TASK COMPLEXITY, THE COGNITION HYPOTHESIS, AND INTERACTION 
IN CMC AND FTF ENVIRONMENTS

A Dissertation 
submitted to the Faculty of the 
Graduate School of Arts and Sciences 
of Georgetown University 
in partial fulfillment of the requirements for the 
degree of 
Doctor of Philosophy 
in the Department of Spanish and Portuguese

by

Melissa Lorrain Baralt, M.S.

Washington, D.C. 
June 9, 2010
ACKNOWLEDGEMENTS

This dissertation process has been a journey whose destination arrival point would not have been possible without the help of many people. The daily encouragement, interaction, stimulation and joy I have received from working in the Department of Spanish and Portuguese at Georgetown University has been one of the greatest gifts during my graduate experience. This ‘home away from home’ and the people in it were a constant source of learning and scholarship. Witnessing daily our undergraduate students at Georgetown learn, engage and produce in a FL served as a constant reminder of how fascinating and important the field of second language acquisition is. Their energy and effort were often the source of my empirical inspiration and pedagogical task ideas.

I am indebted to my dissertation committee for their continuous encouragement and direction. My mentor and ‘Guyanese father,’ Dr. Ron Leow, taught me to ever be ‘aware’ when operationalizing theoretical constructs. I have cherished our research talks as well as our discussions on leadership and language program direction. Ron always sees the bigger picture, and has been an incredible professor, director, mentor and guide. Always putting God, family and health first, I aim to balance as he has done. I would like to thank Dr. Rusan Chen for his assistance with statistical analyses and for all our time together looking at data. He taught me a great deal about SPSS and how to quantify the seemingly impossible. To Dr. Andrea Révézsz, for teaching me about the world of task-based language teaching. Andrea is one of the best critical readers I know, and a stickler on methodology. I am grateful for her friendship and counsel. I am also thankful to Dr. Peter Robinson, the author and mind behind the Cognition Hypothesis. I am thankful for his guidance and for always encouraging me to read outside of the field. All of my
committee members were instrumental in getting me to the end of my journey, and I aspire to be as dedicated of a professor and researcher as they are.

To my other professors: Dr. Alison Mackey, Dr. Cristina Sanz, Dr. Héctor Campos, Dr. Elena Herburger, Dr. Tom Walsh, Dr. Alfonso Morales-Front, and Dr. Rod Ellis, visiting professor to Georgetown, 2009: you are the source of my linguistic training and preparedness, of inspiration and support. I am also grateful for the camaraderie and mentorship of Dr. Michael Ferreira, Dr. Veronica Salles-Reece, Dr. Gwen Kirkpatrick, Dr. Barbara Mujica, Dr. Abelardo Hernando, and Georgetown University’s provost, Dr. James O’Donnell.

To my fellow graduate students, my colleagues and journey buddies, I am eternally appreciative. Ellen Johnson and Germán Zárate, my officemates and co-directors, were my ultimate ambassadors of help in achieving balance. Their friendship and encouragement got me through the toughest journey points. I would not have survived without Sarah Grey’s help in collecting post-test data, but mostly, without our ‘neuro’ talks and afternoon vanilla coffees. Julio Torres, my source of zen, was a partner in brainstorming ways to operationalize task complexity. I will forever cherish our empirical chats. Grant Armstrong, my second ‘brother,’ taught me a great deal about linguistic theory. He, Rebecca Sachs, Mika Hama, Luis Cerezo and Maymona Al-Khalil always helped and encouraged me to think outside of the box. To my other linguists: Nina Moreno, Bea Lado, Almitra Medina, Annie Calderón, Charlie Nagle, Maria Moreno, Sheri Anderson, and to the literature gurus: Michelle Ferrell, Álvaro Baquero, Maisha Mitchell, Clara Pascual-Argente, Lucas Izquierdo, Patricia Soler and José Lara ... you all were guiding steps along my journey, perhaps more than you shall ever realize.

Most deserving of a thank you is my research partner and best friend, Laura Gurzynski-Weiss. I could not have finished this journey without you. You are my sounding board, my best
critic, and my inspiration. I will always be grateful for the hours we spent collecting and analyzing data, consuming coffee, and trying to achieve balance. (Perhaps we’ll figure it out some day!). Laura, you have been my guiding light from start to finish.

I want to conclude by dedicating this dissertation, this culmination of 5 years of work, to my amazing family: my husband, Alejandro Baralt, my parents, Mick and Holly Bednarek, and my three siblings: Chris, Stephanie and Joanna. You guys complete me. Your support, patience, and love were instrumental during this journey. Ale, my other half and Venezuelan counterpart, always helps me to see things from a different perspective. Indeed, he walked this entire journey with me. My parents …I can’t even count the hours that my mom dedicated to me, be it phone encouragement, hot meals, or study break walks. My dad is forever my source of hooah, of motivation and of leadership. Mom and Dad, I thank you for giving me life and language. I could not have done this without all of you. You guys are my foundational rock, and this dissertation is for you.
ABSTRACT

