sounds like. When I listen to those podcast words I thought, wow that’s more professional discussions. So, yeah I thought it was pretty cool. Before taking this course I wasn’t thinking so important, but
after this course, it changed.

In other words, P6 had considered the tokens he learned during this study to belong mainly to written genres and he did not consider the tokens to be the kind of language he would use in spoken discourse. However, during the term his perception changed:

When those premises, those kind of words, those premises on the pdf, or like you taught us to use by the podcast, if those words are used in a discussion it’s more professional but if those words are not included then it’s just kind of like normal talking and it’s not like real discussion.

Gann: I see.

P6: So that is what I was like—those kind of English is what I was using before but by taking this course it feels like those premises and those words will help to have discussions. The people who didn’t get a high education don’t use those words.

Later, P6 returned to this theme saying:

For other people I’m talking with, they can easily understand what I’m saying. Like what I’m agreeing about, or, what’s the point. Or what level I’m agreeing at, or I agree or not. It’s more easier to understand for the other people listening to what I’m saying.

In other words, some words which had seemed predominantly written were recognised as having the potential for genres outside that of strictly written argument.

An interesting case was P5, who did not contribute to two of the five forums and was absent for the In-Class Rewrite. Although he had been as attentive as most of the other participants during the OTRE sessions, all of which he attended, he was also the least articulate of the nine participants when interviewed about the structures which had been modelled:

Gann: What were some of the words used, for example, when we did a class project, I asked you to write a kasetsu.

P5: Hypothesis? [Kasetsu means hypothesis.]

Gann: Sure.

P5: Validity.

Gann: Any specific words.

P5: ‘I think.’

I attempt to guide him to the topic of hypotheses.

Gann: Dealing with hypotheses?
Nevertheless, in the forums and chat rooms in which P5 did participate, his performance was average, and at times above average, except in the area of variety of tokens within specific communicative functions. Unfortunately, because of his absence on the day of the ICR, we do not have that data to further triangulate. Importantly here, although his general averaged level of appropriacy for tokens was low in Chat I (where access to the Master MDM list not permitted) was it increased to a level higher than his group’s GWR level. Moreover, his levels of appropriacy for the construction of conditional statements and hypothetical premises was consistently in the Average range.

Various language learning backgrounds were represented among the participants. It is arguable that that simply having had exposure to words like *because, since, so* and *therefore* does not equate with having developed a metadiscoursal understanding of the use of those words as argument indicators. The responses indicate that during the ISE-1 course, participants underwent an experience which led to a greater awareness of the presence and appreciation for the function of MDMs in argumentative writing or writing which, in their view, expressed their opinions and the reasons for the beliefs that they held. It also suggests that the more engaged students were with the OTREs and the tasks that followed, the greater the degree of awareness reached. It was not necessarily the case that non-participation in forums equated with lower degree of awareness. Nor did the apparent lack of ability to speak articulately about metalinguistic issues correlate with low performance on forums in which he did contribute.

5.6.4. The influence of grammatical cues vs argument indicators

An important issue to explore was the role the sentence breaks played in participants’ completion of the OTREs. There are two grammar-related concerns for the OTRE-maker who wishes to direct students’ attention to premise and conclusion indicators. Japanese students, after years of English education with an emphasis on grammar are typically poised to notice grammatical cues rather than any new cues such as argument indicators that they have been directed to look for. For that reason, sentence breakpoints in OTREs should be distanced from grammatical cues so that students are maximally directed to focus on argument indicators. We see a typical example of collaborative use of grammatical cues during an OTRE session:

P2: “Because of”.
P3: Meishi . . . (Noun).
Here, the way P3 finishes P2’s tentative utterance with a cue regarding grammar further suggests that the social constructivist nature of the interaction is helpful in achieving a mutually held metalinguistic understanding of the text as is further supported by comments from the interviews. At first glance, it seems that the premise indicator, because, was noticed. However, P2 and P3 seem to be more focussed on noting that a noun must follow the preposition of. In this instance they are thinking of the highly localised several-word region rather than on a paragraph-wide level. I had attempted to pre-empt that by including more than one multiple-choice selection beginning with a noun. Due to the application’s random multiple-choice generation, however, the ideal multiple-choice array did not always appear.

A second issue is that grammatically valid cues do not inadvertently occur as a result of an incorrect multiple-choice selection being presented by the application’s random multiple-choice array generation function. While I had worked to pre-empt such occurrences, I wanted to hear from participants about their own experiences.

Gann: How helpful was grammar?
P1: It is not so important, but to do the exercise, it is helpful.
Gann: When I say grammar, I meant like s’ on the ends of verbs or is and are . . .
P1: It’s very useful.

P2 also said:

P2: Yes, it was. I used grammatical aspects to clear the exercise.
Gann: Sometimes or occasionally.
P2: If in choices there is [begin with a] capital [letter].
Gann: Capital letters, sure. Other than that, did you use grammar much or did you use your understanding of argument? See there are two things here—there's understanding of argument and there’s grammar. So, I’m asking, how often did you use your understanding of grammar and how often did you use grammar.
P2: Both.

P4 made an interesting comment about the length of the sentences and how that affected the ability to predict subsequent language content based on grammar.

P4: Grammar or new words?
Gann: Right. Grammar or your new knowledge of the indicator words and the structure of argument.
P4: Structure of argument is more important.
Gann: How about grammar? Was grammar also helpful?
P4: Yes, yes. But long sentences, I can’t answer from grammar. So, I think the [language of] critical thinking is important for it.

This speaks to the particular qualities of WebSequiturs, (see Section 3.6.6). P5 said ‘70 [percent] or more’ of the time he tended to look for cases of the first or third person depending on which had been used up to that point. P6 spoke at length on the matter in a way that contrasted these previous responses:

P6: Well, there was one day actually after the class was done and me and (P5) and we were talking about the WebSequiturs and actually when we did the WebSequiturs in class and after it was done, I remember he was saying when I try to do this with grammar it be many times chigau (wrong). I remember he was saying that sometimes or quite a few times it was wrong.

Gann: Well there could be more than one—looking at grammar alone could lead you to make two or three choices.
P6: Yeah, so maybe those questions if you just think about grammar a lot of times it’ll be wrong. So, it was effective [in directing attention toward the indicators] or not? The answer for me was, it was effective for me.

P7 likewise responded with: ‘Almost [all] indicator words. Few times bunpo (grammar).’ P8 also said: ‘Almost (all) indicator words. Few times bunpo (grammar).’ P9 said, ‘About 50/50.’ [Trans.] ‘Grammar was not so important,” and ‘I think I was conscious of the language listed [in the Master MDM List] on the [LETUS] website.’ However, immediately after this, she also noted using orthographic points, saying, ‘Omoji (capital letter) big.’

Somewhat related here is the role of punctuation. The following are two examples of participants referencing punctuation during an OTREs session. P1: Piriodo tsuiteiru. (About the period).

In another exchange, one participant noted that the correct selection would require a comma:

P9: Kanma tsuiteiru. (With a comma.)
P7: Tsuiteiru. Tsuiteiru. (With. With.)

Just as I had worked to minimise grammatical cues through careful sentence-break placement, I also worked to avoid allowing punctuation to influence students’ decision
making during the multiple-choice selection. Nevertheless, the transcripts of student-student talk during each group’s OTREs include many such utterances. In the case above, the comma was purposely absent at the end of the phrase fragment, but present at the beginning of the correct multiple selection, which was a token taking the form of a relative clause. Again, the interaction between participants working within a zone of proximal development (ZPD) (Vygotsky, 1978) helped participants to reach a successful reconstruction of the text.

5.6.5. The influence of argument indicators
I have shown various cases in which participants focused on components other than the intended argument indicators. In the following sections, I will present cases in which participants demonstrated they had relied upon argumentative structure-based and MDM-based strategies to determine their selection during OTREs.

In the following passage, Group 3 deliberates over their next choice, drawing upon their developing metalinguistic knowledge of argumentative structure:

P8: Hai. C. (Yes)
P7: D wa chigau yo tabun. (Maybe it’s not D.)
P8: Chigau? (It isn’t?)
P7: "Because" wa tsuzukanai desyo tabun. ("Because” doesn’t continue on, I suppose.)
P9: Tsuzukanakumo nakunai? ([Agreeing] It doesn’t continue?)
P7: Sou? Jya wakannai. (Really? Hmm, I don’t know.)
P9: Datte kyoumi motte, click shiterundeshoo? (But, if we want to know, we’re going [have to] click [something].) Sono riyu nobetereba hairuyo. (If you tell me why, I’ll enter it.)
P8: Ikai dete kitayone? (That answer came out one time before.)
P7: Kokorahen ka? (Around here?)
P8: Ah hontoda. (Ah. [agreeing] about there.)
P9: "A" wa chigau yo. (It’s not “A”)
P7: "C"?
P8: Un chotto watashi mo so omota. (I kind of think so too.)
P9: Un ne. (Yeah, yeah.)
P7: "C" dato. ("C”, I think.)
P8: Ja "C" de iko. (Well then, let’s go with C).

What makes this exchange particularly noteworthy is P8’s referencing of an earlier similar
case. This indicates new knowledge which is being applied in a slightly different context.

Another example occurs in Group 2:

P5: ‘Because’ wa nai ne. There’s no ‘because’.

Here, P5 notes that if a premise were coming up soon, that a premise indicator ought to be present—but isn’t. In another example, P5 enthusiastically argues for one selection:

P5: Ah! Kore kana. This one, maybe!


In this also noteworthy exchange, P4 observes that, given the length of the source text, selections with more general vocabulary are likely to follow; and it is too early for the anticipated premise indicator to appear in this conclusion-first, premise-last argument.

The responses made in the interviews and the excerpts of student-student talk taken from the OTRE sessions add weight to the argument made in the preceding sections for the use of text reconstruction in the instruction of AW. Perhaps one of the most powerful aspects of the OTREs was simply the potential for students to work together in a social constructivist context. Another is the collaborative relationship with the computer itself. Yet another is the particular affordances provided by the WebSequitur application.

5.6.6. WebSequitur-specific pedagogic affordances

In response to the question, P6 took an unexpected turn and began commenting on the narrower topic of WebSequiturs. His insightful comments strongly support the claim that WebSequiturs have unique qualities that make them particularly well-suited to teaching the structure of argument and the role of premise and conclusion indicators.

P6: We couldn’t see the whole sentence from the beginning, so we had to guess—we really had to think what kind of—well there are the four choice and what do they all say?

Here P6 touches on the point that I wrote about in Section 3.6.6. The linear quality of the text reconstruction is quite different from a drag and drop exercise where all the fragments are visible. The non-visibility of fragments on the back end of the text passage and the fragment-by-fragment reveal sends the user down quite a different series of steps and the experience is very different from that of any other commercially available application. That is because, as P6 continues:
And what do you think and if you choose this word what kind of sentence do you think will come later? Well, I think maybe this word will come in and, until here, it’s this kind of story and from here, this kind of story will come up and it’s a little like we did talk like that because we don’t know what is the whole story so we just imagine. So, maybe this is the correct one or if this is [. . .] We didn’t see the whole sentence. We like—used our brains.

This statement about how WebSequiturs lead the language learner to imagine, or perhaps more accurately, guess what language content will follow. This is very much in following with Goodman’s (1995) conceptualisation of language learning as a hypothesis testing or guessing game mentioned earlier in Section 3.6.5.

P6 then concluded his comment, reiterating those two main points, and noting how much less effective a teaching tool the OTREs would be, had they resembled a fill-in-the-blank, a drag-and-drop or some other format:

If we can see them all [all the missing words] from the beginning, then we really don’t have to imagine as much as the type of WebSequiturs we did. So, we can see [could have seen] what kind of structure is over here. [And we could have then guessed] So maybe this kind structure is going to come over here.

P6’s comments advocate strongly not only for the social constructivist aspect of group knowledge building and working within a mutual ZPD during the OTRE session, but also for the specific qualities that are provided by the WebSequitur application.

5.6.7. Participants’ perception of MDM’s (token’s) influence
In this last section covering semi-structure interviews, I reflect briefly on the participants’ perception that they had increased their knowledge and skills and, in some cases, reached a higher awareness of language use. P2’s following statement was perhaps more poignant than he knew:

P2: I am not an English speaker, so to have weapons is good to use.

The idea of MDMs as weapons is, at first, amusing, but this conceptualisation is very much in line with the Vygotskian idea of words and utterances as tools. Japanese students often mistake the words weapons and tools. MDMs are tools with which specific communicative functions can be engaged in order to achieve specific communication goals (Vygotsky, 1978).

P3 commented on his developing view of AW from the point of view of how a lack of
MDM-related writing skills would adversely affect the writer’s ability to effectively communicate:

P3: [Trans] In my dissertation, I think that the language of discussion is important. No matter who sees it, they have to bring it to the same conclusion, so if the paper is interpreted differently depending on the person, it will not be possible to accurately convey the results of their research. The value of research is also reduced.

P5 likewise was certain that he would ‘use these words . . . [Trans.] when I become a graduate student.’ P7 noted that, in his opinion, his group had not discussed the OTRE texts carefully enough. Nevertheless, he felt that the OTREs were very important, saying:

[Trans.] If you don’t do the WebSequiturs then you [literally] ‘won’t get those words in you’, and you probably will not be able to use them during the chat room or report.

Here we see the theme of internalisation. That is to say that P7 recognised that through the OTREs, he could move through the developmental steps of interlanguage (see Section 2.7.5), by way of imitation toward automaticity (Lantolf and Thorne, 2007). Following this he expressed his determination to put what he had learned to use:

P7: I think very high possibility [that I will use these MDMs]. Yes, after graduating university, I likely use English, I likely write English report.

Gann: I when you write will you be able to remember the language you studied in my class?

P7: I will use those words.

P8 said ‘I didn't really understand what the language of discussion was, so everything that came in was new.

P8: [Trans.] Before taking this class, I didn't write or read much academic texts, so I thought it was important, but after all, what exactly is the language of the discussion? I didn't know, so it was a vague feeling, but I feel like I've grasped it a little after taking classes.

This suggests that this study’s course design, of which participants agreed OTREs were an important part, have a strong pedagogical basis. Participants acknowledged their efficacy and expressed a sense of having undergone transformational development as a result of its implementation, after which they moved from the other-regulation during the period in which they worked through the OTREs to self-regulation in the GWR and ICR (Aljaafreh
and Lantolf, 1994).

5.7. Discussion
The findings in this Chapter support many of the major claims found in the background literature surrounding OTREs in AW instruction and constitutes potentially important additions to the body of knowledge on their effectiveness. There are distinct challenges in Japan to teaching AW (Wingate, 2007). However, the quantitative findings support the idea that by modelling western conclusion-fronted paragraph structures, WebSequiturs can be used to expand beyond the kishooten-ketsu form which Maynard (1998) frames as culturally and institutionally entrenched, and as found in Kubota (1998) and Kobayashi and Rinnert (2008), that students can develop new written rhetorical habits as a result of instruction. At the same time, the findings do not contrast with sources such as Bao (2014) who points to culturally determinant factors such as a high valuation for politeness and reserve. Participants were able to increase their use of explicitly connective premise and conclusion indicators. This was generally evident as the data analysis showed an increase from the original baseline established in the first pre-OTRE forum and the increased frequency of use and an increased use of unique tokens. A wider variety of implicitly connective premise and conclusion indicators was also seen in the written output following the OTRE sessions. The quantitative findings, therefore, answer RQ1 by showing an association between OTRE sessions and increased use of the OTRE-modelled tokens.

The qualitative findings arrived at through thematic analysis of the OTRE transcripts and the semi-structured interview transcripts. Of the studies involving Hot Potatoes exercises, only Čechovičová (2010) and Rodríguez (2010) asked their students to express their opinions or impressions. Neither monitored student-student talk during their work sessions. Čechovičová (2010) only questioned her students casually. Rodríguez (2010) did distribute a questionnaire. However, the questions, by her own admission, were not rigorously designed; nor was any systematised analysis reported. This study greatly contrasts those two studies, by posing a carefully designed and well-ordered battery questions that probed into how participants experienced specific and discrete aspects of the OTREs, upon which the answers to RQ1 are hinged. The findings add an extra dimension to the body of available research on this subject. They establish that participants’ use of tokens increased in terms of several relevant metrics and strongly support the claim that the influence of the OTREs played a major and perhaps primary role in that increase.
5.8. Summary
The research context in this action research study is complex and it is not always easy to isolate the influence of the OTREs from the influence of the podcasted instruction before the OTREs or the influence of other interlocutors in the forum and chat rooms at the time that each participant was making his or her own contributions to the discussion. In this Chapter, I have attempted to unravel some of those strands.

I have presented both quantitative and qualitative data supporting my argument that the use of OTREs for the purpose of enhancing instruction of MDM use in AW resulted in the increased use of both explicitly and implicitly connective premise and conclusion indicators. I showed that this increase occurred both within specific communicative functions that it extended across a number of communicative functions in such a way that could both be used as premise and conclusion indicators.

I have also presented responses given by participants to interview cues in concert with examples of student-student talk from OTRE sessions which speak to the veracity of those interview responses. Both the interview responses as well as the OTRE session student-student talk excerpts provide qualitative evidence supporting the claim that OTREs can be an effective tool for increasing the use of MDMs.

In the following chapter, I take a more qualitative approach and draw primarily on participants’ written output to demonstrate that OTREs can be an effective tool for modelling the appropriate use of MDMs in AW and for guiding language learners to become better writers of argumentation.
Chapter 6: Appropriacy of MDM (token) use

6.1 Introduction
In this chapter, I address RQ2 of this study as I discuss the data supporting the efficacy of WebSequiturs in guiding participants to use tokens common to argumentative writing appropriately in terms of acceptability within the bounds of academic discourse. I also continue to explore how the qualitative findings help to answer RQ3, that is how participants experience the OTREs and what their perceptions are after having completed the English course. Following on from Chapter 5, where I presented quantitative data, supported by qualitative data, showing that participants increased their frequency of token use after OTRE sessions; in Chapter 6, I show that participants also achieved higher levels of appropriacy in their token use following each OTRE session. Moreover, I present data reflecting three independent raters’ assessments of the written output from all four frames introduced in the previous chapter. These include: Frame 1 (Forums I, II and V), Frame 2 (Forums III and IV), Frame 3 (Chat Rooms I and II) and Frame 4 (Group Written Report and In-Class Rewrite).

Within each frame, I examine trends in participants’ use of premise and conclusion indicators following the first three OTRE sessions. I then look at their constructions of hypothetical premises before and after the fourth and final fifth OTRE sessions in which those structures were presented, the scale (introduced in Chapter 4) used to measure appropriacy of token use in AW, the scale I developed to categorise different approaches to constructing premises, and I explain how that scale developed as these categories emerged through my initial analysis.

Finally, I discuss the findings in terms how they directly address RQ2 and RQ3.

6.2. Independent assessment of appropriacy of token use
In the following sections, I address the qualities of the participants’ written output which respond to RQ2. Figure 6.2 presents two closely related data sets. The first is the average of the levels of appropriacy for token use in the four frames, that is, in all sources of written output by the participants. The second data set is what I have termed the High Points. The need to present these emerged as there were cases in which some participants had made level-3 Skilled or level-4 Exceptional uses of tokens, but because of the much greater number of level-2 Average uses in their assessment score sets, their presence was not apparent. The impact of the fewer Skilled or Exceptional token ratings was typically diminished by the much greater quantity of level-2 assessments of token use.
Table 6.2: Number of assessments for each level by Raters 1, 2 and 3

<table>
<thead>
<tr>
<th>Level</th>
<th>Rater 1</th>
<th>Rater 2</th>
<th>Rater 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>10</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Level 2</td>
<td>797</td>
<td>802</td>
<td>799</td>
</tr>
<tr>
<td>Level 3</td>
<td>83</td>
<td>81</td>
<td>84</td>
</tr>
<tr>
<td>Level 4</td>
<td>10</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

Several such cases where high ratings are lost in the averaging process are evident in Figure 6.2 when the *High Point* is compared to the average during the same activity. The High Point is the highest scoring assessment given to any one token or token structure during an activity. These High Points become more noteworthy when we look at individual cases. In cases where I am not referring to the High Point, but the average of the total of ratings for a given text, I use the term *general averaged level of appropriacy*. In the graphs this is shortened to *Appropriacy of Token Use*.

The average for all nine participants’ levels of appropriacy for token use in all four frames is presented below (see Figure 6.2).

![Figure 6.2 Average of nine participants’ token use by levels of appropriacy](image)

In Figure 6.2, it is noteworthy that, within each frame, (FI, FII, and FV; FIII and FIV; CI and CII; and the GWR and the ICR) there is an increase in the average appropriacy assessment; and there was either maintenance of, or an increase in, each High Point.
assessment—except for the ICR assessment where there was a decrease in both the average assessment and the High Point assessment. This is explained by the lack of access to the Master MDM List, the solitary nature of the task, and the restricted time limit. It was also possible that the decreases might be explained by participants who were not contributing to or otherwise engaging with the GWR task. The ICR content evidence showed that this was not the case and all members of the three triads collaborated actively within their group.

Whereas Forum I is unique in that it was assigned before the first OTRE session, during the following three OTRE sessions, levels of appropriacy rose even when the Total Word Count (see Figure 5.1) and the Token Count (see Figure 5.15) decreased. Similar to Forum I, Forum III was also completed before the next new tokens and structures were presented—in this case during the third and fourth OTRE sessions which presented hypothetical premises. After these two OTRE sessions, token use appropriacy increased both in terms of average assessment and High Point assessment. Meanwhile, in the Chat Rooms, the average for High Points remained the same, but at the Skilled level, and the average for token use appropriacy increased to the highest of all nine output sources. The only downturn in appropriacy was between the ICR and the GWR that preceded it. Even so, the High Point assessment was maintained at the Skilled level. I interpret this maintenance of appropriacy level as suggesting that participants were guided by OTREs to use tokens with greater appropriacy, and that in various genres, and with or without access to a written list of tokens, they were able to increase, or at least maintain, newly attained levels of token use appropriacy.

6.2.1. Examples of appropriacy assessment
In this section, I discuss the factors (see Figure 6.3) involved in the assessments of participants’ written output presented in Section 6.3, which follows. The final assessment was a single rating arrived at by averaging the three values, and in three cases, after conferring with raters and negotiating a consensus. I discuss these factors by way of examples of the output of participants P3, P2, P5. I purposely chose these three examples, because they received higher than average assessments for the presence of appropriately skilled and exceptional use of a variety of tokens. I comment upon them in the following section to distinguish what was required to achieve averaged ratings above the level of Average.
In Figure 6.3, above, are three examples of written output and the rationale used to assess them. Each of these cases were interpreted as indicating a purposeful use of tokens as contrasted with uses which were indistinguishable from other average token uses. As noted in Section 4.2.3, the Roman numerals which I had initially used for coding served to alert the raters to the tokens which required rating. The final ratings were as follows. P3 received an assessment of Level 4 (Exceptional) (for successfully forming a conditional statement and for building a linked structure of tokens across multiple sentences. The word invalid was rated as Skilled as it was newly introduced vocabulary, rarely used and is appropriate for the context. The same rules of evaluation also apply generally for the second passage written by P2, where a structure of tokens, again, related to forming a propositional statement is also extended across multiple sentences. The errant grammar in the latter clause of the final sentence, written by P5, was not considered to be an issue. What was considered important was the extent to which the phrase, is generally believed, was appropriately applied. The phrase in question was adjudged by the raters to fall outside the range of average participants’ normal use. It is also more grammatically advanced with the -ly suffix and its passive voice necessitating the -ed suffix on the main verb. However, the participant appears to have mistaken the second half of the construction, where the phrase that it is was used instead of the correct to be; or where the phrase It is generally believed that minus (negative) ions are could also have been used. For that reason, the use of this token was assessed as a 3 and not a 4.
In these ways, participants were able to achieve higher than Average ratings. In the
cases of participants where this happened only once or twice, the participant’s output was
seen as generally average. Cases in which above average ratings were achieved consistently, a
pattern was observed by which a participant’s level of competence could arguably be inferred
to be at the level of Skilled, as defined in this study. It was noted that achieving an averaged
rating of level-4, Exceptional, was not possible, due to the fact that a level-4 rating structure
was necessarily made up of multiple tokens used at the levels of Skilled and often, Average.
That is to say that the use of the token, because, for example, would most often be rated as
Average; used within a structure however the culmination of tokens used at levels of Average
and Skilled could achieve an additional rating of 4, Exceptional.

It might be questioned, why the structure should not simply be rated as 4 without
rating the tokens which made up its construction. In this way, the cases in which 4s occurred
would make up a more representative presence in the overall average of the whole text. There
are three reasons that I did not arrange for the rating to be done in that way. First, doing so
would disrupt the marking system of Roman numerals I had left from the earlier
categorisation in order to cue raters to the presence of tokens to be rated. Removing those
would have alerted raters to the presence of what I had identified in my initial rating as a
multi-clause structure prejudiced the raters to rate the structure as a 4, since they had been
instructed to rate nearly all such structures as Exceptional. Second, in cases of short
contributions, doing so could have a volatile impact on the overall average; or in cases of long
contributions, the noteworthy achievement of a 4 might be lost amidst a large quantity of 2s
and 3s. Thirdly, eliminating the scores for the tokens making up a multi-clause structure of
meaning would have made it made cases in which the constituent tokens were used skilfully
indistinguishable from cases where their use was Average in appropriacy. For this reason,
instead, the separate High Point rating was introduced so that cases of Exceptional levels of
appropriacy of token use, including multi-clause structures of meaning could also be utilised
as an important component of this study.
6.3. Assessment of levels of appropriacy for premise and conclusion indicators (tokens)

In this section, I present the data, derived from an averaging of the three raters, for the levels of appropriacy for Frames 1 to 4 (see Section 6.1). This data also includes the High Points (see Section 6.2) for each written output source for each participant.

Again, as in Chapter 5, I present bar graphs as visual aids. I present the lowest assessed individual participants’ data when it presents generalisable trends in levels of appropriacy of token use among similarly performing participants. Doing so, I highlight groupings of data, while minimising the number of graphs.

**Frame 1: Forums I, II and V**

The data suggests that, as a group, participants not only gradually increased their token use appropriacy, but also increased their High Point assessments in Forum V.

In Frame 1, after a decrease from an initially elevated High Point, P1’s use is seen to be trending flat (see Figure 6.4) for the High Points in the second and fifth forum, but with some increase in the averaged appropriacy level, within the Average level of appropriacy. Of particular interest is his success, in the pre-OTRE Forum 1, at the formation of a conditional statement: ‘If a doctor tell them patients terrible diseases, they will feel shock and) their mental health will be lose.’ He also wrote with equal success: ‘If he don't tell them patients terrible diseases, they may die without knowing their own condition.’

The theme of hypothetical premises is the focus of the latter half of this chapter, but these two formations are noteworthy as, the first example comes as close to being a proper hypothetical premise as any of the later attempts by any of the participants. In Frame 2 or Frame 4, either sentence structure would have gained the writer a High Point of 4 if used as a hypothetical premise. Although the second clause is in the future tense, here, the construction is used as a conditional statement of fact and not a tool for testing a hypothesis. The second sentence is essentially the same, although the may introduces a lower level of certainty into the statement.
Figure 6.4: Levels of appropriacy of token use for P1 in Forums I, II and V

P2 (see Figure 6.5) was judged to have expressed a similar increase for the High Points and the general averaged level of appropriacy elevated from Average to Skilled. P2 made a strong showing in Forum I with:

I think the doctor is right. It might causes other problems, But to tell the truth is his duty, because he is the only person who knows it. [. . . ]

As to tell the truth is the doctor’s duty, the only thing which he needs to consider is telling the truth in good season.

The excerpt above shows that P2 already knew some of the basic phrases included in the Master MDM List and others that were not included, such as using as to mean because at the beginning of the second entry. However, P2 did not construct a system of tokens across the text passage as P1 did. Hence, P2’s High Point reached the level of Skilled, but not Exceptional.
P3, and P8 (see Figure 6.6) similarly, were judged to have expressed an small increase for the averaged level of appropriacy. P3 wrote:

I don't think it was right because considering Teppei's explanation, Mrs. Takayanagi's disease may get worse by the shock. Furthermore, it was not good that the doctor made a decision by himself with ignoring his advise.

P3 followed that with:

The doctor needed to talk more about her disease and how to tell her it with Teppei and their family if possible. For example, about basical method from medical position and doctor's experience of the similar situation and the result, concrete story about shocked her.

In both cases, a meaningful network of tokens can be seen. In the first, the token (transition marker) (see Appendix 11) related to Additional Information (see Appendix 11) continues the meaning of the first sentence through to the sentence that follows, notwithstanding some redundancy. In the second entry, there was a code gloss (see Appendix 11), for example, where in other words would have been more appropriate. P3’s two text passages, above, although written prior to the first OTRE session, serve as an example of attempts at structures, extending across multiple-clause sentences and short paragraphs, which earned
rater assessments for this token use of *Skilled* and *Exceptional*. The inappropriacy of token use of smaller or greater degrees in these two examples serves to show what prevented some written output from reaching assessments of Exceptional and limited those text passages to assessments of Skilled, or in many cases, where no structure was attempted, assessments of Average.

![Frame 1: Forum I, II, V](image)

Figure 6.6: Levels of appropriacy of token use for P8 in Forums I, II and V

In another example that highlights the importance of presenting both the general averaged levels of appropriacy and the High Points, P4 (see figure 6.7) and P9 were judged to have expressed an increase in both general averaged levels of appropriacy and High Points from Forum I to Forum II. However, in Forum V, in both participants’ cases the High Points increased to 4 and 3.5 respectively, while the general averaged levels of appropriacy decreased a small amount. P4’s MDM use developed during the term, as can be seen in the two following excerpts. In the pre-OTRE Forum I, P4 wrote:

> I think Dr. Spencer had the right to tell Mrs. Takayanagi the truth about her disease because she had the right to know the truth about herself. At first, she may suprize and be embarassed. However, she become able to think what she is going to do next ,and she act positively.
In Forum II, P4 wrote:

In my opinion, it is very difficult and sensitive problem. It is because what we advise may hurt people who are overweight. In short, we should not give advice insensitively. It is important tell exactly to them, But it may be batter not to say clearly. Moreover, we should not Physical details.

In Forum I, the use of the tokens was judged to be competent, and the introductory clauses are prominent. P4’s token use in Forum II, however, was collectively assessed as having a little greater appropriacy. Then, in Forum V, P4 wrote the following passage:

Frankly speaking, I think it is not valid. Because the date of the experiment is too small. Although there is no connection between blood type and personality from a chemical point of view, we assume there is connection. Hence, we believe connection. In my opinion, it is very important. This is because the credentials of author influence many people. Therefore, it is very important to check whether the information is correct or not. I am sure that there is a no relationship. I also think that it depends on how to grow up, culture and custom. In other words, the environment around people has a big influence on their way of thinking.

As reflected in the slightly lower general averaged level of appropriacy, no single token use stood out as more appropriate in Forum II. However, the text passage as a whole was assessed by the raters as containing a network of meaning created by the use of tokens so as to extend argumentation across the text passage.

To note the details of this example, we can see how P4 forms a rudimentary argument through the use of tokens. She begins with Frankly, because, although. She regroups and continues the argument with a mid-argument conclusion indicator, in my opinion. However, the conclusion indicator is somewhat inappropriately used, as is only followed by the statement, this is important, rather than a substantive claim. This momentarily shifts the focus from the argument to P4’s reasoning, but returns to the main thread of the argument again with This is because, and hence, and therefore. The sentence beginning with Therefore would be better placed as a last sentence and so might be moved to the position following what now seemed to be afterthoughts. Nevertheless, the independent raters felt that there was ample evidence for a strong intent to develop a structure which could serve as a foundation for conveying an argument.
Figure 6.7: Levels of appropriacy of token use for P4 in Forums I, II and V

In contrast to P4’s example, P6 (see Figure 6.8) was judged to have expressed a gradual rise in averaged levels of appropriacy, but High Points peaked in Forum II.

Figure 6.8: Levels of appropriacy of token use for P6 in Forums I, II and V

In summary, although the trends in Frame I varied somewhat, there was a general increase in levels of appropriacy of token use and in no cases were the raters’ assessments for levels of appropriacy of token use recorded for Forum V lower than those recorded for the Forum I. There was an unambiguous advance from Unskilled and Average performance to Skilled and Exceptional and this is a quality which we continue seeing between Frame 2
through Frame 4. I comment upon some reasons for this in the Discussion in Section 6.6. In the next section, regarding Frame 2, I present the trends which were observed in Forum posts specifically aimed at hypotheses construction.

**Frame 2: Forums III and IV**

The assessments for this frame include markers for general argument indicators and for conditional statements and hypothetical premises. Because these two forums tasked participants to make predictive statements based on set factors, a large amount of the written output and token use in each text was employed in the formation of hypothetical premises.

The averaged assessments, that is the values resulting from averaging the three raters’ assessments of the appropriacy of token use, (see Section 6.1) suggests that participants increased their token use appropriacy level from a 2.06 to 2.26; and that they increased their High Point assessments given by the raters from low-skilled to high-skilled. The High Point here corresponds to the single highest assessment of appropriacy of token use. For that reason, and because the entries in Forums III and IV were generally only two or three short sentences, the High Points did not tend to diverge a large amount from the general averaged level (see Section 6.2).

In Frame 2, P2 (see Figure 6.9) and P3 were judged to have expressed values with a flat trend, in this case, for both the averaged appropriacy and High Points.

![Frame 2: Forum III, IV](image)

Figure 6.9: Levels of appropriacy of token use for P2 in Forums III and IV
In the output of P6 (see Figure 6.10) and P9; and to a lesser degree in the output of P1, there was an increase in the average token appropriacy assessment given by the raters. There was also a greater increase in raters assessment for High Points. In only one case was there a decrease, in P5 (see Figure 6.11).

Figure 6.10: Level of appropriacy of token use for P6 in Forums III and IV

Figure 6.11: Level of appropriacy of token use for P5 in Forums III and IV
Because the High Points were invariably related to the appropriacy of each participant’s conditional statement and hypothetical premise formation, the written output gathered from Frame 2 is largely presented in Section 6.3, which is on the appropriacy assessments of conditional statement and hypothetical premise formation. However, P6’s output does merit examining due to his Forum III and Forum IV posts, which were considerably longer than the other participants’ posts in those two forums. In the following example, P6 was able to develop the appropriacy of the use of tokens. In Forum III, responding to the simpler question, P6 wrote:

There are three cups and I have only one chance to choose a cup. The total probability of those three cups adds up to 1(100%) and the probability of each cup has to be equal. Because of this, I think the probability of choosing the correct cup is 1/3 (33.33...%). My hypothesis is there is a one in three chance that the correct cup will be chosen.

As with the other participants, this passage suggests that P6 does not yet understand that hypotheses are tools for testing claims. More to the point, his use of token use is Average in appropriacy. The post is long enough to support a multiple clause system of tokens. However, he does not exploit this potential. There are no introductory clauses, nor any other metadiscoursal devices. Other than the conclusion indicator, because of this, which references the premises in the previous two sentences, there is no further structure. In Forum IV, having completed two OTRE sessions, the same participant, P6, constructs a much more metadiscoursally developed structure:

The first case that I have to think about is there are three cups to choose from and the rule is that I must change the cup from the beginning. This means that if I choose the correct cup from the beginning (the probability is 1/3), I definitely would end up choosing the wrong cup. Second, if I choose the wrong cup at the beginning (the probability is 2/3), then Mr. Gann is going to show me the wrong cup which I didn’t choose. And this means that if I change the cup this moment, I can definitely choose the correct cup. For this reason, we can see that the probability of ending on the correct cup is 2/3.

In the paragraph above, there is a first and second listing of the relative facts, i.e., premises. Both premises are followed by the conclusion indicator, this means, and finally at the end, For this reason. The gap between Average and Exceptional, which was narrowed after the
completion of OTRE Sessions 4 and 5, strongly suggests a guiding role in P6’s developing ability to appropriately use discourse markers. The guidance in developing appropriacy touches not only on hypothesis formation, but also the explanation of the context leading up to the hypothetical premise.

Summing up this Frame and the general appropriacy levels for the nine participants, although there were cases of non-participation, the trend in four of the seven cases were judged as increases in both average appropriacy of use and High Points (see Figure 6.2).

In the following section, regarding Frame 3, I present several findings which involve synchronous communication and timed in-class communication.

**Frame 3: Chat Rooms I and II**

In examining the data from the two chat rooms, I return to the theme of general premise and conclusion indicators. In short, the average of the nine participants’ levels of appropriacy for token use (see Figure 6.2) in the chat rooms followed the same trajectory as that in Forum III and IV.

P1 (see Figure 6.1), and similarly, P3, P5, P7 continued to maintain the average level of appropriacy.

P2 (see Figure 6.13), P4 and P8 were judged to have expressed an increase in appropriacy from Average to Skilled; and P6 (see Figure 6.14), who was at the Skilled level in Chat Room I, reached the level of Exceptional in Chat Room II.

![Frame 3: Chat Room I, II](image-url)

Figure 6.12: Level of appropriacy of token use for P1 in Chat Rooms I and II
Figure 6.13: Level of appropriacy of token use for P2 in Chat Rooms I and II

Figure 6.14: Level of appropriacy of token use for P6 in Chat Rooms I and II
P9 (see Figure 6.15) meanwhile, was judged to have expressed a drop from Skilled to Average. As stated earlier, many of the increases made in appropriacy of token use in Frame 2 were repeated in Frame 3. However, on an individual basis, apart from the case of P6, there was no clear trend. A typical example of a Skilled assessment was P2’s response below in which he concurred with P5:

P2: ‘Skills that I used’ means to use English collectly.
P5: That cannot be happen when we learn alone.
P2: That makes sense.

Rather than using an Average level token, such as I think so too or I agree, which many students learn in high school (as evidenced by the Wisdom dictionary and Genius dictionary asterisk system (see Section 3.4.2), P2 used That makes sense, which is not a phrase often used by first-year Japanese university students. Hence, this use gained a level-3, Skilled, appropriacy assessment from the three raters. To receive an assessment of Skilled use, it was not necessary to write an extended response. Using a token infrequently employed by incoming first-year students suggests a meta-awareness of different ways that language can be used to carry out a given communication function. The appropriate use of the phrase suggested a sufficient movement on the spectrum of interlanguage towards a broader lexicon to register as skilled.

Of the three participants, P1, P6 and P9, who made gains in appropriacy of token use
in Frame 2, only P6 repeated those gains with an increase. The single decrease which occurred with P5 in Frame 2, occurred in Frame 3 with a different participant, P9. However, P9 can be seen to be among the participants with the greatest gains, seen later in Frame 4. The reason for the lack of clear trends is likely attributable to the small sample size. A sample size of 30 or more is recommended as a minimum for small-scale school-based research (Winterbottom, 2013). With fewer participants, there may be fewer opportunities for commonalities to rise to the surface. On the individual level, it may also be due to a particular participant’s skillset in synchronous and asynchronous genres, e.g., forums and chat rooms. In Section 2.3.4, I discussed the situationality differing of genres. For example, Forums, like the GWR, allow more carefully crafted constructions, whereas chat requires quicker, on-the-spot, decision making and responses.

One such case of a lack of clear trends is P6 who made a somewhat extended entry in Chat I, gained a level-3, Skilled, average of ratings, but no a level-4 High Point. This was in contrast with P6’s posts in Forum IV in which systems of meaning were extended across multiple sentences in a clearly purposeful manner, gaining High Point assessments of level-4, Exceptional. In Chat I, P6 wrote, on the subject of fat shaming:

Well, if we never say anything to them, we can never help them at all. In my opinion, we must carefully look at how that person is? For instance, he/she has a strong mental or has a weak mental.

Later, during the same chat session, he wrote:

My opinion is going to be we should really be careful how we act around them. We don't want them to feel sad, and I want them to live happily as same as normal weight people.

The use of tokens in both of these passages was assessed as either level-2 or level-3, with My opinion is going to be, but there is no system of tokens running from beginning to end. As noted above, the difference in genre, where less artifice might be expected in chat than in forum posts might be one reason for the lack of a token system. The lack of time might also be a factor. In Chat II, however, P6 wrote the following, near the end of the chat session:

In Japan, majority of the classes are just litnening to the lecture. But not all people are good at just listening the lecture to study. I believe that there are many different types of people to [who] study. For instance, lerning by visual, sound, movement, feeling
and so on. This brings me to the opinion that it is great to have a chance to learn things in a different way.

The passage begins with an introductory clause *In Japan*, which he set off with a comma. Argumentatively, and relative to the next clause, *not all people*, the first sentence functions loosely as a major premise to the minor premise which follows it. It establishes a common practice, then shifts to a contrast of that with *But*. P6 next states a belief with *I believe* and follows that with an example, using the token, *for instance*. Finally, the phrase *This brings me to the opinion* is used to state the conclusion about learning styles. The system of tokens that P6 creates has two flaws. First, the phrase, *I believe* was taught as a conclusion indicator, as was, *in my opinion. Brings me to or leads me to* was P6’s own. The sentence *I believe . . .* essentially repeats the propositional content of the previous sentence and sets up the way for the following sentence which adds the details about different sensory modes. In fact, that sentence and the subsequent introductory clause, *For instance* could be excised without an objective loss of propositional content, but at a cost of the more subjective quality of smoothness or flow. For these reasons, the text was the single passage from any of the participants in both Chat Room I and II to be accessed as *Exceptional*. In these ways, P6’s written output serves as an example of a language learner in the process of proceeding through stages of interlanguage development toward greater appropriacy. In isolation each passage provided limited information. Taking a broader view affords the ability to discern developing skills.

Summing up Frame 3, as seen in Figure 6.2, there was a general increase in both the general averaging of levels of appropriacy and in High Points. However, the limiting factors of time between exchanges and the perceived urgency to respond to other Chat Room members before the conversation moved on to other topics, was most likely a factor in the lack of extended systems of token use extending across text passages.

It is useful to end the discussion of this frame with a comparison of P4 and P6. The two participants’ nearly equal achievements in constructing networks of propositional quality culminating in a short argument demonstrates that differently skilled communicators can sometimes perform at approximate or equal levels of appropriacy in varying contexts and within one genre or another. Throughout this study, P6 exhibited a comparatively high level of confidence and familiarity with English. Nevertheless, as we see in this and previous frames, his writing does not often achieve a networked system of tokens. However, he does often achieve High Points in the Skilled and Exceptional range for appropriacy of use of
individual tokens. Likewise, P4’s increases in appropriacy applied only to isolated tokens, with the exception of the excerpts provided in Frame 2, where a longer forum response was posted. Even so, P4’s achievement of a construction of a chain of tokens linked through a text to create an extended argument leading to a propositional statement was equal to P6’s, despite P4’s generally average proficiency. In Table 6.4, I present utterance counts showing that although P4 made far fewer comments about tokens than P6, P4 was able to nearly match P6’s High Point levels of appropriacy of token use. In this case, one of the major commonalities between the two participants was the completion of the OTRE sessions. This further suggests that OTREs are an effective means of guiding language learners to use tokens with greater appropriacy and to construct cogent arguments with tokens interrelated so as to create extended structures of meaning.

**Frame 4: GWR and ICR**

In examining the data from the GWR and the ICR, I continue with the emphasis on the theme of *general premise and conclusion indicators*. Again, the average of the nine participants’ levels of appropriacy for token use (see Figure 6.2) was reflected in the trends of some, but not all, of the individual participants. For example, P1 (see Figure 6.16) continued to maintain the average level of appropriacy.

![Frame 4: Group Report, In-Class Rewrite](image)

Figure 6.16: Level of appropriacy of token use for P1 in the GWR and the ICR

Differing from P1, P3 (see Figure 6.17), P8 and P9 began with its averaged level of appropriacy of token use at level-2 and High Point at level-3 and repeated that performance in
Differing again, still, P2 (see Figure 6.18), P4 and P6 were judged to have expressed an increase in the averaged levels of appropriacy from low to mid-average, while the High Point increased from Average to Skilled.

Contrasting the above cases of maintenance and gains in appropriacy of token use, P7 (see Figure 6.19) was an unusual case where, while the High Point decreased, the average level of appropriacy of token use increased.
Figure 6.19: Level of appropriacy of token use for P7 in the GWR and the ICR

There were no cases of level-4 assessments of *Exceptional*. This was because there were no sustained attempts to construct systems of meaning through *premise and conclusion indicators*. This may be because the participants’ attentional resources for building such structures were expended on hypothetical premises, which are examined later in this chapter.

P4 seems to have made the most successful attempt at creating such a structure which received an assessment of *Skilled*, but not *Exceptional*:

In the first experiment, the probability of choosing the correct cup is about one third of the time because the number of trials is 30 and the number of times the correct cup selected was 10.35 times on average. In the second experiment, the probability of choosing the correct cup is about two-thirds since the number of trials is 30 and the probability of choosing the correct cup is 18.95 times on average. More accurate data can be obtained by increasing the number of trials. Since the probabilities are always equal, the probability of the first experiment is one-third and the probability of the second experiment is two-thirds.

However, this passage lacks the kind of development seen in earlier Forums I, II and V in Frame I. As noted, a variety of factors may have burdened participants’ attentional resources (Truscott, 1998). For instance, being the end of the term, participants may have been prioritizing their time for other classes and may not have been able to devote the time needed
to create such systems of meta-discourse; or perhaps they were experiencing fatigue with the writing task. As P4 wrote at the end of her ICR, even though there was ample time left during the class period, ‘Sorry, I can not finish this paper.’ She left shortly after submitting the paper, presumably to attend to her other more highly prioritised studies.

Summing up, participants showed increases in their skill in using the tokens which had received instruction through OTREs, which arguably guided participants to use tokens appropriately within the bounds of what the raters considered acceptable for first-year students at TUS. They also showed noticeable and sometimes large increases in High Point assessments as they learned to use one or more specific tokens to greater effect. These increases match the increases seen in Chapter 5 in areas such as Word Count, token Count, Unique token Count, and Range of Communicative Functions. Where these did not match is at the general falloff at the In-Class Rewrite task that we see in Chapter 5 for most of these areas. Here, even though students were separated from their groups, prevented from accessing the Master token List or any other notes, they still largely maintained and, in some cases, increased their levels of appropriate use (see Figure 6.22).

6.4. Assessment of levels of appropriacy for conditional statements and hypothetical premises (tokens)

In this section, I present findings from the written output from the Forum III and IV tasks, the Group Written Report and In-Class Rewrite, each of which tasked participants to construct hypothetical premises. These structures are essential to writing about cause-and-effect, problem solving and decision making, all areas related to basic rhetorical modes in most university writing textbooks such as Cengage’s Basic Steps to Academic Writing: From Paragraph to Essay (Taylor and Kluge, 2012). Finally, irrespective of the reasons for modelling conditional statements and hypothetical premises, these are the structures which the OTREs modelled; and one of the two major aims of this chapter is to determine what role OTREs play in guiding participants to use tokens employed as argument indicators more appropriately.

Forum III serves as a baseline as it occurred prior to the OTRE-delivered modelling of conditional statements and hypothetical premises. Forum III (in Frame 1) was completed before the two OTRE sessions (Sessions 4 and 5) that involved conditional statements and hypothetical premises. Forum IV was completed after OTRE Sessions 4 and 5. For these reasons, in this section I have not drawn upon Forums I and II since they occurred before OTRE Session 4 and do not task participants to construct hypothetical premises. Likewise, I
have not drawn data from the chat room sessions, as Chat Room Session I occurred before OTRE Sessions 4 and 5; and neither Chat Room Session I nor II tasked participants to form hypothetical premises, so the content there is not relevant.

I have drawn upon Forums III and IV, the Group Written Report (GWR) and the In-Class Rewrite (ICR). While each are output sources written in different genres, the unifying theme of probability and the shared experience of the Cups and Coins activity draws these four sources together.

Responding to the task to form a predictive two-clause present tense and future tense sentence in Forums III and IV, participants attempted to write hypotheses accordingly. However, they frequently formed sentences in grammatically erroneous ways or in ways not typically seen in university writing. In Forum IV (post-OTRE Sessions 3 and 4), participants’ use of those two structures improved. However, they continued to write outside the limits of appropriate use which had been set. Moreover, when participants successfully constructed a well-formed hypothetical premise, they often followed that appropriate use with errant use such as they had used in Forum III. This issue is more closely examined in Section 6.6.

6.4.1. Participants’ High Point assessments in Forums III, IV, the GWR and the ICR
In this section, assessments of token use specifically related to the construction of conditional statements and hypothetical premises in Forums III and IV (Frame 2) are presented via graphs. Examples of the participants’ written output is provided to help explain those assessments. The hypotheses were required to make a predictive statement about the first and second iteration of the Cups and Coins activity discussed in Chapter 3. I present the High Points (explained in Section 6.2) specifically and not an average of the assessment values in total. The reasons are that first, the data for isolated tokens, including those used in conditional statements and hypothetical premises was presented in Section 6.2. Second, the assessment of the skill level a given token used within a hypothetical premise relays little about the overall appropriacy of the attempted structure. Third, participants often made only one attempt at writing a hypothetical premise in each of the two forums and so there are very few or no other data to comment on. For these reasons, in most cases High Point and the average would either be identical nearly the same. In cases where there were as many as two attempts at expressing a hypothetical premise, what is deemed important here is whether the participant (or in the GWR, participants) was able to construct one appropriately formed hypothetical premise as modelled. For these reasons, the graphs which follow in this chapter each present a single data set, High Points of Token Use Appropriacy.
In Figure 6.20, the averages for the High Points arrived at through the combined assessments of the three raters is displayed. Although individual cases differed, this table shows that there were no cases, within any single output source, of multiple large decreases. Overall, the High Point rose after Forum III and was maintained afterwards through the three subsequent output sources. In this section, I present those individual cases. Later, I draw upon student-student talk during the OTRE sessions, participants’ responses to interview cues and each participant’s written output in Forum III and Forum IV, the GWR and the ICR.