The construct of cognitive complexity has played an increasingly important role in studies on task design, which aim to explore how increases in the cognitive complexity of tasks differentially mediate interaction and learning outcomes (Kim, 2009; Gilabert, Barón, & Llanes, 2009; Kim & Tracy-Ventura, forthcoming; Nuevo, 2006; Révész, 2009, forthcoming; Révész, Sachs & Mackey, forthcoming; Robinson, 2001, 2007). The Cognition Hypothesis (Robinson 2001a, 2003, 2005a, 2007b, 2010; Robinson & Gilabert, 2007) predicts that more cognitively complex tasks will result in greater incorporation of forms made salient during interaction, and that cognitive individual differences will affect learners’ performance as the tasks increase in complexity. In addition, researchers have posited that modality may play an important and differential role for SLA; however, studies on computer-based interaction have thus far been tangential to task-based research. The research on the effects of increases in cognitive complexity on learning is so far inconclusive, with no study to date comparing its effects in different modes.

The current study sought to fill this gap by operationalizing the Cognition Hypothesis, looking at the effects of increases in task complexity and modality on L2 development alongside the provision of recasts. Learners engaged in two-way interactive tasks for which they had to come up with the intentional reasons of peoples’ actions (+complex) or not (-complex). In addition, learners carried out the task with the researcher in either the face-to-face (FTF) or computer-mediated communication (CMC) mode. 70 intermediate-level learners of Spanish were
randomly assigned to one of the following groups: FTF+C, FTF-C, CMC-C, and CMC+C. The targeted linguistic item was the Spanish past subjunctive. Uptake was explored as a mediating variable for learning, and working memory capacity (WMC, measured via the OSPAN, CSPAN, and RSPAN; cf. Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005) was explored as a moderating variable. Independent measures of task complexity were also collected, including time judgments of the tasks and anxiety and perceived difficulty questionnaires.

Results indicated that engaging in more cognitively complex tasks yielded higher development, but differentially so according to mode. The +complex task resulted in the highest gains for the FTF mode, but hardly any development for the CMC mode. The -complex task in the CMC mode led to the highest amount of development. Contrary to expectations, uptake and WMC did not predict learning, and in fact were negatively and significantly related to development in the FTF+C group.

In order to explicate these findings, a deeper probe into the concurrent processes demonstrated by the participants during interaction was carried out. The follow-up analysis revealed that while some participants noticed the form, others demonstrated exemplars of hypothesis testing and rule formation, features associated with awareness at the level of understanding (cf. Leow, 1997; Rosa & Leow, 2004; Rosa & O’Neill, 1999). In fact, awareness and production of the form during the treatment appeared to be the clinching factors that explained the superior performances of FTF+C and CMC-C.

To conclude, it was found that (1) increases of cognitive complexity in the FTF mode appear to promote deeper processing and subsequent higher level of awareness, which was found to significantly predict L2 development in this study, (2) L2 development in CMC can be extended to the FTF mode, (3) modality and task complexity interact in unique ways for SLA,
(4) neither uptake nor WMC was found to predict L2 development, and (5) a more fine-grained operationalization of what constitutes uptake after feedback may be needed in future research employing this concept.
# TABLE OF CONTENTS

I  STATEMENT OF THE PROBLEM........................................................................................................1

II  HISTORY AND DEVELOPMENT OF THE COGNITION HYPOTHESIS ............................................. 8
  2.1 History of the Cognition Hypothesis: Robinson’s Early Research Agenda ................................. 8
  2.2 The Construct of Cognitive Complexity....................................................................................... 12
  2.3 The Importance of Investigating Individual Differences in Task Design ................................. 15
  2.4 Operationalizing the Task Complexity Differential and Development of a Taxonomy........... 24
    2.4.1 First Proposal for Task Classification ................................................................................. 30
    2.4.2 Second Version: A Triad of Task Complexity, Condition and Difficulty
        Factors .................................................................................................................................. 32
    2.4.3 Third Version: Pedagogic L2 Task Classification ............................................................... 36
  2.5 Robinson’s Cognition Hypothesis and SSARC Model of Pedagogic Sequencing .................. 38
    2.5.1 The Triadic Componential Framework .............................................................................. 39
    2.5.2 Predictions of the Cognition Hypothesis ............................................................................ 41
        In terms of L2 production
        In terms of interaction, noticing, and uptake
        In terms of individual differences
    2.5.3 SSARC Model ..................................................................................................................... 41
  2.6 Other Proposed Schemas for Designing and Investigating Tasks ............................................ 42
  2.7 Benefits of Operationalizing the Cognition Hypothesis .............................................................. 47
  2.8 Summary of the Chapter ........................................................................................................... 48