![Figure 6.20: Average of assessment of appropriacy for nine participants in Forums III and IV and the GWR and the ICR](image)

**Frame 2**

In this section, I present the data for the written output from the two forum tasks, which involved the construction of hypothetical premises. Forum III was completed before the OTRE two sessions (Session 4 and 5) that involved conditional statements and hypothetical premises. Forum IV was completed after two OTRE sessions.

Below are results of the averaged assessments, illustrated by graphs and explained through examples of participants’ writing followed by explication of those examples. I reference individual participants’ graphs which are illustrative of trends common to two or more participants, and then provide examples of the texts which led to the assessments from which the graphs were generated.

In Section 6.4, I connect the relationship between the trends and the written output presented here and the lived experience of the participants who worked through the OTREs. In this way, I answer RQ2, regarding the extent to which OTREs guide participants to use
tokens more appropriately; and RQ3, regarding participants’ perceptions of their experiences completing OTREs.

**Forum III**
Prior to the first OTRE work involving conditional statements and hypothetical premises, participants engaged in forum discussion, discussed below. I have grouped the participants together in *groupings*, according to the trend between Forum III and IV. I present the graphs for both Forum III and IV in the discussion of Forum III.

**First Grouping**
In this first grouping, I look at the written output of participants, where the level of appropriacy of token use was assessed as maintained at the same level between Forum III and Forum IV.

The following forum posts were made in Forum III by P2 (see Figure 6.21), P5 and P6, who maintained an average of level-2. P5 wrote:

> The probability is 1/3. I have two reasons. First, every choice has same probability. Second a cup is correct but the other cups are incorrect when an cup is chosen. So I think the probability is 1/3. Probability is a specific limiting value when an trial is repeatedly done. Hence we can check the probability by repeating the trial.

![Frame 2: Forum III, IV](image)

*Figure 6.21: Average of assessment of appropriacy for P2 in Forums III and IV*
P5, reached a level-2 assessment of appropriacy through the use of a few of the tokens which are typically well-known, frequently used by first-year students at TUS, being widely taught in high school English lessons. However, he failed to construct any two-clause sentences, and, furthermore, neglected to observe the tense rules that had been modelled in the OTREs. There is the use of *when*, but it is not connected in any meaningful way to a predictive clause. Again, the use of *can*, rather than *will be able* is an example of the constructions which convey a sense of a future tense, which participants often created, rather than the forms modelled in the OTREs. P6 also failed to provide a correctly formed hypothetical premise during the discussion of the Cups and Coins enquiry into probability. P8 did not complete Forum III, but was assessed at level-2 in Forum IV.

**Second Grouping**

The second grouping is characterised by increased assessments of appropriacy, in the written output of P1, P4, P3 and P9 in Forum III. P1 (see Figure 6.22) wrote:

> I think it is 1/3. Because there are three way to choose the cups and there is one way to choose the correct cup. The possibility of choosing each of cups is 1/3.

![Frame 2: Forum III, IV](image)

Figure 6.22: Average of assessment of appropriacy for P1 in Forums III and IV

P4 (see Figure 6.23) wrote:

> [. . . ] if you can guess only one cup 500 times, the probability gradually approaches 1/3. My hypothesis are that the probability is always equal.
P4’s above attempt, with the token *if*, is more appropriate than P1’s and P2’s hypothesis writing attempts, but not assessed as level-3, *Skilled*, as the second clause fails to predict what will happen. Instead, it only comments on probability.

P3 (see Figure 6.24) had begun with an assessed appropriacy level of 1, *Unskilled*, with:

I think the probability of the choosing correct cup is 1/3. there is one ¥10,000 and three cup, and I don't know the situation of the money setted. As a result ,I think the probability is 1/3.

This output lacked both the form and function of a hypothetical premise as modelled in the exercises in OTRE Sessions 4 and 5.

Figure 6.23: Average of assessment of appropriacy for P4 in Forums III and IV
P9 achieved level-2 assessments in Forum III (see Figure 6.24). She wrote:

In my opinion, the possibility to choose each cup of them is 1/3 mathematically if I don’t consider favorite colors and so on. If I choose one cup three times, I will choose each cup. But if there are blue, white and pink cups, I believe I won’t choose the pink cup because I don’t like pink, so possibility won’t be 1/3.

This was one assessment on which raters needed to confer. First, P9 did not seem to be responding to the question in an entirely serious manner. The purposeful use of tokens was noted, but the two main tokens, *In my opinion* and *I believe* were used inappropriately. The irrelevant details also caused the line of argument to meander, and this further lowered the assessment.

P9’s performance here was uncharacteristic of her written output in earlier forums and chat rooms. I was unaware of the output in Forum III until after the end the course and so no feedback or response was given to P9 regarding her output in Forum III. However, subsequent to OTRE Sessions 3 and 4, P9 had learned and is able to put into practice the tokens and structures modelled for use in hypothetical premises.
Below in the next part of this section, I follow up on these results in the section on the second grouping of written output for Forum IV.

Forum IV
Subsequent to their completion of OTRE Session 4, participants next again engaged with hypothetical premises in Forum IV. Here they were again tasked to construct a hypothesis, this time employing what knowledge they had gained from the OTREs’ modelling of conditional statements and hypothetical premises. The graphs presenting the levels of appropriacy for these values are positioned above in the previous section, on Forum III.

First Grouping
Included in the first grouping, is the writing of participants characterised by an average maintenance of the level of appropriacy in token use. These included P2, P5 and P6 who each began with a level-2 in Forum III and gained the same assessment level in Forum IV.

Contrasting P5’s lack of fluency during the interview, was his expressiveness in the forum. P5 did not seem unmotivated, using several of the tokens with clear intentionality:

The probability is 1/3. I have two reasons. First, every choice has same probability. Second a cup is correct, but the other cups are incorrect when an cup is chosen. So, I think the probability is 1/3. Probability is a specific limiting value when an trial is repeatedly done. Hence, we can check the probability by repeating the trial.

However, he does not seem to have attempted to write a hypothesis. I will discuss this and
other such cases near the end of this section. He continued:

That's because the probability is 2/3 when I change my choice. When the cup chosen first is correct (the probability is 1/3), I cannot find correct cup, and when it isn't correct cup (the probability is 2/3), I have 100% chance to find it.

Here, P5 comes closer to writing a hypothesis, with three conditional statements, due to the present tense in the second clause. He does suggest an outcome, but one that presupposes a statistical result, rather than events from which a statistical generalisation can be drawn. P6 wrote inconsistently, but more successfully:

[... ] if I choose the correct cup from the beginning (the probability is 1/3), I definitely would end up choosing the wrong cup. Second, if I choose the wrong cup at the beginning (the probability is 2/3), then Mr. Gann is going to show me the wrong cup which I didn’t choose. And this means that if I change the cup this moment, I can definitely choose the correct cup. For this reason, we can see that the probability of ending on the correct cup is 2/3.

However, this is an explanation of the procedure and not a prediction of the outcome. This is an improvement in writing a hypothesis, with the two-clause structure, but again there is the same unorthodox way of suggesting the presence of a future tense through what I describe as conditional potentiality. Also, P6 shifted to a mix of past-perfect tense and past tense implied by would:

If we have chosen one cup without changing, the possibility of choosing the correct cup would be a 1/3.

Worth noting, in Chat II, discussing a different topic, P6 wrote, ‘If not enough people participates, then the class will be terrible.’ P6 used the two-clause form twice, in one sentence after the other. Here we see a correct construction for a hypothetical premise. P6 was clearly capable and yet did not apply that knowledge during Forum IV, the Group Report or In-Class Rewrite. I discuss the possible reasons for participants generally not using the OTRE-modelled present tense, future tense construction later in Section 6.5. In this grouping, although the High Point levels of token use appropriacy did not increase, it was not due to Unskilled performance. The reasons were that (1) Participants in Forum IV had not yet internalised appropriate token use to a level of appropriacy greater than the level (Average) they had achieved during Forum III. Another reason for these participants not having increased their level of appropriacy is that in Forum III, they were already using tokens at or
near the maximum for what the appropriacy scale accounted for, and in Forum IV, they repeated their High Point performance.

Second Grouping
The second grouping of participants increased their appropriacy of use in Forum IV from the level they had achieved in Forum III. The participants in this grouping are P1, P3, P4 and P9. In Forum IV, P1, who reached a High Point assessment of level-2 (Average) in Forum III, but level-4 (Exceptional) in Forum IV (see Figure 6.24), wrote:

If you choose the correct cup first, you won't choose the correct cup on the end. However, If you choose the false cup first, you will choose the correct cup on the end. The probability of choosing the false cup first is 2/3.

The two-clause, present-tense future-tense construction with the modelled use of the modal won’t and will earned a level-4 assessment.

P4, who, like P1 began with a level-2 assessment of appropriacy for token use in Forum III, reached a level-3 in Forum IV (see Figure 25) wrote:

When there is ¥10,000 under A cup, the probability you choose B cup is 1/3. In addition, you change to A cup. Because, the person who know there is ¥10,000 under A cup chooses wrong cup, C cup. The probability you choose C cup is same 1/3. As a result, the probability you choose correct cup is 1/3+1/3=2/3.

P4 followed that with:

I suppose that if you repeat this event many times, you will know the correct probability. The probability you choose the each cup is same.

Here, we see the attempt to express the hypothesis in a more appropriately structured two-clause sentence. However, again, we see use of the present tense in the second clause, when the future tense had been modelled in the OTREs. In the second attempt, P4 used the correct form: If x, (then) will (do) but it did not fulfil the meaningfully predictive function of a hypothesis and was therefore assessed as level-3 appropriacy on the basis that in the first post, metadiscourse was extended across several clauses and two sentences with the phrase, as a result.

P3, who had earned a level-1 assessment in Forum III, increased his appropriacy of use to level-2 in Forum IV (see Figure 6.26), writing:
If I select the incorrect cups, i can understand the place of the correct cup. that means if you choose incorrect cups, then you can choose correct cup? Here, *can* (present tense) means *have the capacity to do (subsequently).* Such flawed attempts at hypothetical premises often emerged from this data during analysis. Rather than concluding that participants had disregarded the instruction they had received in the OTREs, regarding future tense in the second clause, it seems that they often formed novel ways of intimating a future tense without using the modal *will,* which had been modelled in the OTREs.

P9 who began in Forum III with mixed assessments by raters, due to uneven performance, gained a mid-level-4 assessment in Forum IV (see Figure 6.27) subsequent to completing OTRE Sessions 3 and 4. P9 had formed a two-clause construction using *won’t be* in Forum III to indicate the future tense in the second clause. In Forum IV, she continued this use, writing:

If there are three cups which are named A,B,C and you choose A cup first and the answer is the other cups, the probability that money is in A cup will [be] 1/3 and the probability that the money is in B or C will [be] 2/3. So If you change your selection, the possibility that you can get the money will double compared with the beginning.

P9 received a mid-level 4 assessment based on the argumentative structure which extends across the paragraph, and what appears to be the intent to use the *will be* token, which she had used in Forum III but seems to have forgotten here, as she used only *will.*

To sum up Frame 3, findings indicate that students who began at average levels of appropriacy for constructing hypothetical premises were either able to maintain that level or, in an equal number of cases, were able to raise it one or two bands on the appropriacy scale.

In the following section, covering Frame 4, I present data that makes a stronger case for the efficacy of OTREs in guiding first-year science students to produce more appropriately constructed conditional statements and hypothetical premises.

**Frame 4**

In this section, I present the data for the two final output sources, the Group Written Report (GWR) and the In-Class Rewrite (ICR), both of which, like Forum III and Forum IV (in Frame 2) involved the construction of hypothetical premises. Frame 4 is unique among the four frames in that it includes two genres: a group written document produce asynchronously, out-of-class and a document produced individually, in-class, within a single class meeting.
The writings are paired, as the ICR was assigned as a rewrite-from-memory of the GWR. The purpose for this was dual. First, the ICR could identify cases in which a participant had contributed minimally or not at all to the group effort. Second, it could give some indication as to how internalised the language learning had been.

In the following section, I will present participants’ group and individual attempts at writing appropriately phrased and structured hypotheses. It was expected that, in the ICR, without the benefit of the Master MDM List and the support of other group members, some, if not most, participants’ levels of appropriacy might decrease. However, this was not always the case. In the following passages, I will present data showing how participants’ levels of appropriacy was maintained, increased or decreased. As I did with Frame 2, with Frame 4, I have grouped the participants together in groupings, according to the trend between the GWR and the ICR. I present the graphs for both the GWR and the ICR in Frame 4.

First Grouping
In this first grouping, I look at the written output of participants, where the High Point levels of appropriacy of token use were assessed by the raters as neither increasing, nor decreasing, from the time of the GWR to the ICR.

GWR
In the first grouping, P4 (see Figure 6.26) and P8, between their GWR and ICR, were judged to have expressed values with a flat trend for High Points at the level of Average.
Previously, in Forum IV, P4 had been assessed at level-3, Skilled; and P8’s output in Forum IV had been assessed as level-2, Average. In their GWRs, P4’s group, Group 2 achieved level-2, writing a competent enough description of the process for carrying out the experiment, followed by:

We think the probability of pulling a correct cup is one third. Because all the cups are the same, I think the probability of drawing the right cup is equal. We think the probability of drawing the correct cup is two-thirds. Because if the cup chosen first is not the correct cup, the cup you choose after changing is always the correct cup.

This attempt has the common problem of failing to separate the explanation of the novel and counter-intuitive aspect of the activity from the tasked formation of the predictive statement. Furthermore, rather than using a future tense phrase such as will happen, the group members proceeded directly to a claim, in the present tense, about what the probability is, an issue I take up later in the Discussion in Section 6.6.

In Group 3’s attempt, they wrote:

If we choose correct cup at the beginning, we cannot get the correct one because we must choose another cup which is wrong one. Second, if we choose wrong cup at the beginning, the game master will show us the wrong cup which we didn’t choose. Finally, we can get the correct one. Each probability of choosing one cup is equal, 1/3. Then, in
the first case, the probability of choosing correct one is $1/3 \times 0 = 0$ and in the second case, the probability is $2/3 \times 1 = 2/3$. Therefore, the total probability is $2/3$.

There is, as we have seen in earlier examples, as much of an emphasis on the procedure of the experiment as on the outcome. Also, there is the common, implied future tense that is suggested by the potential conditional *we cannot* and what I have termed as a *conditional imperative* with *must* in *must choose*.

In their concluding paragraph they continue using present tense phrases such as *it is* and *we can* writing:

> If we can change cups in the game, the probability of choosing correct cups is $2/3$. So, it is predictable that we can get 2000 yen ($100 \times 30 \times 2/3$) at the end of the game by mathematical thinking. Therefore, we should pay up to 1900 yen to the game master at the beginning of the game.

Taken one-by-one, these phrases merit appropriacy assessments of *Average*. However, there is a clearly purposeful attempt to distribute tokens in such a way as to construct a system of metadiscourse across the entire paragraph which earned the passage a level-3 assessment. What was considered to be holding the passage back from maximal appropriacy was the lack of phrases such as *will (happen)* or *is going to (do)*, and again, the common dependence on present tense constructions to suggest futurity, such as *it is predictable*.

**ICR**

In this next passage, I look at ICRs in the first grouping which followed after GWR to form a flat trend.

In P4’s ICR, the attempts at a hypothetical premise had no examples of exceptionally appropriate constructions. P4 wrote, ‘If he chooses cup before he change cup, he can not choose correct cup’. P4 continued to use the present tense. Moreover, in the second clause she informed the reader what the player cannot do, rather than what will happen.

P8 also expressed the common misunderstanding of the purpose of hypothesis, emphasising the procedure rather than prediction and then stated what the probability *will be* rather than what they predict *will happen* with:

> If player get a correct cup at first choice, he has to change cup after game master shows another wrong cup. For this reason, the probability will be 0 in this case.
As with previous examples provided by other participants, P8 does not seem to understand the difference between data and probability. P8’s hypothesis formation was assessed as equal to his group’s prior attempt at writing a hypothesis in their GWR.

In this grouping, although the High Point levels of token use appropriacy did not increase, it was not due to use that was assessed by raters as *Unskilled*. Participants either had not learned to use tokens at level of appropriacy greater than average or that learning had not yet been internalised to a point of automaticity, and they were unable to access the appropriate tokens of recall the OTRE modelled constructions.

**Second Grouping**

**GWR**

In the second grouping, P6 (see Figure 6.27) P7 and P9 achieved high level-3 and level-4 assessments in the ICR after their groups, Group 2 and Group 3 had achieved a level-2 in the earlier GWR (Grouping 1, GWR, above).

![Figure 6.27: Average of High Point assessment of appropriacy for P6 in the GWR and the ICR](image)

**ICR**

Here, I note the ICRs which followed the GWRs mentioned above to complete the trend of increasing appropriacy of token use.

In his ICR, P6 wrote:
When we try more and more times to see how the possibility will be, it will be more and more closer to the ideal possibility.

P6 also wrote:

If the challenger has chosen the correct cup from the beginning, the result will become miss. If the challenger has chosen the incorrect cup, the result will become hit.

While these three sentences do not form predictive statements exactly as modelled in the OTREs they do make use of the modal will. It is possible that the OTREs that presented the conclusion indicator, as a result may have been an influence in this case. This use of the word become in hypotheses was taught neither in the podcast series, nor the OTREs. The Japanese equivalent construction with the phrase ni narimasu, which means to become, is a likely factor.

P7 came close to reaching a near level-4 assessment, writing:

If we choose correct cup at the beginning, we will end up gaining wrong one. If 100 yen is under each cups, it is predictable that we can get 2000 yen (100*30*2/3) finally because of mathematical premise.

This attempt includes the hitherto unseen end up, which is a way of suggesting future events without constructing the future tense. It was not included in the Master MDM List. In this case, however, the use was redundant as it was preceded by will. That is to say, that although will end up is acceptable, it makes necessary the -ing in gaining where the simpler will gain would have been closer to the form modelled in the OTREs and the previously noted can (do), suggestive of potentiality and future action. P7’s use of will resulted in a somewhat higher assessment, but it should be noted that this was followed by the return to the formation of the somewhat inappropriate future tense with it is predictable that and we can get . . . finally.

P9 wrote to a better effect, ‘If we chose a doubt cups at first, we will be able to choose correct cup.’ Even so, P9’s attempt at a hypothetical premise in her ICR is, like earlier examples, a mix between a prediction and a description of conditions resulting from the procedure. I discuss this common characteristic in Section 6.5.

In both cases, despite the grammatical and idiomatic flaws, there is a two-clause predictive statement, using present tense in the first clause and future tense in the second clause; and that future tense is expressed through a modal will.
Third Grouping

GWRs
Here, I show that in cases where the level of appropriacy for token use fell in the subsequent ICRs.

The third grouping of participants included the three members of Group 1. P1 (see Figure 6.30), P2, P3 were cases characterised by two points.

Figure 6.28: Average of High Point assessment of appropriacy for P1 in the GWR and the ICR

First, these were cases in which the level rater assessment High Points of appropriacy of token use decreased in their ICRs. However, they all three fell from level-3 in the GWR to assessments of no lower than 2.5 in the ICRs. In their Group Report, Group 1 wrote the following:

If we hide a coin under a cup among three cups at random, then ask a tester chose one cup and tell him a cup which have no coin from two cups he didn’t choice, and repeat this as many as we can, the probability of ending on the correct cup become about two thirds.

Notwithstanding the present tense in the second clause, Group 1, overall, made an improvement in their GWR score. P2 and P3 had previously been assessed at level-2 appropriacy in Forum IV. It must also be pointed out that P1 had been assessed at level-4
appropriacy, due to his close-to-model-form hypothesis construction. However, his exceptional performance was not continued in the group effort.

**ICR**

In this final part of this section, I note that in the cases in which the levels at which the raters’ assessment of the appropriacy of token use fell, those levels did not decrease greatly.

In their ICRs, the individual members of Group 1 each decreased in their levels of appropriacy, but all three began in the GWR at level-3 (Skilled) and there were no drops to an Average level.

P1 wrote: ‘If you choose the false cup first, you will choose the correct cup on the end.’ The attempts at a hypothetical premise in P1’s rewrite had no examples of inappropriate constructions. However, while the construction is that of a hypothesis, the function was assessed as more of a conditional statement. A hypothesis should be a statement that provides a means of testing the reliability of a claim. This attempt presupposes, in the first clause, that the chosen cup is one of the two not containing the coin.

P2 wrote: ‘If I am changing the cup, I will lose the money only when I choose the right cup [at the beginning].’ This attempt has a more complex structure, with *will* and *only when* as links between three segments of the sentence. Also, while the construction is that of a hypothesis, the function is that of a conditional statement. It is a declaration of an obvious fact, i.e., conditions, and not a prediction.

Lastly, P3 wrote: ‘If we choose the correct cups then, we counted the number of hit points.’ The participant did not write a hypothetical premise, but description of the procedure for collecting data after the experiment has been completed. The second clause which should be future tense, is past tense. At first glance, this token use appears less skilled than if the participant had maintained present tense throughout. It may be asked, however, if the use of the past tense is not more than just an error, but a misstep along the path of interlanguage (Lantolf and Thorne, 2007) to correct use. It also likely the case that the shift into past tense occurred as P3 shifted, mid-sentence from attempting to construct a hypothetical statement to recounting what had happened during the Cups and Coins activity. Despite these flaws, this was an improvement over what had been done in Forum IV.

In this third and final grouping of participants’ written output, we see cases in which the token use in the writing of the individual group members decreased in its High Point level of appropriacy: as assessed by the three raters. The possible reasons for this are covered in the in Section 6.6.
6.4.2. Summary of appropriacy levels in conditional statements and hypothetical statements

To conclude, the written output of the participants in Forum IV shows an improvement in appropriacy or token use upon what was submitted in Frame 2, Forum III. After the hypothetical premise-focussed OTRE Sessions 4 and 5, skilled use of the two-clause If-then present tense-future tense construction developed from absent to nascent to in-development. In a few cases, in the GWR and even in the ICRs, participants’ interlanguage developed to the point of constructing competent hypothetical premises. In view of these improvements, it may be strongly argued that the OTREs played an important role in fostering and accelerating this development. In this way, the data answers RQ2. In the following sections, I present responses from transcripts of one-on-one semi-structured interviews and utterances made during OTRE sessions. By triangulating these data with the data presented earlier in this chapter, I further respond to RQ2 and RQ3 by adding support to my claim that OTREs guided participants to use tokens at higher levels of appropriacy.

6.5. Participant interview responses

As noted in Chapter 5, in the week following the In-Class Rewrites, I conducted one-on-one semi-structured interviews with the nine participants. In Chapter 5, I discussed the first half of the interviews where the purpose was to learn about participants’ backgrounds in the use of the language of argument and to find out the extent to which the participants felt the OTREs had influenced them to use tokens with greater frequency and variety. Using the mixed-methods research technique of triangulation (see Section 3.3.3) I also drew upon excerpts from the transcripts of participants’ talk during OTRE sessions, to further the claim of the OTREs influence and thereby answer RQ1.

As noted in Section 3.3.3, triangulation of quantitative data and qualitative data can make points salient for the researcher which either method of analysis on its own might not (Boeije, et al., 2013). Also, as Evans et al. (2013) notes, in a small-scale study, triangulation involving qualitatively derived data can help to determine or reject observations of emergent themes by comparison with simultaneously derived variable-oriented, quantitative data. Correspondingly, in this section, I draw upon semi-structured interview responses to add weight to the claim that the OTREs were an important part of participants’ interlanguage progress, thereby furthering my response to RQ2 and RQ3.

Without the use of triangulation, the case for a causative relationship between OTRE sessions and gains in appropriacy of token use could not be as strongly supported. By
assessing the participants’ semi-structured interview responses as complementary or contradictory. The participants’ responses also provide a window into their otherwise hidden experience, as they completed the OTREs and later drew upon the language learning they had gained there in the forums, and chat rooms. Drawing upon participants’ insightful remarks, I am thereby aided in answering RQ3. These data add to the argument that OTREs influenced participants to use tokens with greater frequency and support the claim that OTREs played a major role in guiding participants’ development of their skills at using the tokens they had studied with greater appropriacy; and that the interactive technological and socially interactive elements in their OTRE experiences were important in their language skills development.

6.5.1. Participants’ perceptions of the general effectiveness of OTREs
Both P1 and P2 confirmed that that the OTREs had aided them in using tokens more frequently and with greater appropriacy. P1 said briefly, ‘Yes, I think, use more and more’ and P2 said:

P2: Yes, I used.
Gann: Where did you use it?
P2: We discussed online. So, I used, agree[ment] (and other categories) with the members in the forum and chatroom.
Gann: Anywhere else? We had the forum and chatroom and--
P2: The final.
Gann: Yes, the final report.
P2: Yes, I used, of course. I use.

Each OTRE begins with a sample text with the arrow of cause and effect the opposite of that found in the sample text (see Section 3.6.7). When asked how much of their learning was a result of completing the text reconstruction as opposed to the initial reading of the sample text, participants spoke expressively in favor of the OTREs. P3 provided the following analogy of recognizing the gap:

The difference between reading and doing . . . By doing WebSequiturs I learned the language of argument and practically so if I study by reading texts, I will not use answer. [Trans.] If you just read it, you probably won't actually use it as you read it, so you can't study efficiently. For example, if you just read the score of the piano, you can't play it, but if you actually play it while looking at the score, it's easy to understand what's wrong and it's easy to improve. It was easy to understand.
When I asked him for some examples of what he had learned he responded thus:

First, the theme of discuss, and second the reasons for this discussion and the finally conclusion.

Gann: Was there one more that we studied that was very important with the—

P3: Conditional and hypothetical statement.

P3 confided that he was still struggling with the structure for conditional statements, saying, ‘I cannot use language of this argument effectively, but I can use these language a little, so . . . uhh.’

P6 explained in detail:

Well, if I like only read only once or so then I’m not totally still 100% understand, then when like we discuss together with group members, I can listen to what they’re thinking and sometimes they can tell me this part means maybe like this and I can say, oh, now I understand. And these kinds of things will happen.

P4 spoke of how the OTRE sessions had improved her reading speed:

I can read the sentence much faster than before because my group members is--have so high skills for English so I can learn from them, so and I can read English sentence and I can learn grammar.

P5 indicated that he felt the OTREs had helped him:

If I didn't do OTREs, the term probably wouldn't have stuck with me as much, so I would hardly be able to use it when it comes to actually using the report [recalling items from the Master MDM List] in the chat room.

He said this, although his was one of the more sporadic performances in the forums and chat rooms. He failed to participate in Forum II and V and was absent on the day of the ICR. However, when he participated, the word count and token count in his forum entries and chat performances were average or above average as were his levels of appropriacy of token use.

P5 continued:

[Trans.] Can I know a few, or some words, including how to use them, I wonder? I can read . . . I think, but [when] I saw the rest of the words, I [would] get stuck when I actually read [them]. I saw them, but it's hard to recall them, I can't remember them.

Yet the data indicates that P5 was able to recall the tokens and in some cases his performance exceeded others. He later described himself as lazy student: ‘Tadaina seito nande’. His
contributions were uneven and inconsistent might be more a more accurate description. An important point is that triangulation of data shows that missed forum and chat room opportunities did not decrease P5’s frequency of use, nor his appropriacy of use. Moreover, his response further supports the claim that OTREs played an important role in influencing P5 to write argumentatively with greater frequency of token use and guided him to use tokens more appropriately. In this way, participants’ reports provided during the semi-structured interviews of their experiences during the OTREs sessions and helped to answer RQ2 and RQ3. Their responses indicate that participants felt that the OTREs had increased their knowledge of how to use the tokens appropriately; and showed that participants valued the experience of actively co-constructing the OTRE text passages, rather than passively reading them.

6.5.2. Efficacy of OTREs over printed handouts

Two threads that I attempted to separate through dialogue with participants were: (1) the guidance of the OTREs and (2) that of traditional handout such as students have often received in traditional classes. Participants’ responses helped me to further discover the finer points of the participants’ experiences and thereby address RQ3.

When queried about this P2 said, ‘I learned a lot by exercises. I can use my brain more hareshiku (pleasantly) in exercise’. I asked which is a better way of learning: to work through an OTRE session or to receive a handout, a photocopy with examples of correct use of tokens. P2 responded candidly, ‘If you just give me paper, I . . . I might not read it.’

I questioned P1 about his mental perception of the image of the tokens:

Gann: When you were trying to do the chat with other people--when you were trying to access a word, in your mind, (gesturing) did you have an image of the forum in your mind or did you have a memory of doing the OTREs with your partners.
P1: Mostly pdf files.
Gann: Mostly?
P1: Un. (Yes).
Gann: Did you sometimes think about the WebSequiturs?
P1: Yes.
Gann: How important were the WebSequiturs compared to the text, itself.
P1: I think the WebSequiturs is more important.
Gann: . . . Why?
P1: Because, if we use WebSequiturs we can think [about] the questions by ourselves
[with our own minds], but if we use textbook we often only see and don’t think, so WebSequiturs is more important.

This was an important answer, because it made me later question the presupposition that the object of a participant’s mental imaging at the time of the attempt to recall the tokens was necessarily what the participant would consider to be the most important. Instead, the pdf, i.e. the Master MDM List was the object she had most recently engaged with, so for her the list had primacy over the OTREs. However, the list would not have been so easily recalled, had she only worked passively from a textbook. Thus, the OTREs had primacy over traditional textbook styled lessons. To this point, P4 indicated that, although she was more directly bound to the Master MDM List, the OTREs had been an important part of the total experience:

Gann: As you try to think of the words. Are you thinking about the pdf or are you thinking about the WebSequiturs? [I repeat the question in Japanese.]:

P4: Pdf.
Gann: Is it difficult to think of the pdf or memorise the pdf?
P4: When I did the forum I saw (looked at) the pdf. So, if I see it, I can remember this word, but I cannot remember soon without seeing the pdf.
Gann: Some people learn easily by memorizing a paper and other people learn by remembering an experience. Do you have an opinion about which kind of person you are?

P4: Mmm. I can’t memorise the pdf. I did experience . . . Mmmm
Gann: What kind of experience?
P4: Forum and of the WebSequiturs. So, my class members send their message. The message, conclusion, the key word phrase. So, I can learn to see it [by looking at] my class members’ comment forum.

Although P4’s response is mixed, it is noteworthy that she notes the OTRE session, which seems to have played a supporting role to her use of the Master MDM List, which, without the OTRE experience is merely a collection of words.

 Conversely, P7 indicated that the OTREs had played a direct part in his learning:

At first, I often mistake the reconstruction, but [Trans.] As I repeated exercises, I became able to be aware of the connection of sentences, and I think that I gradually gained the power to express my opinion.

Gann: What if I didn’t have the WebSequiturs and I just gave you the same text here on
a piece of paper and I said study these and then try to write the same way? How would
the result be in terms of your learning?

P7: [Trans.] In that case, I would just read the sentences without thinking about
anything. But because there was a rearrangement [or reconstruction] of the
WebSequiturs, I came to understand the connection of the sentences, so that I wonder
if I could cultivate the ability to connect and say that in the [correct] way [if I did not
do the OTREs].

P8 said similarly: [Trans.] ‘If I hadn't done WebSequiturs, I think that word wouldn't be well-
established in me.’

P9 responded likewise:

Only handout maybe we will we want to talk about . . . [Trans.] but by using
computer, computer is interesting for us to see so we can more easily to be interesting
in English more. Statements. So more deep study we could do.

Here, it is noteworthy that P8 spoke of the token being well-established. He uses the Japanese
word, teichaku, which can also be translated as fixed in the sense of a thing being securely
fastened and it is often used in the sense of a something taking root, in a positive sense,
through practice or habit. Thus, P8’s comment suggests an understanding of OTREs as a path
towards internalisation (Lantolf and Thorne, 2007) (see Section 2.7.5).

The participants’ responses in this section further opened up their experience of
OTREs and address RQ3. The overall consensus here seems to be that the OTREs were an
important component in the series of steps leading to participants’ developing skills, even if it
was not always what the participants directly drew upon.

6.5.3. Participants’ perception of technological and social constructivist advantages
of OTREs

In this section, participants’ responses to cues set in the semi-structured interviews completes
the discovery of the participants’ perceptions of their experiences of the OTREs. Participants’
responses add to the previous section’s observations by discussing the social constructivist
dimension of the OTRE experience and allows me to address RQ3 to the extent allowed by
the limits of this study.

In addition to the two threads in the previous section, two other threads need to be
separated. The first of these is the attributes conferred by the application itself, that is the
linear progression created as the text is reconstructed from beginning to end via cues of
logical flow. The second thread involves the benefits derived from working with peers within a zone of proximal development (ZPD) (Vygotsky, 1978) discussed in Section 2.7.5. The social constructivist element is also an important part of the OTREs experience.

I asked participants to compare their past experience of working individually with their experience of completing the OTREs in triads.

P3: If I do WebSequiturs alone, I can't share my thoughts, and when I run into a difficult problem, I get stuck.
Gann: Sorry, what?
P3: If I do WebSequiturs by myself, I cannot share my opinion with others and if I see difficult question I cannot others to . . . I cannot call
Gann: ask . . . call
P3: [Long gap]
Gann: You can say it in Japanese.
P3: [Trans] When you come up to a difficult problem, even if you don't know what to say or what to do, you can talk with the people in the group to hear the opinions of others and see the problem from a different angle, so you can solve the problem. Therefore, I think the purpose of this WebSequiturs is to discuss with the group and find one answer because it will be easier to solve the problem.

P1 also commented on the importance of the social aspect of the OTRE work:

Gann: And now imagine I have you working in groups but without the WebSequiturs kind of exercise and the building up of the paragraph. How important is the social aspect and how important was the WebSequiturs application’s format?
P1: When we discuss about the questions with my group members we cooperate to solve the questions—so each of my group members have their own ideas, so we could have a good time.

Although P4 had indicated that she connected the tokens more to the pdf of the Master MDM List, she nevertheless felt that the social component of the OTRE sessions was important.

P4: I like doing the WebSequiturs because we can discuss with group members. Other members give advice to me and we can discuss, we can do critical thinking and so it’s very good . . . I can read the sentence much faster than before because my group members is--have so high skills for English so I can learn from them, so and I can read English sentence and I can learn grammar.
Gann: You said a moment ago your other group members are higher than you. Your skills are quite good. Could you compare the importance of process of using the software and the social aspect? If you had to tell me how important by percentage?

P4: I think the social situation is very important. About 70%

Gann: 70%. 70-30?

P4: Yes, because I had to communicate with my group member, so if I use the computer alone, I can learn more reading skills and grammar skills, but I can’t— I want to learn speaking skills and communication—priority communication. P5 also estimated the ration of importance at 70% for social interaction and 30% for the affordances of the application.

In the following excerpt, P6 speaks on how, in at least one instance, learning was connected to some random but memorable event that occurred in a social context during an OTRE session. During his interview, P6 related a learning event in which he had been working in his chat room trying to remember a token which he was unable, at first, to mentally access. He then had remembered a moment during which his group had been working on a WebSequitur exercise. He remembered that I had noticed that another student several rows away was playing a videogame instead of doing the work and that I was mildly remonstrating that student, asking him questions about his class work:

P6: After that thing happened, I saw the WebSequitur. I do still remember the WebSequitur. Yeah, the trigger was what happened with [that student].

Gann: Well, that’s the way language learning works sometimes. As a Japanese language learner, I remember Japanese words and phrases that I remember because of the social circumstance. I don’t remember them from a page. They’re anchored to some kind of event.

P6: Yeah, well maybe this time they were anchored to what happened.

In other words, an event that entered his awareness, while he was actively engaged in a language learning activity helped him to mentally access the token, which, he indicated, would likely have otherwise remained out of reach at that time. As in Lantolf and Thorne (2007), social exchanges between triad members might also be associated with learning moments experienced during OTRE sessions; and such lived experiences could aid in the recall of the metadiscourse markers presented in the OTREs.

P8 responded similarly:
[Trans.] In group work, you can exchange opinions when sorting, so when you do not understand yourself, other people's thoughts come out, so I think that is very important.

P7 could not seem to separate the threads of WebSequiturs and the idea of working in groups, but said with certainty: ‘I think that my English skills improved when I used a computer that was close to me rather than alone’.

A discussion with P9 about the likelihood of being able to use the tokens we had studied in future contexts neatly distinguished the social and technological threads of the OTRE experience and also turned back to the threads of paper-based and computerised ways of teaching MDM use.

P9: Today we have a speaking test. In speaking test, me and other two students used the things we learned in critical thinking or WebSequiturs and indicator word, so, but in other class English class, without ishiki.
Gann: Willpower?
P9: Without ishiki or will we can’t use.
Gann: What kind of will do you need to use [indicator words]?
P9: To make sentences we must have will to make good sentences [Trans.]
In order to write a dissertation you will have to make a good sentence, and at that time, I think that you have to think about the structures of the sentences you learned.
Gann: How much of that learning then was a result of doing the exercises and how much was a result of reading the example texts
P9: Only handout maybe we will we want to talk about more [Trans.] but by using computer, computer is interesting for us to see so we are easy to be interesting in English more. So more deep study we could do.
Gann: What is it about the computer that gives you more motivation?
P9: [Trans.] Now, for example, smart phone or computer and many tools are coming to be, so students are interested in computers, so . . .
Gann: What is it about the WebSequiturs application that you liked or was motivating?
P9: If Mr. Gann gave us a handout each other [separately/individually] . . . [P9’s speech trailed off here but the implication was that she was comparing an inferior situation to the following one], but with one computer display we feel that we are doing with three students, so this is a good point.
Gann: Okay, so I could just take one handout and put it on the screen and have three members looking at the screen. Right?

P9: In computer, [Trans.] a button will bring up a new question, or it's usually convenient and convenient, so I think the computer is better in that respect.

Again, this expresses the view, that the OTREs weave both technological and social constructivist advantages together.

P2 alone says that he had an less than satisfactory experience, recalling that he did the exercises while the other two members watched, but agrees that the two other members may have benefitted from working with him.

Gann: So, the next question is about how much of your learning was a result of working in a social situation? I mean I could give the WebSequiturs individually and say, okay everybody, do your own work. Don’t look at other people’s screens. But I put you into groups. How did that affect your learning experience?

P2: I did almost all the questions.

Gann: You were the most talkative of the group?

P2: So, I almost individual.

Gann and P2 [laugh].

Gann: So, for you it was almost the same as individual. Do you think it was helpful to your other two group members to work with you?

P2: I tried to get [them] to know my idea and they understood—they seemed to understand. So, [that was] good.

Although there were different opinions on the matter, the general consensus was that the use of the OTREs had been an important factor in the educational experience. The responses of participants opens their experience for further discovery. They respond to RQ3 and inform us that the utility of the technological aspect of their OTRE experience was acknowledged by the participants, who understood, as shown in Section 6.4.1, the purpose of the linear construction feature of the WebSequitur application. However, what the participants reported most valuing was the social constructivist aspect of the OTREs. The OTRE was the virtual learning environment (VLE) in which the collaborative learning took place. Participants expressed that they not only found working with their classmates motivating; they reported that they valued the advantages of working within their and their peers’ ZPD. The linear reconstruction of the text was a feature of that environment.
6.6. Participants’ talk during OTREs

In this section, I present excerpts from the transcript of the OTRE sessions. I present the findings of my thematic analysis of that talk. This analysis is directed towards triangulation of data presented in Chapters 5, to further respond to RQ1 on changes in the frequencies in token use; data presented in Chapter 6, to further respond to RQ2 on changes in levels of appropriacy of token use; and qualitative data presented in both Chapter 5 and Chapter 6, in the form of participants’ written output as well as interview responses, in order to answer RQ3, regarding participants’ impressions of their experience. This third area involves discovering how motivating the exercises were and pulling apart threads of the OTRE experience to understand in what ways the OTREs were motivating, or demotivating.

Much of the transcript consists of participants reading the exercise passage, reading the multiple-choice selections and short one to three-word utterances, often no more than the letters ‘A’, ‘B’, ‘C’ and ‘D’. Much of this talk was already presented in the previous two chapters, where it was relevant and poignant, as well as earlier in this chapter. The talk that remains is repetitive and, on the face of it, few utterances, on their own, add further insight into the language learning which occurred. For that reason, following the presentation of participant-participant talk, I present a quantitative accounting of the whole of the transcript. This quantification includes the four areas of participants’ discourse which indicate metalinguistic awareness, or awareness of tokens and structures creating a system of metadiscourse. These include premise and conclusions indicators and references to the presence of premises and conclusions, and conditional statements and hypothetical premises and the presence of conditional statements and hypothetical premises. In addition, utterances indicating references to grammar and punctuation are also presented. The presence of such utterances is interpreted as suggesting either flaws in the design of the exercise, which allowed for the usefulness of grammatical or orthographic cues in completing an exercise; or a lack of understanding on the part of the participants regarding the purpose and aims of the text reconstruction activity.

Transcript excerpts and commentary

In the following passages I present, group by group, examples of participants’ talk during OTRE sessions, the forum chat room and Group Report writing. I interpret each group’s case through participants’ comments and responses made during one-on-one semi-structured interviews. Lastly, I will look at each participant’s In-Class Rewrite and look for signs of consistency or contrast (Hammond and Wiriyapinit, 2005).
During the first of these two sessions (OTRE Sessions 4 and 5) there was no discussion at all by any group of the structures of conditional statements and hypothetical premises; and only minimal discussion in the second. In the following examples, it is clear that varying levels of metalinguistic discussion were occurring. There is little or no in-depth discussion. Perhaps the participants tended to work under an assumption of shared knowledge. Perhaps non-verbal communication played a role, although personal observation revealed nothing more than occasional pointing at the screen and mouse cursor movement. As will be shown below, the most explicit and extended discussion of the matter came from Group 3.

**Group 1**

During their OTRE sessions, Group 1, demonstrated that they were aware of the focus on conditional statements which I had tasked them to apply toward the OTREs, and in some cases they were aware of the utility of understanding the form of conditional statements in recognizing their presence. Below is a typical exchange:

P3: Conditional . . .
P1: All check?
P2: All check . . . All check dekiruno (Can I do (click on) it?)

The phrase, *all check* was used in more than one group and seems to be the participant controlling the mouse asking the other two members if they agree on the selection he or she is hovering the cursor over.

P1: Wakannai. *I don’t know.*
P3: Sakino bunshou minagara janai to omou. *When you look at it and the sentence that came before, I don’t think so.*
P2: ‘How many?’ ‘How many?’ Let’s thinking! Oh, kore douyu imi? *What does this mean?*
P2: Conditional statement.

**Group 1**

During the OTRE Session 5 that followed, Group 1 had the following short exchange:

P1: Conditional.
P2: Conditional.
P3: Jouken mitaina. *Like a conditional.*

The word *conditional* was not written in the instructions to the OTRE itself. The participants
learned the word either from the podcast or from my minimal in-class instruction. The participants were applying what they had learned and were searching the multiple-choice selections for a conditional statement. Thus, the participants of Group 1 demonstrated that they were aware of the focus on conditional statements which I had tasked them to apply toward the OTREs, and in some cases they were aware of the utility of understanding the form of conditional statements in recognizing their presence.

**Group 2**
During their OTRE sessions, Group 2 discussed conditional statements as they progressed through the OTREs. A typical example is: P4: Kore ‘conditional’. *This (is) conditional.* No reason was given for this claim and none was asked by other members; and this was typical among the three groups. Perhaps, once pointed out, the veracity of the identification seemed obvious and did not require discussion. This demonstrated that they were aware of the focus on conditional statements which I had tasked them to apply toward the OTREs, and in some cases they were aware of the utility of understanding the form of conditional statements in recognizing their presence. Still, they were not successful in producing the appropriate construction.

**Group 3**
I follow up the preceding section, with its examples and content with quantitative data, also gathered from the transcripts of OTRE sessions, which completes the analysis of the participants’ talk and adds to the argument for a causal relationship between engagement with the OTREs and increases in appropriacy of token use. Following Bryman (1992), cited in Flick (2014) commonalities can be drawn out of qualitative data with the addition of quantitative findings. Therefore, I also applied a quantitative approach, further facilitating the triangulation of data related to utterances and data representing subsequent levels of appropriacy of token use. Doing so enhanced my response to RQ2, as I could better distinguish utterances which originated from metadiscoursal awareness from utterances which may have been incidental.

In order to determine if there was an association between relevant utterances made during the OTRE sessions and the level of a participant’s subsequent written output, I compared the number of utterances relating to premises and conclusions made by each participant in the first three OTRE sessions with the levels of appropriacy of tokens used as premise and conclusion indicators in Forum V. I repeated this process for the fourth and fifth
OTRE sessions, comparing the number of utterances related to conditional statements and hypothetical premises made by each participant to the levels of appropriacy of use for tokens used as conditional statements and hypothetical premises indicators in Forum IV, the forum following the OTREs focusing on conditional statements and hypothetical premises.

I chose to use the data from Forums V and IV) and not other sources because both followed immediately after the two sets of OTREs; as forums, both were asynchronous; in both cases, the Master MDM List was accessible; and except for P5 in Forum V, and P7 in Forum IV they both received contributions from all participants.

In the table below (see Table 6.2), I present the record of relevant utterances of each participant. I separated the count of participant talk into two main parts: OTRE Sessions 1-3, which focused on premise and conclusion indicators, and Sessions 4-5, which focused on conditional statements and hypothetical premises. The upper and lower portion are each divided left to right into columns denoting Groups 1-3, and are subdivided by participant, 1-9.

The next two columns represent utterances regarding grammatical or orthographic details, elements of the OTRE text that I did not intend participants to utilise as cues. The next column which follows presents instances in which a participant vocalised a premise and/or premise indicator, which they identified in the OTRE sample text, or anticipated in the reconstruction text. In the final column are cases in which a participant uttered the term for an indicator such as premise, reason, conclusion, or Japanese equivalents. Utterances in which a participant was simply reading the sample text or one of the multiple-choice selections were not counted, as this would give a falsely elevated indication of utterances specifically noting tokens. The last two columns repeated the categorisation of the previous two, but were concerned with the tokens used in conditional statements and hypothetical premises, e.g., if . . . then, unless, will be/do; and the terms conditional, hypothetical and Japanese equivalents. Again, cases in which a participant was simply reading the sample text or one of the multiple-choice selections were not counted.
Next, I separated the participants into groupings of similar numbers of utterances relevant to argument indicators among participants in the OTRE Sessions 1-3 and then in 4-5. The purpose of doing this was to search for associations between numbers of utterances and frequency and appropriacy of token use.

Previously, I had separated the counts for grammatical and orthographic comments, and for tokens and references to tokens as it seemed plausible that these distinctions might become relevant to one case or another. From this point, however, maintaining these distinctions needlessly complicated viewing the tables and discussing the data. For this reason, in Table 6.3, I combined the utterance counts for grammatical and orthographic comments, as these were not relevant to argument indicators. I also combined the utterance counts for argument indicators and references to argument indicators, doing this separately for premise and conclusion indicators and for hypothetical premises. Compressing six columns (columns 4 and 5, columns 6 and 7, and columns 8 and 9) into three columns (see Table 6.3) made viewing the data and discussing the data less burdensome.

Table 6.3: Utterance counts of participants by category

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Group</th>
<th>Participant</th>
<th>Grammatical</th>
<th>Orthographic</th>
<th>Premise/Conclusion</th>
<th>Terms/References</th>
<th>Hypotheses</th>
<th>Terms/References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

263
Table 6.4: Participants’ utterance counts grouped (categories condensed)

I ranked the groupings top to bottom by number of utterances relevant to premises and conclusions in the upper half of the table; and conditional statements and hypothetical premises in the lower half of the table.

In Sessions 1-3, I have identified groupings, as follows. Because of the large number of relevant utterances, I have isolated P6 as a grouping of one. Next, I have grouped P1, P2, P3 and P9. I have grouped P4 and P7 together. P4 only made about half the utterances of the others and all of P4’s utterances relate to terms referring to premise and conclusion indicators rather than to the indicators (tokens) themselves. I note that P8 made no utterances relating to either tokens or terms, and only a single utterance referring to an orthographic element. In addition, within this grouping of five participants, P1, P2 and P7 also spoke about orthographic (P7) and orthographic and grammatical (P1, P2 and P7) elements, whereas the others did not.

As explained in Section 3.6.7, I designed the OTREs so that grammatical and orthographic elements could not be used to select the correct multiple-choice selection as readily as the tokens toward which I intended to direct participants’ attention. Hence, the presence of many utterances in these columns cannot be viewed as simply negative or positive. On one hand, comments about grammatical or orthographic details, especially in the absence of utterances about argument indicators suggest that a participant did not understand the purpose of the exercise or did not notice the token in the text. On the other hand, such comments were evidence that a participant was actively engaging with the text passage. Moreover, such comments, when made in addition to comments more directly relevant to argument indicators, suggest that the participant was using what he or she viewed as all available information at hand.

In Sessions 4-5, I isolated P6, in a grouping of his own, with his ten utterances about terms referring to conditional statements and hypothetical premises and three, about premises

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Groupings</th>
<th>Participant(s)</th>
<th>Grammatical and Orthographic</th>
<th>Premise/Conclusion and Terms</th>
<th>Hypotheses and Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 3</td>
<td>1</td>
<td>1, 2, 3, 7, 9</td>
<td>0</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4 - 5</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2, 3, 5, 8, 9</td>
<td>4</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>10</td>
</tr>
</tbody>
</table>
and conclusion indicators. Note that P1 also made ten utterances and nearly the same number of orthographically related comments. However, whereas P6’s ten comments were concerned with terms relating to conditional statements and hypothetical premises, P1, on the other hand, made no such references, but continued to note terms and references to premises and conclusions.

With the similarly performing participants grouped and ranked, each grouping’s average values can be compared with each member’s levels appropriacy of use in Forum V (premises and conclusions) and Forum IV (conditional statements and hypothetical conclusions). To make this comparison, a common scale was devised. Beginning the utterances, I chose the terms Low (L), Lower Medium (LM), Upper Medium (UM), and High (H). For example, I differentiated between P4 and P7, both in the Medium ranking for relevant utterance count, by also comparing their comments about grammatical and orthographic elements. P7, who had made additional comments regarding grammatical and orthographic elements was ranked as Upper Medium, while P4 who had made no such comments was placed in the Lower Middle.

In order to apply this to the number of utterances, I noted the highest number of utterances by a single participant in the OTRE Sessions 1-3 (relating to premise and conclusions) and 4-5 (relating to conditional statements and hypothetical premise): 9 and 10 respectively. I then divided these utterance counts into thirds, for OTRE Sessions 1-3, as 0-3 as L, 4-6 as M, and 7-9 as H. I used the same division of the utterance counts for OTRE Sessions 4-5, but set H as 7-10. I regarded P6’s utterance count of 34 in OTRE Sessions 1-3 as an outlier and grouped him separately as H, ranked above the others, also ranked H, but one row lower.

Next, for the purpose of corresponding the utterance count scale of L, M, H to the levels of appropriacy of token use in Forums V (premise and conclusion indicators) and IV (conditional statements and hypothetical premises), I equated Unskilled with L, Average and Skilled with M and Exceptional as H. I used the High Point as the primary value to rank participants and place them into groupings; and used the Averaged level of appropriacy of token use for fine tuning of each rank. Further to this, I included the frequency of use of explicitly connective indicators compared with implicitly connective indicators as it was the use of the former which was emphasised in the OTREs. By using this scale to rank the appropriacy of token use on the left and relevant utterance count on the right side of Table 6.4, I am able to present the two data sets of this small-scale study together to determine the degree of association between active engagement and communication with group members.
during OTRE sessions and subsequent performance in written argument, in this case, in asynchronous forum-based discussions.

In Table 6.5, above, we can see cases of consistency between utterance counts and subsequent levels of appropriacy for token use; and cases of lack of consistency and moreover, contrast. I will now examine the case of each participant and discuss the degree of consistency, starting with the groupings designated as High.

Beginning with P6, we see that although he was enthusiastic about communicating with his group members, he used the terms referring to argument indicators in a meaningful way, identifying and referring to premises and conclusions, their function and placement (see Section 6.5). During the interview, P6 also demonstrated that he clearly understood the aim of the activity and intent behind the lesson. Yet, in Forum V, there was a large contrast between P6’s utterance count and level of appropriacy of token use. It bears noting that P6’s performance in Forum V was not representative of his performance in other forums and other written output sources.

In the second grouping of participants (P1, P2, P3 and P9), there was a large contrast (Denscombe, 2014) between P1’s High number of utterances and Medium level of appropriacy of token use in terms of his High Point and Medium Averaged level. P2, P3 and P9 however showed varying degrees of consistency (Denscombe, 2014). P2 and P3 expressed consistency falling more closely within the Medium range in their utterance counts and within
the Medium rating for appropriacy of token use, differing only in P2’s more frequent use of explicitly connective indicators, a factor, which, as mentioned above, I used for fine tuning. P9’s Medium utterance count was matched with a High Point of 4, in comparison with P2’s and P3’s Medium High Points of 3 and 2, respectively. P7 also expressed a strong consistency, differing from P2 only in terms of a lower frequency of explicitly connective premise and conclusion indicators.

P4’s case is contrastive and the inverse of P1’s, with a Low utterance count matched with a High rating of 4 for her High Point and a high frequency of use explicitly connective argument indicators.

Looking at the relationship between the number of utterances made by participants during OTRE sessions focused on explicitly connective premise and conclusion indicators, there is a strong consistency between the two. Next, I examine this relationship in respect to conditional statements and hypothetical premises.

Before noting any specific participant, it can be seen that, whereas in the analysis of utterances related to premises and conclusions, in which the participants grouped in the High and Upper Medium ranking; in the analysis of the conditional statements and hypothetical premises, the participants grouped in the Lower Medium and Low utterance counts. This suggests that participants found the two-part if . . . then structure more difficult to identify and/or refer to than premise and conclusion indicators; and this notion is supported by the higher number of participants in the Low ranking for High Points in the levels of appropriacy of use of tokens.

Beginning with P6, he again made the largest number of utterances, 10; and he matched his utterance count with a rating of 4 for his High Point for level of appropriacy of token use.

P8, who previously achieved only a low utterance count and a low level of appropriacy of token use, elevated his utterance count to Medium, but again wrote at a Medium level of appropriacy for tokens used in conditional statements and hypothetical premises. This Medium level was ranked lower than other cases of Medium levels of appropriacy owing to the less frequent use of explicitly connective argument indicators. Although, the relationship between utterance count and appropriacy of token use is not contrastive, the degree of consistency is not strong.

P2, P3 and P4, however, show a strong consistency. They each had low utterance counts (regarding relevant tokens) and achieved only Medium levels of appropriacy of token use with P2 and P3 ranked the lowest with low frequencies of explicitly and implicitly argument
indicators and P4 ranked one rank higher, because of a high frequency of explicitly connective indicator use, but still with P2 and P3 in the Medium rating for appropriacy of token use. It should be noted that P2 and P3 both made a high number of comments about grammatical and orthographic elements, while P4 made a low number of such comments. As I have written (see Section 3.6.6), I designed the OTREs so that grammatical and orthographic elements in the reconstruction text provide as few cues as possible. This was done in order to direct learners’ attention to the points relevant to the lesson, that is the tokens acting as relevant argument indicators. I have noted above that in the presence of other utterances more relevant to the focus of the lesson, comments on grammatical and orthographic elements can suggest that the participant was attempting to make maximum use of the text. However, in the absence of comments concerning relevant tokens, utterances concerning grammatical and orthographic elements may suggest that the participant did not fully comprehend the aim of the lesson or could not detect the presence of, or reliably identify, the relevant tokens. This seems born out in P2 and P3, low use of tokens used in explicitly connective hypothetical premises and P4’s lower number of utterances concerning grammatical and orthographic elements and a corresponding high frequency use of explicitly connective tokens. It is further born out by statements made by P2 and P4 during the semi-structured interviews. P2, who, like P3, made many references to grammatical and orthographic elements and who remained at a Low-Medium level of appropriacy for token use also remained in the Medium rating, she did make frequent use of explicitly connective hypothetical premises. Moreover, her response to cues during the semi-structured interviews were consistent with that written output and utterances:

P2: I used grammatical aspects to clear the exercise.
Gann: Sometimes or occasionally.
P2: If in choices there is capital [letter].

His responses were consistent with his written output in terms of appropriacy of token use and frequency of use. P4, by contrast made only a few remarks concerning grammatical and orthographic elements, and while her level of appropriacy for token use also remained in the Medium rating, she did make frequent use of explicitly connective hypothetical premises. Moreover, her response to cues during the semi-structured interviews were consistent with that written output and utterances:

P4: Grammar or new words?
Gann: Right. Grammar or your new knowledge of the indicator words and the structure of argument.
P4: Structure of argument is more important.
Gann: How about grammar. Was grammar also helpful?
P4: Yes, yes. But long sentences, I can’t answer from grammar. So I think the critical thinking is important for it.

Gann: And how will you use it can you tell me something about that?

P4: I use it—especially the structure of the sentence. So, *kasetsu* [hypothesis], *kinkyou* [current conditions], *kekka* [results]. A series of it is very important for writing a ronbun or report and I have to write about the experiment and the report and ronbun in the future so I can use the structure of the sentence and the key words, for example the conclusion and it’s very good for me.

Contrasting P4’s case, P1 and P9 both of whom made a low and medium number of relevant utterances during the OTRE sessions, respectively, and a high number of comments referencing grammatical and orthographic elements, nevertheless achieved level-4 High Points of appropriacy for token use in their written output in Forum IV, which emphasised the use of hypothetical premises. It should be noted that P1 was an outlier, having made a total of 18 non-relevant utterances, 11 of those relating to premises and conclusions (not conditional statements), and, as noted earlier, there seems to be an overlap between participants who were highly attentive to any elements, relevant or non-relevant, and high ratings for appropriacy of token use. Perhaps as important is the commonality shared by P1 and P9: high frequencies of explicitly connective argument indicators. This serves as a reminder that not only utterance count, but other factors such as frequency of token use may play a role in the levels of appropriacy of token use achieved by language learners.

Lastly, P5, notwithstanding two comments about orthographic elements, expressed strong consistency between a Medium number of utterances related to conditional statements and hypothetical premises and a level-4 High point and High rating for his Averaged level of appropriacy for token use.

Beyond the analysis of individual, the issue of vicarious learning (Sutton, 2001) arises (see Section 3.10.1). The question is: is collaborating, comparatively quietly and passively, in a group, in this case a triad, where one or more of the other members are making a high number or relevant utterances, beneficial to language learning? To what degree can working within a group of peers, within one’s zone of proximal development (Vygotsky, 1978) increase one’s own metadiscoursal performance? I examine each group below.

First, during OTRE Sessions 1-3, despite P2’s impression expressed during the semi-structured interview (see Section 6.4.3) that he had been the main, if not nearly sole contributor during the completion of the OTREs, Group 1’s members made about an equal
number of relevant utterances among all three members, in the Upper-Medium range, as well as comments from each member about grammatical and orthographic elements. In their subsequent Forum V discussion, they separated but remained in the middle range. P3, the least forthcoming member of Group 1 said on this matter during our semi-structured interview:

When you hit a difficult problem, even if you don't know what to say or what to do, you can talk with the people in the group to hear the opinions of others and see the problem from a different angle, so you can solve the problem.

P2’s comment during the semi-structured interview that the other two members may have been advantaged by the collaborative work is arguably born out, albeit to a limited extent, in the performance of the quieter members of Groups 2 and 3.

Analyzing Group 2 was problematic since P5 did not participate in Forum V. However, P4, who made a Lower-Medium number of utterances performed at a High level of appropriacy of token use along with P6. P4 had said during our semi-structured interview: ‘I like doing the OTRES because we can discuss with group members. Other members give advice to me and we can discuss.’ The use of the word ‘discuss’ may be misleading. There were very few cases throughout the OTREs of any group engaging in what would be commonly regarded as discussion. As I noted at the beginning of this section, most communication involved very short utterances of several words with responses from other group members consisting of simple acknowledgement of the statement, agreement or doubt. This suggests that the relationship between utterances made within a group and elevated levels of appropriacy (and frequency) of token use, could be causative. In other words, comparatively passive participants could be learning vicariously. However, following OTRE Sessions 4 and 5, where the focus was on conditional statements and hypothetical premises, in Forum IV, there was no indication of vicarious learning. The less outspoken members, P4 and P5, performed at a much lower level of appropriacy than P6.

Group 3’s data was similar to Group 2. In OTRE Sessions 1-3, P7 and P9 were in the Upper-Medium and High ranking for the utterance count, while P8 was in the lowest having made only a single comment about an orthographic element in the reconstruction text. Yet, P8 reached a level of High appropriacy in Forum V surpassing P7 and joining P9. Contrasting that, after OTRE Sessions 4 and 5, in Forum IV, P8 made a greater number of relevant utterances than did P9. However, P8’s levels of appropriacy of token use were among the lowest, while P9 maintained her level-4 High Point. P7 did not contribute to this forum.
In summary, the data seems to suggest that when the appropriate use of the tokens was not problematic or difficult for the participant, as in the case of premise and conclusion indicators, then he or she performed at a level which suggests vicarious learning may have played a role in his or her performance. However, when the more challenging structure of the conditional statement and hypothetical premise was tasked, the participant’s unskilled performance suggests that, in cases where explicit instruction is necessary, its potential may be limited.

6.7. Discussion
In this discussion, I cover the following points: rhetoric, noticing, interlanguage, OTREs. In a small-scale study such as this, the findings do not always reflect a smooth and unwavering advance forward. Many factors, such as midterm and final exams in other more highly prioritised classes can affect learners’ development through interlanguage, limiting the amount of focus participants are able to allocate. The following are the main points in the findings for Chapter 6. I also consider the findings in respect to the literature.

Rhetoric
Rhetorical styles are often shaped and constrained by societal norms which may differ between cultures (Chen and Baker, 2010). Hyland (2017) refers to reader’s expectations regarding rhetoric as their recipient filter. While the participant-written texts in the forums and chat rooms were not long enough to observe the influence of the ki-shoo-ten-ketsu pattern (Maynard, 1998), or a move away from it, participants responses in the semi-structured interviews indicated that they noticed the gap between what they had learned during their previous instruction in high school and the instruction in university writing that they gained from the text passages modelled in the OTREs. Participant acknowledged that they had held the attitudes prevalent in Japanese rhetoric (Hirose, 2003; Kubota, 1998; Maynard, 1998; Hinds 1987), and which Jennings (2016) found in his interviews with TUS students; but they explicitly stated that by the end of this study they had experienced a shift in their viewpoints about writing, as they learned how the writer bears an equal or greater share of the responsibility for guiding readers’ meaning-making. The result of this shift was more measured and appropriate use of tokens by the participants and this increase in appropriacy contributes to answering RQ2.
Noticing, internalisation, interlanguage

Schmidt (1990) writes that noticing is a necessary condition for input and intake. Participants’ reconstruction of texts modelling tokens was arguably a catalyst for noticing and intake. Through these means, as noted in (Brett, 1994; Thornbury, 1997), language learners activate implicit and explicit knowledge. As with Sanford (2012) and Cheng and Steffensen (1996), as participants’ use of tokens increased in frequency levels, level of appropriacy of token use also increased.

The increase in appropriacy of token use which was observed after each OTRE session is consistent with the findings of Lai and Zhao (2006) where language learners self-corrected their errors; and Swain (1985) which argues that output pushes learners to further process language. Consistent with Cheng and Steffensen (1996), by self-report during semi-structured interview responses and through private utterances whilst working with peers during OTRE sessions, participants in this study demonstrated that they had become aware of the pragmatic and rhetorical utility of MDMs, and the appropriacy of their MDM use tended to increase.

Thus the themes of noticing, internalisations and interlanguage are borne out in participants token use, utterances and semi-structured interview responses as important parts of this study’s response to RQ2.

As shown in (Basturkmen and Randow, 2014; Leedham and Cai, 2013; Hyland, 2008; Williams, 1992), novice English writers often overuse MDMs, which leads to inappropriate use. However, in this study, unskilled use of MDMs was not typically caused by overuse, but instead by general cases of misuse. Participants who were rated on the lower end of appropriacy, had not tended to use a greater number of tokens than other participants. The participants in this study were first-year students, younger and at a lower level of proficiency than the students in the aforementioned studies. These factors may have had a part in these differences as participants were coping with more basic language learning issues.

In addition to studying the effects of the OTREs during and soon after each OTRE session, data was also gathered as many as six weeks after the fifth and final OTRE session. This data was gathered to study the longer term effects of the OTREs within the time constraints allowed by this study.

The findings of the research into RQ2 show that participants experienced a process in their interlanguage development in which they first mimicked the OTRE modelled texts and then later internalised those L2 practices.

Participants responded that working through the OTREs had made this difference experiential and meaningful for them. For example, Participant 3 drew an analogy involving a
piano score: ‘If you just read the score of the piano, you can't play it, but if you actually play it while looking at the score, it's easy to understand . . .’. This recalls Cheng and Steffensen’s (1996) observation that greater awareness of MDMs as rhetorical and pragmatic communicative tools resulted in more skillful and measured use of those MDMs.

The question of why the participants did not respond as positively to the OTRE Sessions 4 and 5 as they did to Sessions 1-3 has a direct bearing on RQ2. Below, I offer three possible explanations.

First, as the formation of English conditional statements is difficult for many Asian second language learners, particularly because of the conjugation of verbs. Second, outside factors may have played a role. The OTRE sessions involving conditional statements and hypothetical premises and the forum that followed occurred during the midterm period of the course. Participants may have been either preoccupied with midterm exams in their science classes, and had their attentional resources (Robinson, 1995) depleted, which could have interfered with the internalisation (Lantolf and Thorne, 2007) of the relevant structures. Third, perhaps the two-clause present tense, future tense construction is more difficult than it appears from a native English speaker perspective. Fourth, first language interference may play a role. The participant’s sentence, *If we change cups the probability becomes 66%* is a direct translation from the Japanese *Kappu o kōkan suru to, kakuritsu wa 66% ni narimasu* . . . It is also likely that participants used machine translation, which often translates statements without nuance or care for cultural difference. A fifth explanation is that, as first year science students, the participants do not sufficiently understand the purpose and function of hypotheses; and students in the Mechanical Engineering department do not typically receive instruction concerning hypotheses until their third year, a point I learned after the completion of this study.

It may be that more than one of these explanations is true for one participant or another. A final sixth possibility is that WebSequiturs are not as well-suited to teaching conditional statements as they are for teaching premises and conclusions. Future iterations of this study should focus on this issue.

**OTREs and WebSequiturs**

The findings in Chapter 6 also were consistent with Higgins and Whites’ (1999) remarks about the importance of breaking sentences at specific points which circumvented users making use of grammatical or syntactical features of the text. During the OTRE sessions
participant made few utterances relating to grammar and syntax and instead focused on the argumentative structure and the tokens that signalled them.

Other small-scale studies with which my findings are consistent are Gánem (2006), which involved collaborative work using Hot Potatoes to model phrases and produced L2-related utterances, during the Hot Potatoes sessions; and Mohammed, Assam and Saidi (2020), which studied the use of Web 2.0., including Web Sequitur, to promote the internalisation collocations and common phrases. The participants in my study, likewise, demonstrated that they had internalised the tokens and could use them later without access to the Master MDM List. As Participant 5 said, ‘If I didn't do OTREs, the term probably wouldn't have stuck with me as much’.

6.8. Summary
The findings of this chapter build upon those in Chapter 5. I have reported the findings related to RQ2, which were concerned with the role OTREs play in guiding language learners to use tokens appropriately within their future academic or professional discourse communities.

This chapter first contributes the means to interpreting appropriacy. I wrote in Section 4.2.2 of developing and deciding upon my four-level scale. Previous studies such as Bax et al. (2019), Noble (2010) and Intaraprawat and Stefferson (1995) influenced me to form a small group of independent assessors who were familiar with the research context and able to use a shared criteria to rate each word or phrase, item by item.

Second, I looked at the students’ attempts to employ premise and conclusion indicators before and after the introduction of the first OTRE sessions; and later to construct and employ hypothetical premises both before and after completing the fourth and fifth OTRE sessions. Using the scale which I introduced in Chapter 4, three independent native English speakers assessed the written output of the 9 participants in the main study. This written output included forum discussions, chat room discussions, a Group Report and an individually written In-Class Rewrite. I analysed these assessments for averages of appropriacy for the tokens used in each text and each High Point of token use. I also divided this analysis between an assessment of the MDMs used to express premises and conclusions and an assessment of MDMs used to express conditional statements and hypothetical premises.

Third, I provided examples of the student-student talk during the OTREs. I noted that the talk was often not illuminating but did at times establish that the participants were engaging in metalinguistic discourse. Some student-student talk did provide a view into the strategies participants used to complete the exercises. I categorised the variety of comments
and quantified the number of comments relevant to token placement as well as comments not
relevant to token placement. I quantified comments by category for each participant. Then, I
analysed the quantisation, looking for areas in which the most relevant comments grouped.
Based on this, I developed a scale of quantity similar to the four-level scale of appropriacy. I
did this so that, in a single table, the quantisation of OTRE session utterances could be
compared with the subsequent level of appropriacy of token use achieved during following
forum discussion. In this way, I searched for cases of consistency and contrast between
utterance counts and appropriacy levels. This scale consisted of Low, Medium and High
quantities of relevant utterances. The scale differed in that the Average and Skilled rank of
appropriacy were both matched to the Medium utterance count. When relevant I referred
Lower-Medium and Upper-Medium. I also further refined the appropriacy ranking by noting
cases where there had been low and high frequency of token use, with a priority on explicitly
connective argument indicators, which was the main focus of the OTREs and of RQs in this
study.

In closing, to add to the explanatory potential of this data, I referenced participants’
responses to the semi structured interviews. This use fulfilled several functions, among these
being evidence of vicarious learning in cases where low utterance counts were followed by
Upper-Medium and High appropriacy of token use. Along with the appropriacy ratings and
the utterance count, the semi-structured interviews completed the means for triangulation in
this mixed-methods study. Through a mixed-methods approach and triangulation of these data
sources I could conclude that there is evidence for the efficacy of OTREs in guiding language
learners to use tokens appropriately. Moreover, in their own words and their own way of
expressing the concepts discussed in Section 2.7.5, participants showed that they grasped and
concurred with the model of imitation, interlanguage and internalisation and that they
recognised high quality interaction with their group members, within their own ZPDs, as both
a tool for learning, and as evidence of their having achieved greater automaticity and
appropriacy with the OTRE modelled tokens. Thus, the triangulation of these data sources
enabled me to answer RQ2.
Chapter 7: Conclusions, implications, limitations and future research

7.1. Introduction
This study came about as a result of my having perceived, as early as 2010, the need for materials for and approaches to critical thinking instruction in EFL contexts which would focus on the language of basic CT skills as much as expressions of critical dispositions. A confluence of three points, informed a shift in my teaching practice: (1) at the first-year university level, a basic organisation of arguments and the language used to indicate premises and conclusions; (2) blended learning techniques such as the flipped classroom, podcast delivered lectures and asynchronous online communication to preserve class time for student-student interaction; and (3) OTREs providing a structured path for language-focused collaborative learning and student-centred class meetings.

In the following years at TUS, the faculty and administration underwent a shift, as it responded to mandates from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for CT skills to be taught at the university level (MEXT, n.d; Stroupe, 2006); and the corresponding response within TUS, calling for those mandates to be satisfied (Tokyo University of Science, n.d.; 2016) and within the curriculum for the TUS English program (see Appendices 3-5). Sources cited in the literature section of this thesis (see Section 2.6) recommend a focus on MDMs in undergraduate English courses at TUS and suggest that this might be applied in other science universities in Japan as a way of responding to those mandates and institutional policies.

In this concluding chapter, I both review the aims of this small-scale study and the analytical methods applied to the various data. I then summarise key findings and discuss the extent to which the findings are relevant in my classroom, the classrooms of other teachers at TUS, and teachers in similar science university settings’ first-year EFL English courses. I note how this study points out a way toward satisfying the institutional demands and national educational reform. I propose that this thesis advances knowledge useful to both action researchers in the EFL field and potentially to decision makers operating at the institutional and national level.

7.2. Review of study aims
This study focused on the written output of nine Mechanical Engineering undergraduate students in first-year, second semester, four-skills English course at a leading science university in Japan. Through my small-scale study, I investigated how the pedagogy of
written argument, specifically the use of MDMs used to express premises, conclusions and conditional statements, particularly future predictive, i.e., hypothetical premises could be improved through online text reconstruction exercises. This research had two major aims. The first was to discover the extent to which the use of online text reconstruction exercises (OTREs) influence participants to use tokens with greater frequency and more appropriacy. The second aim was to look into the otherwise hidden experience of participants’ individual and social experiences during and after the OTRE sessions to better understand their perceptions of the purpose, utility and effectiveness of the exercises.

To explore these issues, I developed the following research questions (RQs):

1. To what extent does completing OTREs and particularly WebSequiturs influence language learners to use selected tokens in their written arguments?
2. To what extent does completing OTREs and particularly WebSequiturs guide language learners to use tokens appropriately in such a way that indicates skilful signalling of academic argumentative structures in written discourse?
3. What are participants’ perceptions of their experiences with OTREs and particularly WebSequiturs?

While I addressed RQ1 and RQ2 separately in Chapters 5 and 6, respectively, RQ3 was addressed in both chapters due to overlapping themes. These first two RQs were determined by considering what metrics would be most closely related to the improvements in students’ argumentative writing and interlanguage development. RQ3 was included, as, in this interpretivist study, participants’ subjective perspectives held insightful qualitative data, by which I was able to further triangulate other qualitative data and quantitative data. In this way, participants possessed agency and were valued co-constructors of knowledge (Kvale, 2007). In some cases, I took a doxastic approach and limited myself to recording the beliefs expressed by the participants (Brinkmann, 2007). In other cases, I challenged participants to join in a thread, and to explain their responses (Curato, 2012). This resulted in a more robust response to my RQs than many previous studies in this area have achieved.

In addition to learning about the resulting output, I also wanted to hear what participants thought, upon reflection, regarding the utility and effectiveness of OTREs in developing their AW skills, and particularly, their skills in MDM use.

To answer RQ1 and RQ2, I used a mixed-methods approach (Denscombe, 2008). I collected quantitative data, which I analysed via several parameters discussed below. Afterwards, participants’ lived experiences were explored through a thematic analysis of
transcripts of their talk during OTRE sessions and later through semi-structured one-on-one interviews. In such cases, the answers to RQ3, which were also qualitative, were useful in further nuancing the interpretations of the data in Chapters 5 and 6. In other cases, RQ3 brought issues to the fore and revealed possibilities in the interpretation of the data that I had not considered until near the end of the study. Comparing the types of utterances to the subsequent levels of appropriacy of token use was one such instance. This made clear the value of remaining flexible in order to gain the greatest possible understanding from the available data.

While other studies have looked at the general frequency of use, such as Bax et al., (2019) and Khedri et al. (2013), I found no other studies that have examined the use of these types of MDMs by first-year university students of English as a foreign language. Moreover, none of these studies draw upon participants’ utterances during the OTRE sessions in order to understand the situated use of MDMs; nor do they include the responses of participants from semi-structured interviews reflecting on the tasks used to elicit argumentative language. To that extent then, this study presents pioneering research in the field of EFL and AW pedagogy.

As I pointed out in Section 2.10, although text reconstruction exercises (TREs) have been promoted in language instruction (Brett, 1994; Thornbury, 1997), empirical studies of text reconstruction exercises used to develop students’ skills at using MDM in AW are scant. A handful of studies were available to guide my exploration of this topic. Although it used paper-based sentence-combining and did not touch upon MDM use, Kameen (1978) supported the value I placed on collaboration in text reconstruction. Although Rodríguez’s (2010) use of WebSequitur supported my own use of that application, again, MDMs and AW were not the focus of that study. Čechovičová’s (2010) use of Hot Potatoes and her acknowledgement of her students’ immediately increased motivation resonated with what I observed in my classes. Moreover, her noting that WebSequits can be used to increase awareness of textual coherence further validated my interest in the application. Specifically, only Shintani and Aubrey (2016) used text reconstruction, like this study, for online collaborative reconstruction of texts with a focus on the grammatically and syntactically correct formation of conditional statements.

7.3. Findings
This section presents the main findings and implications from data that was collected. This study, being interpretivist, relied first on a foundation of quantitative data gathered subsequent
to each forum, chat room and both group and individually tasked writing assignments in order to answer RQ1 and RQ2. Those data were augmented with thematic analysis of transcripts of talk between participants and data then co-constructed during semi-structured interviews. An overview of the findings in relation to each research question is displayed in table below.

<table>
<thead>
<tr>
<th>RQ</th>
<th>Research findings</th>
<th>Main unit of analysis</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Word count</td>
<td>word</td>
<td>Post-OTRE general increase or maintenance of the Word Count except in the ICR where there was a general decrease</td>
</tr>
<tr>
<td></td>
<td>Token count</td>
<td>token</td>
<td>General increase or maintenance in token count within each frame</td>
</tr>
<tr>
<td></td>
<td>Unique uses</td>
<td>token</td>
<td>General increase in unique uses of unique uses of tokens within each frame</td>
</tr>
<tr>
<td></td>
<td>Communicative functions</td>
<td>communicative function</td>
<td>General increase in number of relevant communicative functions, within each frame</td>
</tr>
<tr>
<td></td>
<td>Premise and conclusions indicators</td>
<td>token</td>
<td>General increase in the number of premise and conclusion indicators, except Forum IV (in Frame 2)</td>
</tr>
<tr>
<td></td>
<td>Factual conditionals and future predictives</td>
<td>token</td>
<td>General increase or maintenance in the number of tokens used in conditional statements (factual conditionals); and hypothetical premises (future predictives)</td>
</tr>
<tr>
<td></td>
<td>Implicit and explicit premise and conclusion indicators</td>
<td>token</td>
<td>Inversely proportional relationship between changes in frequency of implicit and explicit premise and conclusion indicators</td>
</tr>
<tr>
<td></td>
<td>Influence of genre difference on token use frequency</td>
<td>token</td>
<td>Differences in genre did not affect the increases and maintenance of the frequency of the various variables</td>
</tr>
<tr>
<td>2</td>
<td>Levels of appropriacy of token use (average)</td>
<td>assessment levels of appropriacy of token use</td>
<td>All participants began at an averaged level of Average (Level-2 or nearly Level-2)</td>
</tr>
<tr>
<td></td>
<td>Levels of appropriacy of token use (high points)</td>
<td></td>
<td>General increase or maintenance of Averaged and High Point token appropriacy levels for indicators, except in Frame 4 where there was decrease after the GWR in the ICR in the Averaged appropriacy but maintenance of High Points in the ICR</td>
</tr>
<tr>
<td></td>
<td>Association between participants’ utterance counts and level of appropriacy</td>
<td>OTRE utterance</td>
<td>In OTRE Sessions 1-3, focussing on premise and conclusion indicators, a high token-relevant utterance-count was closely associated with subsequent high assessments for token use appropriacy.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In the OTRE Sessions 4-5, focussing on conditional statements and hypothetical premises, the medium to low token-relevant utterance-count closely associated with subsequent medium to low assessments for token use appropriacy.</td>
</tr>
</tbody>
</table>

David A. Gann (C8846053)
| Participants’ understanding of OTRE’s purpose | Interview response | Participants demonstrated an understanding of the OTRE’s purpose as a focus on the language of argument, i.e., the expression of opinions, with support. Participants demonstrated an understanding of the OTRE’s purpose as providing collaborative support in their language learning. Participants demonstrated an understanding of the OTRE’s purpose as providing experience in collaborative work situations. |
| Contents and import of participants’ utterances during OTRE sessions | OTRE utterance | Participants during OTRE sessions focused on the following points: Negotiation of answers / disagreement / correction and clarification Identification of premises, conclusions, conditional statements and hypothetical premises Identification of tokens functioning as indicators of premises, conclusions, conditional statements and hypothetical premises |
| Commonalities in participants’ past knowledge of tokens | Interview response | Participants’ experience with argument indicators varied, but all were familiar with many tokens. Most were not acquainted with thinking about the tokens within multi-clause structures. They were able to recognise logically connective tokens as such. |
| Participants’ reliance upon grammatical features during OTRE sessions | Interview response | Participants generally found grammatical features unhelpful in completing OTREs and attempts to use grammatical features in the OTREs focussing on premises and conclusions quickly decreased. Participants resorted to grammatical and orthographic features during the OTREs focussing on conditional statements and hypothetical premises when they found the reconstruction of these two-clause structures too challenging. |
| Generalisation or transfer of knowledge | OTRE session remark | There was evidence of knowledge transfer as what was learned was reapplied in a different multiple-choice selection and the participant noted the transfer verbally. |
| Participants’ utilisation of textual features as tools | Interview response | Participants reported that the linear (beginning to end) sequence of the reconstruction assisted in focussing their attention on argument indicators and logical flow within argumentative structures. Participants reported using language as a tool, suggesting a metalinguistic understanding of language. |

Table 7.1: Findings organised by RQs 1-3

7.3.1. Research Question 1: OTREs and frequency of MDM (token) use in AW

In this section, I summarise the findings regarding RQ1 (Chapter 5). In order to determine the extent to which OTREs influence participants to increase their frequency of token use, I draw upon quantitative data from the nine participants’ writing in which they used tokens to indicate premises, conclusions, conditional statements and hypothetical premises. The writing drawn upon for quantitative data was generated from forums, chat rooms, a Group Written Report (GWR) and an In-Class Rewrite (ICR) from memory of the GWR (see Section 5.2). I drew upon personal observation of participants while they were completing the OTREs. I also drew upon thematic analysis of transcripts of talk between participants during the OTRE.
sessions and participants’ responses during semi-structured interviews to further support the claim of a causative link between OTREs and increased frequency of token use (see Table 3.5).

As seen in Table 6.1, a quantitative analysis showed that in the forums and chatrooms there was a general increase in the frequency of token use, in the use of unique tokens within the relevant communicative functions. In terms of the number of unique, i.e., single, uses of tokens when the differences in the amount of writing in each forum, chat room and writing assignment are taken into consideration there was a clear increase in two important areas of participants’ written output. First, the variety of phrases used to express premises and conclusions within a given communicative function increased; and the range of communicative functions increased. In both areas, these increases were observed in both the explicitly connective and the implicitly connective tokens that students pressed into service to present their arguments, much like MDM use increased in Bax et al. (2019). The increases in frequency of token use and the other aspects of token use continued throughout the four frames until the GWR, after which there was a decrease in the frequency of those aspects in the ICR. Qualitative data, in the form of semi-structured interview responses suggested that the drop in word count in the ICR might be connected to the absence of the social component provided through peer collaboration.

The number of tokens used was also often exceeded by the number of words used in tokens to a degree which showed that participants were also expanding their repertoire of tokens from those initially introduced, to tokens, further down the Master MDM List. These tokens further down the list were often made up of more words. Hence, an exploration of the list could sometimes be result in an increased number of words used in tokens. More central to this thesis, notwithstanding low numbers due to short responses in Forum III and IV and the brevity of some of the Group Report submissions, there was a general increase in the use of explicitly connective premise and conclusion premise and conclusion indicators (EC PC). It was also seen that implicitly connective premise and conclusion indicators (IC PC) were often used in ways that suggest or presume logical connections and that these EC PC indicators and IC PC indicators often compete. In other words, both usually did not increase in use at the same time; instead one increased at the expense of the other.

The findings suggest that students in this undergraduate first-year English course, having completed OTREs collaboratively, were influenced to increase their use of tokens appreciably; and this conclusion was supported by participants’ remarks and responses during the semi-structured interviews. Since no previous studies involving Hot Potatoes or
WebSequiturs inquired into participants’ experiences with the exercises, it is left to future studies with similarly interpretivist perspectives to further this line of inquiry.

In contrast, findings related to participants’ use of the tokens deployed to form conditional statements and hypothetical premises did not as clearly suggest the efficacy of OTREs. Participants frequently did not use the construction that had been modelled and instead invented novel ways to express a future tense in the latter half of the hypothetical premise. While this was not seen as wholly negative, it had to be viewed as a failure to adequately emulate or internalise the writing practices of the discourse communities (Khedri, Heng and Ebrahimi, 2013) that many of the TUS students will soon wish to join. This should, however, be viewed in light of the observation that language learners’ path through interlanguage to competence is not always a straight one (Shrestha, 2020).

To sum up, semi-structured interview responses completed the triangulation between the quantitative and qualitative data sources and provided answers to RQ1. The following three findings were noted. Participants’ responses during semi-structured interviews expressed consistency with earlier quantitative data from the written output and indicated that OTREs were effective in influencing students to use MDMs. The findings of Chapter 5 strongly support the conclusion that OTREs can influence students positively to employ MDMs more frequently with greater variety of words and phrases, and with an expanded range of communicative functions. The findings also suggests that OTREs have a strong social constructivist function, as they appear to be influential when language learners work collaboratively to complete them (Vygotsky, 1978).

7.3.2. Research Question 2: OTREs and appropriate use of MDMs (tokens) in AW

Following my argument in Chapter 5 that OTREs are effective in influencing language learners to increase their frequency of token use, I showed, in Chapter 6, that participants who completed the series of OTRE sessions also used tokens with increased appropriacy. I drew upon quantitative data from the nine participants’ writing in which they used tokens to indicate premises, conclusions, conditional statements and hypothetical premises. The writing drawn upon for data here was the same as that drawn upon in order to address RQ1. I later drew upon thematic analysis of transcripts of talk between participants during the OTRE sessions and participants’ responses during semi-structured interviews to further support the claim of a causative link between OTREs and increased appropriacy of token use (see Table 3.5).

A review of the literature (see Chapter 2) indicated studies that one of the strongest
factors in determining a high score for effective argumentation and being appraised as publishable in academic journals was not only the frequency of token use, but also the appropriacy of that use in context (Intaraprawat and Steffensen, 1995). With this in view, using a scale of 1-4 (see Section 4.2.2), three qualified and trained independent readers assessed the appropriacy, in context, of roughly 750 tokens used by participants over the course of five forums, two chat rooms, a Group Report and an In-Class Rewrite. The analysis shows, unambiguously, that the participants’ token use increased in terms of appropriacy. However, the findings also revealed a tendency by participants to deploy implicitly connective tokens when explicitly connective ones would have made clearer and more forceful arguments. Participants’ metalinguistic talk that occurred during OTRE sessions and chat room discussions provided some explanations of this phenomenon. Both the quantitative findings and the responses provided in the semi-structured interviews suggest that WebSequiturs in concert with explicit instruction leads language learners to notice (Schmidt, 1990) tokens, and provides motivation needed to form an intent to reach a level of automaticity (Schmidt, 1990) in token use in the long term. As Thornbury (1997) notes, text reconstruction exercises provide the means for students to restructure their linguistic understanding by noticing of the gap between their current stage in interlanguage and the correct forms modelled in the exercises. As with Mohammed, Assam and Saidi (2020), this was borne out through the findings which indicate that OTREs played an important role in guiding students to appropriately use MDMs functioning as premise and conclusion indicators.

The findings involving conditional statements and hypothetical premises were less conclusive. Participants’ written output subsequent to the OTREs focusing on hypothetical premises, was inconsistent in terms of their demonstrating mastery of the present tense + future tense, i.e., *future conditional*, structure. Yossatorn et al. (2022) point out the difficulty many Japanese have with forming conditional statements (see Section 3.4.2). Some participants showed improvement between Forum III and Forum IV, and, as with Shintani and Aubrey (2016) and Katip and Gammper (2016), found hypothetical premises difficult, and did not improve. Participants who showed improvement in Forum IV did not always reproduce the form in the Group Report and/or the In-Class Rewrite. However, in semi-structured interviews, participants who had been the most successful in that area responded to interview cues in ways that were generally supportive of the idea that OTREs had been helpful. As with RQ1, the question remains whether more OTRE work or better designed OTREs may correct against ‘inappropriate usage of L1 linguistic knowledge during L2 production’ (Yossatorn et
It is also a matter of conjecture as to whether there may be a delay between instruction and improvement, or competence and performance (Ellis, 1994) as some students may, after the completion of the class, encounter appropriately structured hypothetical premises, which may resonate with what they first formatively learned through the OTREs. The foundation having been laid, these more difficult structures may require longer correction from language instructors and other teachers, and influence through reading science journals or other texts where these structures are frequently used.

7.3.3. Research Question 3: Participant perceptions of OTREs

RQ3 was concerned with participants’ perceptions of their experience completing the OTREs and was addressed entirely though qualitative data. Thematic analysis was applied to transcripts of talk between participants during OTRE sessions. As noted earlier, there is an overlap in the areas covered by RQ3 and the first two RQs. Other qualitative data were collected specifically to understand the experience of the participants.

Irrespective of the varying extent of demonstrable gains individual participants made, during the semi-structure interviews, participants roundly agreed that OTREs had been effective in influencing them to use tokens more frequently. Thematic analyses of participants’ utterances during OTRE sessions as well as responses given during semi-structured interviews indicate that participants understood the purpose and aims of the OTREs. Whereas my primary interest was in increases in token use and appropriacy, participants indicated that they viewed gaining collaborative social skills to be as or nearly as important. However, as in (Andujar and Salaberri-Ramiro, 2021), participants also commented during semi-structured interviews, often at length, on the role of OTREs in focusing their attention on appropriate token use. Also, as with (Bailey, Almusharraf and Hatcher, 2021) participants said that the online peer work, was motivating and gave them the opportunity to work with others.

Participants’ early response to the Likert scale questionnaire and the later semi-structured interview responses indicated that many of the tokens in the OTREs were new, but that some were familiar. As in Rinnert and Kobayashi (2009) where Japanese students of English gained a greater appreciation for MDMs in argument, participants in my study likewise noted that, through the OTRE group work, they had come to think of tokens, not only as expressions of what they generally felt and wanted to say, but as tools to organise their ideas into logical structures that gain them admission into discourse communities and empower them by supporting their efforts to be understood during dialogic discourse.
Due to the design of the OTREs, participants reported that they were largely unable to use grammatical or orthographic cues. Thematic analysis of transcripts of utterances made during the OTRE sessions added further evidence to support the claim that good design of the OTREs could direct learners’ attention toward premise and conclusion indicators and away from grammatical and orthographic features. Participants’ comments on these features decreased as participants continued to find them unhelpful. There were cases of evident learning transfer in which participants made suggestions during the OTREs, which they explicitly noted were based on earlier structures of AW. However, when the task became too challenging at their interlanguage development level at that time, as in the case of the conditional statements and hypothetical premises, participants reverted to attempting to use grammatical and orthographic cues.

During semi-structured interviews, participants were asked about the role of the three aspects of the activity: (1) the interactive nature of the text reconstruction as compared with simply reading completed exemplary texts; (2) the computer-based nature of the activity as opposed to paper-based text reconstruction; (3) and the collaborative dimension of the activity contrasted with each student completing the exercises individually. It has been noted already in various parts of the two previous chapters, but to restate here concisely, participants’ responses during semi-structured interviews indicated that they were aware of the differences between the top-down teacher-centred or textbook-centred instructional style to which they were accustomed; that they both enjoyed the novelty of the text-reconstruction and the interactive nature of the computerisation; and that they also enjoyed and valued the presence and support of their peers during these sessions. Only one, out of the nine participants, was critical of her experience, noting that it was not always clear why a selection was right or wrong.

7.4. Implications for faculty and teachers, administration, and policy makers
In this section, I discuss the implications that this study may have in the larger spheres of other practitioners, administrators seeking ways to meet governmental mandates and policy makers issuing those mandates.

7.4.1. Implications for faculty and teachers
In this section, I discuss how OTREs can help faculty fulfill curriculum goals, help instructors meet teaching goals and help students reach learning goals.

Ideally, all teachers employing OTREs to the same effect as I have in this study would
have the same information made available to them regarding the pedagogical underpinnings of their use and perhaps ongoing departmental discussions led by knowledgeable faculty members could be helpful in reaching shared understandings of the use of OTREs; and teachers could be encouraged to treat their students with the same agency as have the students in this study. Hence, one key to maximizing gains in the frequency and appropriacy of students’ MDM use is students’ perception of their instructor being knowledgeable and confident about the use of the tools and applications they are tasking students to use. It is therefore critical that teachers have a working understanding of the theoretical underpinnings (Humphries and Burns, 2015) supporting the practice of utilizing OTREs.

Convincing practitioners of OTRE’s pedagogical utility may not always be easy. Several obstacles impede their introduction at the classroom level. In his research on educational reform at TUS, Jennings (2018) found a general scepticism among teachers regarding the effort required to reform their teaching practice justified by students’ need for instruction of the kind which supports learning English used in dialogic discourse.

Perceived threats to teachers’ identity as specialists are likely to be a source of resistance to the implementation of new teaching practices. As the top-down approach is replaced by approaches more aligned with Communicative Language Teaching (CLT) practices and collaborative learning, Japanese teachers may feel unqualified in a position in which hitherto they have felt comfortable and unquestioned (Cheung, 2015) and may feel that their sense of ‘teacher identity’ (Yayli, 2015) has been threatened. Many teachers will be reticent to divulge such private reservations and anxieties. Whether or not the case for the implementation of a curriculum can be successfully made depends to a great degree on engaging with these often-unspoken belief systems (Maskit, 2011). Raising teachers’ awareness and understanding how OTREs can work in concert with issues-based textbook reading may ease reticence toward that change.

Reticence against educational reform, e.g. the introduction of OTREs, may also be felt by students. A review of literature (Humphries and Burns, 2015) suggests that traditional educational practices secondary school can be carried over into university by students whose secondary education was entrenched in the practice of yakudoku (Takase and Otsuki, 2012). These students may also be reticent to express themselves orally. This reticence may also be found in text format discussions such as forums and chat rooms. One participant in this study expressed the sense that commenting in the forums would obligate other students to respond. The deeply socially embedded Confucianism noted in Section 1.3 may be an impediment in motivating students to engage in the critical questioning of their peers involved in the use of
some of the MDMs in this study. University students may avoid discussions that involve disagreement and asking for reasons, as, in Japanese culture, these may be seen as confrontational (Bao, 2014).

Participants in this study tended to participate actively, enjoying the challenge and the collaborative nature of the task; and they tended to take the activities seriously. Students, like teachers, may exert a sort of passive resistance to the course design I have described in this study, by failing to participate in the post-TRE forums. However, this study has demonstrated that students were motivated by being treated as active agents in their development of their own vocabulary. Whenever possible, I employed a method of explaining, both in class and through the podcast, what our class goals were, why we were pursuing those particular goals, how we were going to go about doing that, and at times informing them about the background literature and prior studies which informed, supported and justified my approach.

7.4.2. Implications for administration

In this section, I discuss implications that this study presents for administrations seeking to respond to the mandates for educational reform, by teaching basic critical thinking skills, the language of AW in English instruction while also introducing tasks in lessons which foster students’ collaborative learning and cooperative work skills.

Many institutions like TUS invite reform, but also seek to preserve the nihonjinron notion of Japan’s cultural uniqueness (Hoffman, 2013). Such institutions wish to internationalise so that they can play a contributing role in supporting Japan’s continued relevance in the current and future era of globalisation. At the same time, as mentioned in Section 1.3, they continue to promote and enforce Confucian values such as obligation, duty and social cohesion (Tjeldvoll, 2011; Hu, 2012; Gray, 1998). Institutions that wish to instil these values in students must find a balance between pre-reform era values and the changes that educational reform will necessitate. This can create tensions at the administrative level. These changes promise to usher in changes in the long undisturbed social stratification so that students, and future leaders in international contexts have the language skills to sometimes interrupt, make suggestions, challenge claims, and ask for reasons (Kinsella, 1997). They will require the language skills to do that in a way that is socially acceptable and which matches the norms of the discourse communities that they intend to join.

This brings us full circle back to the teachers. Butler (2011) points out that administrators calling for the reform often do not have a clear idea what that reform should or could look like in curricula or in the classroom. Teachers must take the final initiative to
implement reforms within socially constrained expectations of EFL classrooms and within administrative guidelines. OTREs provide a means for doing so, while effectively responding to MEXT mandates for reform. Beyond serving the needs of the single practitioner in the classroom, they can be serve as a basis through students first two years of core courses, designed with points that connect with the requirements of the curriculum.

7.4.3. Implications for national policy makers
In this section I discuss what implications national policy might be based on the findings of this study. When policy makers in Japan responding to the demands of increasing internationalisation consider the issue of English language teaching materials, many are naturally concerned with ensuring that educational reform occurs which satisfies corporate demands for workers capable of working and researching autonomously (DeWaelsche, 2015; Yonezawa, 2011). Stakeholders in markets with increasing competition will have an interest in maintaining and improving the instruction of English communication skills (Kubota, 2011; Takayama, 2009) so that the upcoming generation of Japan’s graduates are prepared to make a contribution, where English communication is concerned, to the companies that hire them (Grove, 2012).

Both administrators and policy makers should therefore encourage student-centred learning and methods of teaching which promotes critical thinking decision making in addition to the development of AW skills. Development of OTREs specific to each context should be carried out through a bottom-up approach, but policy makers can provided funding for training and support on a national level.

Because OTRE platforms like Hot Potatoes are free to use, are open source and can be plugged into many university Moodle Learning Management Systems (LMS) platforms, they cost virtually nothing to implement. As for logistical concerns, some universities, like TUS, which do not have a sufficient number of computer labs to support the campus-wide use of OTREs, might be supported with the funding needed to purchase a limited number of tablets which teachers could check out from a central storage area. Alternatively, students can use their own laptop computers, which increasingly, most students have.

7.5. Limitations and future iterations
In this section I discuss limitations and problems that limit the reliability of some conclusions that may be drawn. I point out ways that future iterations will improve on this one.
First, as a small-scale study, this study was limited by number of participants involved. I was able to group participant performance in the forums, chat rooms, the GWR and the ICR by trends in token use along various metrics. Had the number of participants been fewer, then my ability to note these generalities would have been greatly decreased. Yet the amount of data produced by nine participants presented enormous challenges for one researcher to process and analyse. Increasing the number of participants could produce more robust expressions of those trends and might also allow other trends to become salient which were not detectable in this study.

The study was also limited to students from one department, the department of mechanical engineering. TUS, however, has 12 departments and the student population of each department has characteristics which teachers who have taught at TUS for some years generally find to hold true. Students from some department may be more motivated than others; or are more outgoing and talkative; or excel in individual written tasks more so than in oral communication. Some departments may also have greater expectations than others about how important English communication skills are likely to be in their careers. Moreover, the TUS is divided into two colleges: the College of Science and Technology and the College of Pharmaceutical Science. The campus of each is separated by a several minute walk and the cultures between these two colleges is quite distinct. A large study involving as many of these departments as possible could possibly yield a far richer diversity of results that could greatly increase the body of knowledge of the use of OTREs in English courses with students from various scientific disciplines.

Another limitation of this study involved the narrow amount of time in which the English course occurred. During the 15-week course the first OTRE session had to follow the pre-OTRE initial writing assignment by which the baseline frequency and appropriacy of tokens use for each participant were determined. The last OTRE session had to occur before the time during which participants had their GWP underway. This left a ten-week span in which, at most, time for five OTRE sessions could be allocated. This schedule raises several questions.

First, were five sessions sufficient to bring participants’ token use to a level of automaticity in their use of the tokens so that meaningful measurement of frequency could be made? In other words, how much of the change in the frequency of token use was due to the influence of the OTREs and how much was due to participants’ perception that an increase in token use frequency was desired and expected?

Also, is there a lag between learning that occurs through OTREs, for some
participants increase their frequency of token use or their level of appropriacy of token use at a later time?

Second, how long is learning through OTREs retained? Does an increase in the number of OTREs in a one-year span result in increased gains in frequency and or appropriacy of MDM use? At what point does a continuation of OTRE sessions no longer influence learners to use MDMs with higher frequency?

Third, how does an increase in the number of OTRE sessions and/or a continuation of OTRE sessions affect levels of appropriacy. Can OTREs be used indefinitely to model appropriate use of a broadening array of MDMs? The gains made through a two-year program might be far greater than those made in 15 weeks. And whereas the residual effects of the 15-week course might be limited, students passing through a two-year program might retain a far greater number of the MDMs, might use them more frequently, use them more appropriately, and this in turn would further bolster the argument for the implementation of OTREs. Coordinating a two-year research project over the four core English courses at TUS (Freshman English and Integrated Skills in English 1, 2A and 2B) could help to answer these questions.

In addition to these questions, there were several questions which I later wished I could have asked participants during the semi-structured interviews. However, I did not have these questions in mind at the time, because I had not yet analysed the data. For example, I would like to have asked several participants why they were inconsistent in forming hypothetical premises, sometimes constructing them as modelled in the OTREs and other times using inappropriate constructions. By the time that data analysis had been carried out many weeks and months had passed, and participants’ memories would have been too questionable to consider their responses to be reliable. However, by noting these questions, I can be prepared to explore them again in subsequent iterations of this study.

In addition to unanswered questions this study had limitations caused by flaws in the research design. One of these concerned semi-structured interview cues, some of which were potentially leading. It was later noted that participants might intuit the answer which they perceived was what the researcher would desire. Although it was not the case that I wished to elicit anything other than an answer that expressed a participants’ genuine response, and this sentiment was expressed to each participant before the beginning of their interview, the possibility and likelihood existed that a participant would answer in a biased way. In Chapters 5 and 6, I attempted to minimise my use of these questions and rely on interview cues for which there was less possibility for bias to influence their responses. I should note that in
cases where such a possibility for bias did exist, responses did not typically suggest the influence of bias. Because of the teacher-student relationship which I had fostered participants appeared comfortable in expressing describing their experience in both positive and negative terms. For example, one participant expressed dissatisfaction with not being informed during the course of completing an OTRE, why a given response was incorrect. Another participant, P2, shared that he felt that collaborative nature of the OTRE sessions resulted in him doing most of the decision making—although an examination of the transcripts indicated that this was not quite the case. It should be noted however that through triangulation of data, cases of probable or biased responses can be detected. For example, when a response is given which would clearly strengthen the claim of the researcher, does not comport with the other triangulated data, then that response may be viewed as possibly compromised and lacking credibility (Clark and Ivankova, 2016). In this study, as noted in the findings, no cases of contradiction between semi-structured interviews and written output or OTRE performance was detected. Nevertheless, future iterations of this study must include emended interview cues to maximally obviate the opportunity for bias to enter the data collection.

Lastly, this study has been somewhat limited by my coming to learn about the work of Halliday only after I was well into this study. Had I read his work earlier, my research tools might have been more refined, benefiting from the greater depth and breadth of his categorisation of ideational, textual and interpersonal meanings (Halliday, 2009). Also, the categorisation by Hyland (2005a) of particular types of phrases would have also informed the design of my analytical tool. As my position on metadiscourse markers was compatible with what I later learned from Halliday and Hyland, the tools I developed were sufficient to the task. Future iterations of this study will incorporate these works into the research design.

The use of OTREs and particularly the kind used in this study brings both problems and solutions for practitioners, administrators, and policy makers. In this study, there is no assumption that an effective teaching approach and well-designed WebSequiturus will result in unwaveringly increased MDM use. However, these results indicate that this study merits further enquiry. I hope through this study and future publications and networking with other researchers with similar interests to contribute to educational reform at TUS and in other contextually similar institutions.
References


opinion polling’, *Qualitative Inquiry*, vol. 13, no. 8, pp. 1116-1138.


Educational Testing Service. (n.d.) *TOEIC score vs CEFR*, [Online]. Available at


Informatics, pp. 92-99.


learner academic writing’, vol. 9, no. 2, pp. 145-169.


pp. 181-205.


Swain, J. (2018) *A Hybrid Approach to Thematic Analysis in Qualitative Research: Using a*
Practical Example, London, SAGE.


Tokyo University of Science (n.d.) *Center for Faculty Development* [Online]. Available at https://oae.tus.ac.jp/fd/node/384 (Accessed 5 June 2017).


Tokyo University of Science (2019a) *FE-1 Syllabus* [Online]. Available at: https://letus.ed.tus.ac.jp/2019/course/info.php?id=104416

Tokyo University of Science (2019b) *ISE-1 Syllabus* [Online]. Available at: https://letus.ed.tus.ac.jp/2019/course/info.php?id=105161


Appendices

Appendix 1: Cover page of the document titled English Language Proficiency Strategies for Students

Cover page of the document titled Tokyo Rika Daigakku Noda Chiku ni okeru Gakusei no Eigo Nouryoku Koujou no Housaku ni Tsuite (English Language Proficiency Strategies for Students in the District)

Appendix 2: Section of the above document
Appendix 3: Detail of the first two paragraphs of the section in Appendix 2.

Below is a translation of the text, provided by my colleague Stephen Jennings:

In recent years, it has become clear that the age of Globalisation has made us rethink educational policy and what is best for our students. In this regard, it is remains clear that English communications skills are the basis of international relations. Moreover, these skills are becoming a basic necessity for job seekers and for those hoping to pass post-graduate school exams. It is within this environment that we find ourselves with a need to strengthen English education, with particular emphasis on communication skills.

However, it may be fair to say that students in our university lack English skills. In fact, according to English teachers, students lack skills in writing and listening. Students also lack the reading skills and the presentation skills necessary when attending international conferences - vital at the post-graduate level. Science teachers also feel this keenly on a regular basis.
Appendix 4: Cover page of the TUS 2014 Syllabus Production Guidelines Handbook
This image shows the front page of the 2014 (Syllabus Sakusei Youryou) Syllabus Production Guidelines Handbook use at Tokyo Rikadai.

This image shows page 6 of the Syllabus Guideline Handbook which was circulated on November 6 of 2014. Circled on this page are two features to be included in all department’s curriculum designs: logical thinking and ethics.
Appendix 6: Core curriculum

Pictured here is the entire curriculum for English courses. The four boxes at the top represent the required courses, Freshman English and the Integrated Skills (ISE) courses ISE-A, ISE-B, and ISE-C.
Appendix 7: Page One of the syllabus for a First-Year English course
## Appendix 8: Core course description (Tokyo University of Science, 2016)

<table>
<thead>
<tr>
<th>1st Year - semester 1</th>
<th>1st Year - semester 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman English (Small)</td>
<td>Integrated Skills in English A</td>
</tr>
<tr>
<td>6–8 Classes per department</td>
<td>8–9 classes (2 departments per class)</td>
</tr>
<tr>
<td>2 credits</td>
<td>1 credit</td>
</tr>
<tr>
<td>Reading, Writing, Listening and Speaking skills development with a focus on Listening and Speaking. Students will become better able to actively use latent vocabulary in discussion; and writing at the sentence level.</td>
<td>Reading, Writing, Listening and Speaking skills development with a focus on discussion and paragraph writing. Students will exchange reasoned opinions on a variety of topics connected with society.</td>
</tr>
<tr>
<td>Class placement test: Early April</td>
<td>Class placement test: Mid-December</td>
</tr>
<tr>
<td>In-class test: (Small: 40% Joint: 50%) &amp; Placement Test: 10%</td>
<td>In-class assessment: 90% &amp; Placement Test: 10%</td>
</tr>
<tr>
<td>Chosen from the book list</td>
<td>Chosen from the book list</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2nd Year - semester 1</th>
<th>2nd Year - semester 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated Skills in English B</td>
<td>Integrated Skills in English C</td>
</tr>
<tr>
<td>8–9 classes (2 departments per class)</td>
<td>8–9 classes (2 departments per class)</td>
</tr>
<tr>
<td>1 credit</td>
<td>1 credit</td>
</tr>
<tr>
<td>Reading, Writing, Listening and Speaking skills development with a focus on holding discussions at paragraph length on a variety of topics. Discussions should consist of introduction/ body/ conclusion and be the basis for writing and presentation skills.</td>
<td>Reading, Writing, Listening and Speaking skills development with a focus on reacting to the opinions of others. Students will express opinions using appropriate phrases, orally and in writing, that make clear their reasons for agreeing or disagreeing.</td>
</tr>
<tr>
<td>Class placement test: Early August</td>
<td>Ability appraisal test: Mid-January</td>
</tr>
<tr>
<td>In-class assessment: 90% &amp; Placement Test: 10%</td>
<td>In-class assessment: 90% &amp; Ability appraisal Test: 10%</td>
</tr>
<tr>
<td>Chosen from the book list</td>
<td>Chosen from the book list</td>
</tr>
</tbody>
</table>
Appendix 9: TUS webpage with mention of critical thinking

The TUS website page pictured here describes the education goals. The two circled words here, including the phrase *hihanteki*, in pink and green highlighting, refer to the development of the *strength* of students’ critical thinking skills. (Tokyo University of Science, n.d.)
Appendix 10: TUS webpage on topic of active learning (Tokyo University of Science, n.d.)
Appendix:  
Metadiscourse items investigated

These are the search items used in this book as potentially realizing metadiscourse functions. It must be remembered, of course, that all items can realize either propositional or metadiscoursal meanings and that many can express either interactive or interpersonal meanings. Every instance should therefore be studied in its sentential co-text.

**Interactive Metadiscourse**

*Code Glosses*

- ( )
- as a matter of fact
- called
- defined as
- e.g.
- for example
- for instance
- I mean
- i.e.
- in fact
- in other words
- indeed
- known as
- namely
- or X
- put another way
- say
- specifically
- such as
- that is

*that is to say*

*that means*

*this means*

*viz*

*which means*

**Endophoric Markers**

(In) Chapter X
(In) Part X
(In) Section X
(In) the X chapter
(In) the X part
(In) the X section
(In) This chapter
(In) This part
(In) This section

Example X
Fig. X
Figure X
P. X

Appendix 11: Hyland (2005a) metadiscourse marker list
Appendix 1: Hyland (2005b) metadiscourse marker list (cont.)

b) label stages
- all in all
- at this point
- at this stage
- by far
- for the moment
- in brief
- in conclusion
- in short
- in sum
- in summary
- now
- on the whole
- overall
- so far
- thus far
- to conclude
- to repeat
- to sum up
- to summarize

c) announce goals
- (in) this chapter
- (in) this part
- (in) this section
- aim
- desire to
- focus
- goal
- intend to
- intention
- objective
- purpose
- seek to
- want to
- wish to
- would like to
Appendix 11: Hyland (2005b) metadiscourse marker list (cont.)
Appendix 11: Hyland (2005b) metadiscourse marker list (cont.)

<table>
<thead>
<tr>
<th>Desirably</th>
<th>Unexpectedly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disappointed</td>
<td>Unexpectedly</td>
</tr>
<tr>
<td>Disappointing</td>
<td>Unfortunately</td>
</tr>
<tr>
<td>Disappointingly</td>
<td>Unusually</td>
</tr>
<tr>
<td>Disagree</td>
<td>Usually</td>
</tr>
<tr>
<td>Disagreed</td>
<td></td>
</tr>
<tr>
<td>Disagrees</td>
<td></td>
</tr>
<tr>
<td>Dramatic</td>
<td></td>
</tr>
<tr>
<td>Dramatically</td>
<td></td>
</tr>
<tr>
<td>Essential</td>
<td></td>
</tr>
<tr>
<td>Essentially</td>
<td></td>
</tr>
<tr>
<td>Even x</td>
<td></td>
</tr>
<tr>
<td>Expected</td>
<td></td>
</tr>
<tr>
<td>Expectedly</td>
<td></td>
</tr>
<tr>
<td>Fortunate</td>
<td></td>
</tr>
<tr>
<td>Fortunately</td>
<td></td>
</tr>
<tr>
<td>Hopeful</td>
<td></td>
</tr>
<tr>
<td>Hopefully</td>
<td></td>
</tr>
<tr>
<td>Important</td>
<td></td>
</tr>
<tr>
<td>Importantly</td>
<td></td>
</tr>
<tr>
<td>Inappropriate</td>
<td></td>
</tr>
<tr>
<td>Inappropriately</td>
<td></td>
</tr>
<tr>
<td>Interesting</td>
<td></td>
</tr>
<tr>
<td>Interestingly</td>
<td></td>
</tr>
<tr>
<td>Prefer:</td>
<td></td>
</tr>
<tr>
<td>Preferable</td>
<td></td>
</tr>
<tr>
<td>Preferably</td>
<td></td>
</tr>
<tr>
<td>Preferred</td>
<td></td>
</tr>
<tr>
<td>Remarkable</td>
<td></td>
</tr>
<tr>
<td>Remarkably</td>
<td></td>
</tr>
<tr>
<td>Shocked</td>
<td></td>
</tr>
<tr>
<td>Shocking</td>
<td></td>
</tr>
<tr>
<td>Shockingly</td>
<td></td>
</tr>
<tr>
<td>Striking</td>
<td></td>
</tr>
<tr>
<td>Strikingly</td>
<td></td>
</tr>
<tr>
<td>Surprised</td>
<td></td>
</tr>
<tr>
<td>Surprising</td>
<td></td>
</tr>
<tr>
<td>Surprisingly</td>
<td></td>
</tr>
<tr>
<td>Unbelievable</td>
<td></td>
</tr>
<tr>
<td>Unbelievably</td>
<td></td>
</tr>
<tr>
<td>Understandable</td>
<td></td>
</tr>
<tr>
<td>Understandably</td>
<td></td>
</tr>
<tr>
<td>Unexpected</td>
<td></td>
</tr>
<tr>
<td>Known</td>
<td></td>
</tr>
<tr>
<td>Known</td>
<td></td>
</tr>
<tr>
<td>must (<em>possibility</em>)</td>
<td>Engagement Markers</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>never</td>
<td>(</td>
</tr>
<tr>
<td>no doubt</td>
<td>?</td>
</tr>
<tr>
<td>obvious</td>
<td>(the) reader's</td>
</tr>
<tr>
<td>obviously</td>
<td>add</td>
</tr>
<tr>
<td>of course</td>
<td>allow</td>
</tr>
<tr>
<td>prove</td>
<td>analyse</td>
</tr>
<tr>
<td>proved</td>
<td>apply</td>
</tr>
<tr>
<td>proves</td>
<td>arrange</td>
</tr>
<tr>
<td>realize</td>
<td>assess</td>
</tr>
<tr>
<td>realized</td>
<td>assume</td>
</tr>
<tr>
<td>realizes</td>
<td>by the way</td>
</tr>
<tr>
<td>really</td>
<td>calculate</td>
</tr>
<tr>
<td>show</td>
<td>choose</td>
</tr>
<tr>
<td>showed</td>
<td>classify</td>
</tr>
<tr>
<td>shown</td>
<td>compare</td>
</tr>
<tr>
<td>shows</td>
<td>connect</td>
</tr>
<tr>
<td>sure</td>
<td>consider</td>
</tr>
<tr>
<td>surely</td>
<td>consult</td>
</tr>
<tr>
<td>think</td>
<td>contrast</td>
</tr>
<tr>
<td>thinks</td>
<td>define</td>
</tr>
<tr>
<td>thought</td>
<td>demonstrate</td>
</tr>
<tr>
<td>truly</td>
<td>determine</td>
</tr>
<tr>
<td>true</td>
<td>do not</td>
</tr>
<tr>
<td>undeniable</td>
<td>develop</td>
</tr>
<tr>
<td>undeniably</td>
<td>employ</td>
</tr>
<tr>
<td>undisputedly</td>
<td>ensure</td>
</tr>
<tr>
<td>undoubtedly</td>
<td>estimate</td>
</tr>
<tr>
<td>without doubt</td>
<td>evaluate</td>
</tr>
</tbody>
</table>

### Self Mention

- I
- we
- me
- my
- our
- mine
- us
- the author
- the author's
- the writer
- the writer's

---

Appendix 11: Hyland (2005b) metadiscourse marker list (cont.)
| look at | apparent |
| mark | apparently |
| measure | appear |
| mount | appeared |
| must | appears |
| need to | approximately |
| note | argue |
| notice | argued |
| observe | argues |
| one's | around |
| order | assume |
| ought | assumed |
| our (inclusive) | broadly |
| pay | certain amount |
| picture | certain extent |
| prepare | certain level |
| recall | claim |
| recover | claimed |
| refer | claims |
| regard | could |
| remember | couldn't |
| remove | doubt |
| review | doubtful |
| see | essentially |
| select | estimate |
| set | estimated |
| should | fairly |
| show | feel |
| suppose | feels |
| state | felt |
| take (a look/as example) | frequently |
| think about | from my perspective |
| think of | from our perspective |
| turn | from this perspective |
| us (inclusive) | generally |
| use | guess |
| we (inclusive) | indicate |
| you | indicated |
| your | indicates |
| Hedges | in general |
| about | in most cases |
| almost | in most instances |
| | in my opinion |

Appendix 11: Hyland (2005b) metadiscourse marker list (cont.)
in my view relatively
in this view roughly
in our opinion seems
in our view should
largely sometimes
likely somewhat
mainly suggest
may suggested
maybe suggests
might suppose
mostly supposed
often supposes
on the whole suspect
ought suspects
perhaps tend to
tend to
tended to
tends to
plausible
to my knowledge
plausibly typically
possible typical
possibly uncertain
postulate uncertainly
postulates unclearly
postulated
typical
postulates un lith ly
presumable
presumably
unlikely
probable usually
probably
quite would
rather x wouldn’t
Table 7-10: Categories of enhancement and principal markers

<table>
<thead>
<tr>
<th>Category</th>
<th>Meaning</th>
<th>Paratactic</th>
<th>Hypotactic</th>
<th>non-finite: conjunction</th>
<th>non-finite: preposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) temporal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>same time</td>
<td>A meanwhile B</td>
<td>(and) meanwhile; (when)</td>
<td>[extent] as, while while in (the course/process of)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[point] when, as soon as, the moment when on</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[spread] whenever, every time – –</td>
<td></td>
<td></td>
</tr>
<tr>
<td>different time: later</td>
<td>A subsequently B</td>
<td>(and) then; and + afterwards</td>
<td>after, since since after</td>
<td></td>
<td></td>
</tr>
<tr>
<td>different time: earlier</td>
<td>A previously B</td>
<td>and/ but + before that/ first</td>
<td>before, until/ till until before</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(ii) spatial</td>
<td></td>
<td>C there D</td>
<td>and there</td>
<td>[extent] as far as – –</td>
<td></td>
</tr>
<tr>
<td>same place</td>
<td></td>
<td></td>
<td>[point] where – –</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[spread] wherever, everywhere – –</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iii) manner</td>
<td></td>
<td>N is via/ by means of M</td>
<td>and + in that way; (and) thus – –</td>
<td>by (means of)</td>
<td></td>
</tr>
<tr>
<td>means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comparison</td>
<td>N is like M</td>
<td>and + similarly; (and) so, thus</td>
<td>as, as if, like, the way like</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(iv) causal-conditional</td>
<td>because P so result Q</td>
<td>(cause ^ effect) (and) so; and + therefore</td>
<td>[effect ^ cause] for; (because) because, as, since, in case, seeing that, considering with, through, by at, as a result, because of, in case of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cause: reason</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cause: purpose</td>
<td>because intention Q so action P</td>
<td>–</td>
<td>in order that, so that –</td>
<td>(in order/ so as) to; for (the sake of), with the aim of, for fear of</td>
<td></td>
</tr>
<tr>
<td>(v) causal-conditional</td>
<td>cause: result</td>
<td></td>
<td>so that – –</td>
<td></td>
<td></td>
</tr>
<tr>
<td>condition: positive</td>
<td>if P then Q</td>
<td>(and) then; and + in that case</td>
<td>if, provided that, as long as if</td>
<td>in the event of</td>
<td></td>
</tr>
<tr>
<td>condition: negative</td>
<td>if not P then Q</td>
<td>or else; (or) otherwise</td>
<td>unless unless</td>
<td>but for, without</td>
<td></td>
</tr>
</tbody>
</table>
## Set I

### I. Premise and Conclusion Indicators

#### Premise Indicators

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Reference Example</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Because</td>
<td>Because</td>
<td>Reason why</td>
</tr>
<tr>
<td>Since</td>
<td>Since</td>
<td>Given the fact that</td>
</tr>
<tr>
<td>In addition</td>
<td>Furthermore</td>
<td>Not only that, but</td>
</tr>
<tr>
<td>Furthermore</td>
<td>Most importantly</td>
<td>The reason why</td>
</tr>
<tr>
<td>Considering the fact that</td>
<td></td>
<td>Given the fact that</td>
</tr>
<tr>
<td>Not only that, but</td>
<td></td>
<td>We have noticed that</td>
</tr>
<tr>
<td>Next</td>
<td>Then</td>
<td>He/She noted that</td>
</tr>
<tr>
<td>Lastly</td>
<td>Finally</td>
<td>We know that</td>
</tr>
<tr>
<td>It appears that</td>
<td>In other words</td>
<td>It is common knowledge</td>
</tr>
<tr>
<td></td>
<td>The data reveals that</td>
<td>While it is true that</td>
</tr>
<tr>
<td>hinges upon</td>
<td>expand on</td>
<td></td>
</tr>
</tbody>
</table>

#### Conclusion indicators

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Reference Example</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>So,</td>
<td>Therefore,</td>
<td>I think/ feel / believe</td>
</tr>
<tr>
<td>Therefore,</td>
<td></td>
<td>indicates that</td>
</tr>
<tr>
<td>Hence,</td>
<td>Consequently</td>
<td>There is no reason to believe that</td>
</tr>
<tr>
<td>As a result,</td>
<td></td>
<td>We must conclude that</td>
</tr>
<tr>
<td>Anyway,</td>
<td>That’s why I believe</td>
<td>Because of this,</td>
</tr>
<tr>
<td></td>
<td>For this reason,</td>
<td>It’s my opinion that</td>
</tr>
<tr>
<td></td>
<td>It’s my opinion that</td>
<td></td>
</tr>
<tr>
<td>In short</td>
<td>Summing up</td>
<td>It may be argued,</td>
</tr>
<tr>
<td></td>
<td>It’s my opinion that</td>
<td>points to</td>
</tr>
<tr>
<td>In brief</td>
<td>In this way,</td>
<td>It makes good sense</td>
</tr>
<tr>
<td></td>
<td>We can see that</td>
<td>We have decided to</td>
</tr>
<tr>
<td>In this way,</td>
<td>It is clear that</td>
<td>As far as I’m concerned</td>
</tr>
<tr>
<td></td>
<td>In the final analysis,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>As for myself,</td>
<td>It looks like x is the case/ happened.</td>
</tr>
</tbody>
</table>

#### Additional Premises/Reasons

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Reference Example</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>And</td>
<td>besides,</td>
<td></td>
</tr>
<tr>
<td>Also,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In addition,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Another thing is</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not to mention the fact that</td>
<td>Plus the fact that</td>
<td></td>
</tr>
<tr>
<td>Furthermore,</td>
<td>Moreover,</td>
<td>To add to the counterargument,</td>
</tr>
<tr>
<td>Not only that, but</td>
<td>Coupled with x</td>
<td></td>
</tr>
<tr>
<td>At the same time,</td>
<td>As well as x</td>
<td></td>
</tr>
<tr>
<td></td>
<td>is just one more</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What’s more,</td>
<td></td>
</tr>
</tbody>
</table>
IV. Creating Focus
Pressing the Matter

<table>
<thead>
<tr>
<th>Above all,</th>
<th>That’s just the point.</th>
<th>particularly</th>
<th>especially</th>
</tr>
</thead>
<tbody>
<tr>
<td>chiefly,</td>
<td>You just don’t get it, do you?</td>
<td>The real question is</td>
<td>Can’t you see that . . .?</td>
</tr>
<tr>
<td>Mark my words,</td>
<td></td>
<td>. . . and I repeat</td>
<td></td>
</tr>
</tbody>
</table>

V. Interruption/Interjection, Resuming a Previous Topic, Changing the Subject
Interrupting

<table>
<thead>
<tr>
<th>Pardon/excuse me, but.</th>
<th>Sorry/Excuse me for interrupting, but</th>
<th>May I ask a question?</th>
</tr>
</thead>
<tbody>
<tr>
<td>May I add something?</td>
<td>I’d like to comment on that</td>
<td>Could I just jump in here?</td>
</tr>
<tr>
<td>Can I add that</td>
<td>I’d like to say something here</td>
<td>What about . . .?</td>
</tr>
</tbody>
</table>

Resuming discussion

<table>
<thead>
<tr>
<th>In any case,</th>
<th>To return to X,</th>
<th>To get back to X,</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anyway,</td>
<td>Where was I?</td>
<td>One last point,</td>
</tr>
<tr>
<td>To get back to what I was saying,</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Changing the subject

| By the way,            | That reminds me,                    | It’s a little off the subject, but |

Appendix 13: Master MDM List (cont.)
Set III

**VI. Validity and Soundness**

<table>
<thead>
<tr>
<th>That’s a false claim.</th>
<th>That is a valid point.</th>
<th>That point is invalid, because</th>
</tr>
</thead>
<tbody>
<tr>
<td>That’s unsound because</td>
<td>You have a sound point.</td>
<td></td>
</tr>
</tbody>
</table>

**VII. Clarification and Confirmation**

**SELF-CLARIFYING**

<table>
<thead>
<tr>
<th>What I mean is</th>
<th>Let me put it another way.</th>
<th>What I’m saying is</th>
</tr>
</thead>
<tbody>
<tr>
<td>What I’m trying to say is</td>
<td>In other words,</td>
<td>I didn’t mean to say that.</td>
</tr>
</tbody>
</table>

**CLARIFYING OTHERS and REQUESTING CLARIFICATION**

<table>
<thead>
<tr>
<th>What you mean then is</th>
<th>Could you explain that in more detail?</th>
<th>What you’re saying is</th>
</tr>
</thead>
<tbody>
<tr>
<td>I don't see what you mean.</td>
<td>What do you mean by . . . ?</td>
<td>Are you sure?</td>
</tr>
<tr>
<td>Could you be more specific?</td>
<td>If I understand you correctly,</td>
<td>Do you mean . . . ?</td>
</tr>
<tr>
<td>So, are you saying that . . . ?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CONFIRMING UNDERSTANDING**

<table>
<thead>
<tr>
<th>Are you following me?</th>
<th>Are you still with me?</th>
<th>Have you got it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you follow?</td>
<td>Okay so far?</td>
<td>Is that clear?</td>
</tr>
</tbody>
</table>

**EXPressing LACK OF UNDERSTANDING**

| Sorry, you’ve lost me. | I don’t follow you. | I’m not sure what you mean. | I don’t get it. |

**VIII. Accusations and Insinuations // Defenses and Explanations**

**Accusations**

<table>
<thead>
<tr>
<th>You led us to believe that . . .</th>
<th>You said that “xxxxx”.</th>
<th>You’re to blame for . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>It’s because of you / your decision to xxx that we’re having this problem / in this mess.</td>
<td>You’re the one who wanted to . . .</td>
<td>It’s your fault that . . .</td>
</tr>
<tr>
<td>You brought this on yourself.</td>
<td>The reason for (x happening) is that you (did y).</td>
<td>You’re responsible for . . .</td>
</tr>
</tbody>
</table>
Set IV

X. Examples and Illustrations/Analogies

<table>
<thead>
<tr>
<th>For example,</th>
<th>Take ______, for instance.</th>
<th>Look at the way</th>
<th>To illustrate the point,</th>
</tr>
</thead>
</table>

XI. Circumlocution

It is very difficult for learner to say or write something when they do not know the correct word (or can’t remember the word) that they want to use.

- Try to think of another way to convey the message. Describe the concept.
- Explain who uses it, why it’s used, or where. Use a synonym.
- Explain what the thing is and what the thing is not.
- Point and use gestures, actions, and sounds.

Example

**Mark:** Do you have a corkscrew?

**Yoshi:** A what?

**Mark:** Oh, … you know… it’s one of those things, um, to get the … cork (while gesturing) out.

**Yoshi:** Oh, a wine-opener, sure, there’s one in the bottom drawer over there.

Appendix 13: Master MDM List (cont.)
Set V

XII. Conditional Statements and Hypothetical Premise Indicators

<table>
<thead>
<tr>
<th>If x, then y.</th>
<th>If I do x, then y will happen.</th>
</tr>
</thead>
<tbody>
<tr>
<td>If x happens, then I will do y.</td>
<td>Whenever x, then y.</td>
</tr>
<tr>
<td>Unless we do x, then y will happen.</td>
<td>Unless x happens, we should continue doing y.</td>
</tr>
</tbody>
</table>

XIII. Side-by-Side Premises and Chain Premise Arguments

Side-by-Side Premises These are usually indicators like:
- Also
- In addition
- Furthermore
- Moreover

Chain Premise Arguments These usually use transitional phrases which use the conclusion of the preceding argument as the first premise for a subsequent connected argument. They therefore use many of the same conclusion indictors as listed in Set I, Premise Indicators. These might also include:
- Clearly then,
- Following this further,
- In light of x
- It is easy to see then that

Lastly, they can also be expressed as hypothetical If x, then y premises discussed in the Critically Minded Podcast Episode 6 and 6.2.

XIV. Suggestions

Making Suggestions

<table>
<thead>
<tr>
<th>Let’s . . .</th>
<th>Don’t you think . . .?</th>
<th>I suggest that we . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>May I make a suggestion?</td>
<td>Why don’t you try . . .?</td>
<td>You might want to consider . . .</td>
</tr>
<tr>
<td>Wouldn’t it be better if . . .?</td>
<td>Are you sure . . .?</td>
<td>Have you thought about . . .?</td>
</tr>
<tr>
<td>Perhaps we could</td>
<td>Why not (do) . . .?</td>
<td>How about . . .?</td>
</tr>
<tr>
<td>I propose that we</td>
<td>I have an idea.</td>
<td>Why don’t we . . .?</td>
</tr>
<tr>
<td>Please consider (doing)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inviting Responses/Suggestions and Rejecting Suggestions

INVITING

<table>
<thead>
<tr>
<th>Do you have any suggestions?</th>
<th>What should we do?</th>
<th>What would you do?</th>
</tr>
</thead>
</table>

Appendix 13: Master MDM List (cont.)
Set VI

XV. Pointing Out Commonly Held False Premises

<table>
<thead>
<tr>
<th>Many people think x, but in fact</th>
<th>Some people say that x, but actually</th>
</tr>
</thead>
<tbody>
<tr>
<td>It may seem that x, but</td>
<td>A commonly held belief is x, but</td>
</tr>
<tr>
<td>At first glance it looks like x, but in reality</td>
<td>We take it for granted that . . . , but the truth of the matter is</td>
</tr>
<tr>
<td>I feel I must point out an error.</td>
<td>I suspect you are thinking about x, which was</td>
</tr>
</tbody>
</table>

XVI. Rules and Exceptions to the Rule

Generalizing

<table>
<thead>
<tr>
<th>In general, most</th>
<th>Generally,</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a rule,</td>
<td>By and large,</td>
</tr>
<tr>
<td>In most cases,</td>
<td>Most of the time,</td>
</tr>
</tbody>
</table>

Exceptions

| There are exceptions, of course. For example, | One exception is |

XVII. Major and Minor Premise Indicators

Part of Speech

| adjectives | verb phrases used as nouns, e.g. sleeping dogs (as contrasted with simply all dogs). |

Major Premise

<table>
<thead>
<tr>
<th>all</th>
<th>only</th>
<th>most</th>
<th>always</th>
<th>never</th>
<th>everybody</th>
<th>nobody</th>
<th>completely</th>
<th>totally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Either p is q or x is y.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Minor Premise

<table>
<thead>
<tr>
<th>the</th>
<th>this</th>
<th>that</th>
<th>these</th>
<th>those</th>
<th>I</th>
<th>he</th>
<th>she</th>
<th>it</th>
<th>his</th>
<th>her</th>
<th>my</th>
</tr>
</thead>
<tbody>
<tr>
<td>now</td>
<td>sometimes</td>
<td>excepting</td>
<td>except for</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is useful to think of major and minor premise arguments as often being similar to chain premise arguments (see Critically Minded Podcast Episode 3 pdf. Set IV). Where as chain arguments often connect premises causally (i.e. cause and effect), major and minor premise arguments are concerned with definition and how different persons, objects, concepts and events fit into sets and subsets.

Appendix 13: Master MDM List (cont.)
### Set VII

#### XVIII. Issues Indicators

<table>
<thead>
<tr>
<th><strong>Descriptive</strong></th>
<th>is...</th>
<th>isn’t...</th>
<th>was...</th>
<th>wasn’t...</th>
<th>do...</th>
<th>doesn’t...</th>
<th>What?</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th><strong>Prescriptive</strong></th>
<th>should</th>
<th>ought to</th>
<th>had better</th>
<th>must</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mustn’t</td>
<td>need(s) to</td>
<td>have to / has to</td>
<td></td>
</tr>
</tbody>
</table>

**(Normative)**

<table>
<thead>
<tr>
<th>good</th>
<th>better</th>
<th>beat</th>
<th>had</th>
<th>worse</th>
<th>worst</th>
<th>too much</th>
<th>not enough</th>
</tr>
</thead>
</table>

Appendix 13: Master MDM List (cont.)
Facility of Wellbeing, Education & Language Studies

Consent form for persons participating in a research project

Title: Investigating the Efficacy of Online Text Reconstruction Exercises for Facilitating the Use of Linguistic Features of Written Argumentation

Name of participant:

Name of principal investigator(s): David A. Gann

This purpose of this project is to find out about how the online English exercises we do in this class help you students to write better academic arguments in English. I hope to learn about the changes in your writing skills and about your personal experiences using the online resources.

1. I consent to participate in this project, the details of which have been explained to me, and I have been provided with a written statement in plain language to keep.

2. I understand that my participation will take place in a Tokyo University of Science Noda Campus computer lab and will involve:
   a. data anonymised by using numbers in place of participants’ names
   b. language learning activities in the form of multiple choice selection during online collaborative exercises identical to other classmates’. These will be completed alongside other classmates during the usual scheduled lessons and will involve:
      i. participation with two other classmates working together on a single computer
      ii. audio recording of participants’ talk (in Japanese and/or English) while completing the exercises
      iii. video recording of computer screen activity made during the completion of the exercise, and
      iv. digital storage of participants’ answer selections
   c. out-of-class discussion with classmates of textbook fronted issues in an online university forum
   d. in-class discussion with other classmates of textbook fronted issues in an online university chatroom
   e. interviews at the end of the term, which will:
      i. be conducted on campus
      ii. be conducted individually
      iii. involve some questions about what you experienced during the exercises and if or how these exercises reinforced what you learned earlier; and if or how they influenced your later language use in the forum and chatroom
      iv. last a maximum of 20 minutes

3. I acknowledge that:
   a. the possible effects of participating in this research have been explained to me.
   b. I have been informed that I am free to withdraw from the project without explanation or prejudice and to request the destruction of any data that have been gathered from me until it is anonymized at the point of transcription on 9 January, 2018. After this point data will have been processed and it will not be possible to withdraw any unprocessed data I have provided.

May 2017

http://www.open.ac.uk/research/ethics/human-research

Appendix 14: Participant consent form, pages 1.
c. the project is for the purpose of research.

d. I have been informed that the confidentiality of the information I provide will be safeguarded subject to any legal requirements.

e. I have been informed that with my consent the data generated will be stored in a password-protected file and will be destroyed after 3 years.

f. I have been informed that anonymized research data may be made available to other members of the research community for the purpose of extending the scope of this research inquiry for a period of 3 years.

g. If necessary, any data from me will be referred to by a pseudonym in any publications arising from the research.

h. I have been given contact details for a person whom I can contact if I have any concerns about the way in which this research project is being conducted.

i. I have been informed that a summary copy of the research findings will be forwarded to me, should I request this.

I consent to the activity during the text reconstruction exercises being audio-taped and on-screen cursor movement and answer selection being recorded □ yes □ no (please delete as appropriate) (please tick)

I wish to receive a copy of the summary project report on research findings □ yes □ no (please tick)

Email or postal address to which a summary should be sent:

Participant signature: ___________________________ Date: ________________

Contact details for the Principal Investigator (PI) and Research organisation and Faculty:

David A. Gann: email: dagann@rsi.us.ac.jp

Contact details for an alternative contact if you have any concerns about the way the research project is being conducted:

Dr. Prithvi Shrestha: prithvi.shrestha@open.ac.uk

This research has been reviewed by, and received a favourable opinion, from the OU Human Research Ethics Committee - HREC reference number: __________ (http://www.open.ac.uk/research/ethics/).

May 2017

http://www.open.ac.uk/research/ethics/human-research

Appendix 14: Participant consent form, page 2.
Appendix 15: An example of participants’ text chat in a chat room
<table>
<thead>
<tr>
<th>Interview question</th>
<th>Research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. これらのエクササイズ中にあなたが勉強したことは、あなたに新しい言語または、すでに議論の言語について知っていましたか？</td>
<td>1. To what extent does completing the OTREs develop metalinguistic awareness?</td>
</tr>
<tr>
<td>2. このコースの前に、学術的な文章の中で議論の言語がどれくらい重要だと思いましたか？それは変わりましたか？また、どのように？</td>
<td></td>
</tr>
<tr>
<td>3. 演習の目的について、あなたの意見は何ですか？</td>
<td></td>
</tr>
<tr>
<td>4. 演習の焦点はどのようなものでしたか？</td>
<td></td>
</tr>
<tr>
<td>5. その後、フォーラムとチャットルームで、演習中に勉強した言葉を使用しましたか？言葉換えれば、実際の学習はありませんか？</td>
<td></td>
</tr>
<tr>
<td>6. グループ活動を行った結果、あなたの学習成果はどれくらいだったのですか？そして、サンプルテキストを読んだ結果、どれくらいの学習成果でしたか？</td>
<td>To what extent does completing the OTREs guide language learners to use selected MDMs in written argument?</td>
</tr>
<tr>
<td>7. あなたの学習は、多肢選択式質問と演習のシーケンシャルなデザインの結果でしたか？あなたのグループのメンバーとの社会的な状況で活動をした結果、どのくらいの学習をしたのですか？</td>
<td></td>
</tr>
<tr>
<td>8. これらの演習では、目標言語に注意を向けるのに効果的な文章がありましたか？</td>
<td></td>
</tr>
<tr>
<td>9. あなたの将来の他の授業で、書く時や論文のように学術論文を書く時に、これらの言語項目を使う可能性はどれくらいありますか？</td>
<td>What, if any, are the likely longer term benefits of these exercises?</td>
</tr>
</tbody>
</table>

Appendix 16: Japanese translation of interview questions
The following is an example of a typical interview that was conducted in this study. There was a pre-written explanation of the purpose and procedure of the interview after which it continued thus:

P2: Is it about CT?
Gann: Well yes, about CT and specifically about the language of argument.
P2: I’ve heard the word CT but I didn’t know what about CT.
Gann: And how about the language of argument?
P2: I knew a little, but the experience to—the sheet, the pdf sheet. I did not have the kind of sheet. ( )
Gann: What did you already know. Can you tell me some examples of things that you already knew, before the class?
P2: I think . .
Gann: Okay?
P2: I think
Gann: G&O (laugh)
P2: I knew . . .
Gann: Of course.
P2: Mochiron [of course]
Gann: Anything else, hoka ni [anything else]?
P2: Mmm. I cannot agree with you anymore.
Gann: Okay. (laughs) I couldn’t agree with you more. And of course, you knew words like so--
P2: Yeah. I . . . It’s my opinion that . .
Gann: Mmmm.
P2: I like this. (laughs)
Gann: So you had some experience with this kind of language. little bit anyway. Alright question number two then: [READ]
P2: Mmmm (not understanding, at first)
Gann: So you have academic writing and inside that, one part of that is argument. Before this course, how important did you think argument would be in academic writing?
P2: I have not opened (begun) my report, but I think it is important, but before this class, argument was [not so important]. Gann: Has that changed?
P2: Yes, yes.
Gann: How so?
P2: Before this study I did not know academic argument words . . . so I can take it. (So I can get it; or internalise the phrases.)
Gann: [READ #3]
P2: To use languages we learned at home, and teichaku—
Gann: Teichaku?
P2: *Teichaku*. To take in.
Gann: Oh I see, *tei*, set or decide, and *chaku*, to pull or take in. Thank you. [I demonstrate an explicit awareness of equality and meaning co-construction.] [READ #4] Now I don’t mean vocabulary. I mean the category of communication. [These are expressions I have often used in class.]
P2: To use in argument?
Gann: Can you tell me two or three or the categories we learned about?
P2: We learned about—to explain how I agree.
Gann: Agreement, very good.
P2: To get to know, I start next talk, *so, then*. (changing the subject)
Gann: *So, then* . . .
P2: Etcetera.
Gann: [READ #5]
P2: Yes, I used.
Gann: You used it?
P2: Yes.
Gann: Where did you use it?
P2: We discussed online. So I used, agree(ment) (and other categories) with the members in the forum and chatroom.
Gann: Anywhere else? We had the forum and chatroom and—
P2: The Final.
Gann: Yes, the final report.
P2: Yes, I used, of course I use.
Gann: [READ #6] How much of your learning . . .? When you did the WebSequiturs you had the example text up here (pointing at a prepared example) and you read this first and you had the exercise below. How much of your learning was doing the WebSequiturs exercise and how much or your learning was just reading the text?
P2: Mmm. I learned a lot by exercises. I can use my brain more hareshiku (pleasantly) in exercise.
Gann: Let’s say for example, I didn’t give you the WebSequiturs exercises, and I just gave you the sample texts and I said “Here, read these and try to write the same kind of writing yourself.” Would that be an effective way to teach? What would happen?
P2: If you just give me paper, I . . . I might not read it.
Gann & P2: (laugh)
Gann: Now, I have another similar question. [READ #7]
P2: I can tell, different type of questions but that was good to use my brain. Gann: Can we go back. I didn’t understand, “I can tell different type—”
P2: I can’t tell different kind of, kind of, tell from . . .
Gann: Oh, tell from. Distinguish between.
P2: Yes, I cannot tell between different types of questions. So, I cannot say what type of question is good, but that was good to use my brain.
Gann: So the next question is about how of your learning was a result of working in a social situation? I mean I could give the WebSequiturs individually and say, okay everybody, do your own work. Don’t look at other people’s screens. But I put you into groups. How did that affect your learning experience?
P2: I did almost all the questions.
Gann: You were the most talkative of the group?
P2: So, I almost individual.
Gann & P2: (laugh)
Gann: So it was almost the same as individual. Do you think it was helpful to your other two group members to work with you?
P2: I tried to get [them] to know my idea and they understood—they seemed to understand, so good!
Gann: Okay. [READ #8] (sentence breaks]
P2: Yes, it was. I used grammatical aspects to clear the exercise.
Gann: Sometimes or occasionally.
P2: If in choices there is Capital.
Gann: Capital letters, sure. Other than that, did you use grammar much or did you use your understanding of argument? See there are two things here—there's understanding of argument and there’s grammar. So I’m asking how often did you use your understanding of grammar and how often did you use grammar.
P2: Both.
Gann: Both. Good. [READ #9]
P2: Yes, I'm going to use.
Gann: Do you understand why it is important to use?
P2: Yes, I am not an English speaker, so to have weapons is good to use.
Gann: Alright, thank you. So, to finish up the interview is there anything you would like to tell me about WebSequiturs.
P2: [The student told me about a technical issue that was unrelated to the WebSequitur exercises.]

Appendix 17: An example of a complete interview transcript
Appendix 18: An example of a complete interview transcript with themes in coding process.

O: Yes, it was. I used grammatical aspects to clear the exercise.

Gann: Sometimes or occasionally.

O: If in choices there is Capital.

Gann: Capital letters, sure. Other than that, did you use grammar much or did you use your understanding of argument? See there are two things here—there's understanding of argument and there's grammar. So I'm asking how often did you use your understanding of grammar and how often did you use grammar.

O: Both.

Gann: Both. Good. [READ #9]

O: Yes, I'm going to use.

Gann: Do you understand why it is important to use?

O: Yes, I am not an English speaker, so to have weapons is good to use.

Gann: Alright, thank you. So, to finish up the interview is there anything you would like to tell me about Hot Potatoes.

O: [The student told me about a technical issue that was unrelated to the Web Sequitur exercises.]
<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>An ellipsis denoted a pause</td>
</tr>
<tr>
<td>XXX</td>
<td>A series of Xs indicates that the speech was unintelligible</td>
</tr>
<tr>
<td>[Read #‘#’]</td>
<td>Reads the question from the printout corresponding to the same numbered question in Appendices 14 and 15</td>
</tr>
<tr>
<td>( )</td>
<td>Used to express a gesture or non-verbal utterance</td>
</tr>
<tr>
<td>–</td>
<td>A long dash denotes a pause in the talk.</td>
</tr>
<tr>
<td>[...]</td>
<td>An ellipsis inside brackets indicates talk that was deleted such as stammers, collecting thought in Japanese and short departures, clarifications that did not add to the interview.</td>
</tr>
<tr>
<td><em>italics</em></td>
<td>Italics denote either repeating a word or words that the previous speaker has spoken; or speaking a discourse marker that had received instruction</td>
</tr>
</tbody>
</table>

Appendix 19: Notation system for interview transcriptions
The Three Cups and Coins

Introduction:

We did two experiments with different cases to see how the possibility will change.

Method:

Part one

First, we will have three cups, a die and ¥1,000. Each cup has a number 1 or 2 or 3.

Second, roll a die and decide which cup we should put the ¥1,000 under the cup 1. If 3 or 4, put under cup 2. And, if it’s 5 or 6, put under cup 3. We did this for 30 times and we had three group members so each member did 10 times each. Finally, we will have other group’s member to come and have let the person choose a cup which he/she thinks that the cup is correct. In this experiment, it is not allowed to change the cup to a different one.

Part two

In this experiment, we used a ¥100 coin besides the ¥1,000. Until deciding which
position to put the ¥100 is same as the first experiment. What the difference is when the
person chose the one which he/she thinks that is correct, our group member who is
doing the dealer will show the incorrect cup which the challenger didn’t choose. And
then, the dealer will ask to the challenger “Would you like to stay with your choice or
change it to a different one?” In this experiment, we will think only the case that the
challenger will change the cup.

Discussion:

In part one, in our groups possibility were a little bit off. It was 17/30. However, when
we add up the results with the other groups, the possibility became closer to one-thirds.
There are three cups and only one cup is the correct one, so we thought the possibility
will be one-thirds. It means, this is an experience that when we try more and more times
to see how the possibility will be, it will be more and more closer to the ideal
possibility.

In part two, the result of our group became two-thirds. And, we thought the possibility
will become two-thirds. That is because, in this experiment, there are three cups and
only one is the correct cup. If this is the case, the possibility that the challenger will
choose the correct cup from the beginning is one-thirds. We can tell this from the first
experiment. But in this experiment, the challenger must change the cup. The dealer will

Appendix 20: Rewrite: Rewrite from Memory example 1 (cntd.)
show the incorrect cup which is not chosen to the challenger, so this means if the
challenger has chosen the correct cup from the beginning, the result will become miss,
and if the challenger has chosen the incorrect cup, the result will become hit.

Conclusion:

Again, we could see that when we try more and more times to see how the possibility
will be, it will be more and more closer to the ideal possibility. The possibility in the
first experiment will be one-thirds, and the second experiment, two thirds.

Post Conclusion:

It was very interesting to see that when we try more and more times to see how the
possibility will be, it will be more and more closer to the ideal possibility. It was very
nice to see and check if it is true about the possibility we learned at the high school.
1. The probability is 2/3.
2. If I’m changing the cup, I will lose the money only when I choose the right cup. So I can get the money two times by three times.
3. If I’m changing the cup I choose from beginning I will get the money two times for three times.
4. From the discussion I have, the expectation is 100*2/3=67. So if I’ll pay 66 or less, I’ll get money.
5. First we sat a seat and cups. And we chose a number by a dice. We had decided when the number was 1 or 2, we set a coin under the top cup, when it was 3 or 4, under the center cup, when it was 5 or 6, under the down cup. In the study part 1, we didn’t let a tester change the choice. And in the study part two, we got the tester change the choice.
6. The first score was about 1/3 and the second score was about 2/3.
7. In my opinion, the hypothesis I made was right.

Appendix 21: Rewrite from Memory example 2

**Discussion**

First, we think from data the probability choose correct cup is one third. Because the cups are same.

Second, we think from data the probability choose correct cup is two third. If he chooses cup before he change cup, he cannot choose correct cup.

Sorry, I cannot finish this paper...

Appendix 22: Rewrite from Memory example 3