III  RESEARCH ON THE TASK COMPLEXITY DIFFERENTIAL.............................................................. 49
  3.1 Studies Operationalizing the Cognition Hypothesis ................................................................. 50
    3.2 Strand 1: The Effect of Increases in Task Complexity on Interaction .................................. 51
        3.2.1 Summary of Cognitive Complexity and the Effect on Interaction ............................... 70
  3.3 Strand 2: The Effect of Increases in Task Complexity on L2 Development .......................... 71
        3.3.1 Summary of the Cognitive Complexity and the Effect on L2 Development ............. 88
  3.4 Measuring L2 Development with +/- Complex Tasks: Nuevo’s Call for
      Customized Assessments ........................................................................................................ 88
        3.4.1 Benefits of Customized Assessments and the Cognition Hypothesis ......................... 89
        3.4.2 Cons ............................................................................................................................... 89
        3.4.3 A Mixed-Method Approach ......................................................................................... 90
  3.5 Measuring Cognitive Complexity: Norris and Ortega’s Call for Independent Measures ....... 90
        3.5.1 Self-Reported Data: Quantitative and Qualitative ......................................................... 91
        3.5.2 Time Judgments of Tasks ............................................................................................ 92
  3.6 Summary of the Chapter ........................................................................................................... 94

IV  THE CONSTRUCT OF WORKING MEMORY CAPACITY ................................................................ 96
  4.1 On the History of Memory ......................................................................................................... 96
  4.2 Two Theories of Working Memory Capacity ........................................................................... 97
    4.2.1 Baddeley’s Multi-Component Model ............................................................................... 97
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1.1</td>
<td>Phonological loop</td>
<td>98</td>
</tr>
<tr>
<td>4.2.1.2</td>
<td>Visuospatial sketchpad</td>
<td>100</td>
</tr>
<tr>
<td>4.2.1.3</td>
<td>Central executive</td>
<td>101</td>
</tr>
<tr>
<td>4.2.1.4</td>
<td>Episodic buffer</td>
<td>102</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Just and Carpenter’s Constrained Capacity Theory</td>
<td>103</td>
</tr>
<tr>
<td>4.2.2.1</td>
<td>Simultaneous activation and exceeding capacity</td>
<td>105</td>
</tr>
<tr>
<td>4.3</td>
<td>Similarities and Differences Between the Two Models</td>
<td>106</td>
</tr>
<tr>
<td>4.4</td>
<td>Research on the Construct of WMC</td>
<td>106</td>
</tr>
<tr>
<td>4.5</td>
<td>Different Definitions of WMC</td>
<td>107</td>
</tr>
<tr>
<td>4.6</td>
<td>Measuring Working Memory Capacity</td>
<td>108</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Recommendations of Conway et al. (2005)</td>
<td>109</td>
</tr>
<tr>
<td>4.6.1.1</td>
<td>OSPAN (Turner &amp; Engle, 1989)</td>
<td>110</td>
</tr>
<tr>
<td>4.6.1.2</td>
<td>RSPAN (Daneman &amp; Carpenter, 1980)</td>
<td>111</td>
</tr>
<tr>
<td>4.6.1.3</td>
<td>CSPAN (Engle, Tuholski et al., 1999)</td>
<td>111</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Grading WMC Tests</td>
<td>112</td>
</tr>
<tr>
<td>4.6.2.1</td>
<td>Partial-credit unit scoring</td>
<td>113</td>
</tr>
<tr>
<td>4.7</td>
<td>Exploring Working Memory in the Field of SLA</td>
<td>114</td>
</tr>
<tr>
<td>4.7.1</td>
<td>Working Memory Capacity in the FTF Mode</td>
<td>114</td>
</tr>
<tr>
<td>4.7.2</td>
<td>Working Memory Capacity in the CMC Mode</td>
<td>119</td>
</tr>
<tr>
<td>4.8</td>
<td>Summary of the Chapter</td>
<td>124</td>
</tr>
</tbody>
</table>

V THE ENVIRONMENT OF COMPUTER-MEDIATED COMMUNICATION ....127
5.1 Definition and Description of CMC ....128
5.2 Theoretical Premise for Incorporating Tasks in CMC ....128
5.3 Research on CMC ....129
5.3.1 CMC and Interaction ....129
5.3.1.2 Comparing Interaction in CMC versus FTF ....133
5.3.2 CMC and Learning ....135
5.3.3 The Provision of Feedback in CMC to Promote Learning ....142
5.4 Extending the Study of Task Design to CMC ....144
5.5 The Cognition Hypothesis, Feedback, and CMC versus FTF ....144
5.6 Summary of the Chapter ....145
5.7 Purpose of the Study ....146
5.8 Research Questions ....148
5.8.1 Research Questions 1-3: On learning ....149
5.8.2 Research Question 4: On uptake ....149
5.8.3 Research Question 5: On working memory capacity ....150
5.8.4 Research Question 6: On independent measures of cognitive complexity ....150

VI STUDY DESIGN AND METHODOLOGY ....151
6.1 Explanation of Terms ....151
6.1.1 Operationalization of Recasts ....151
6.1.2 Operationalization of Uptake ................................................................. 152
6.1.3 Operationalization of Task Complexity ............................................... 153
6.2 Justification for Linguistic Item ............................................................... 154
6.2.1 The Spanish Past Subjunctive ............................................................. 154
6.2.2 Acquisition of the Spanish Subjunctive and Developmental Readiness .... 157
6.3 Independent Measures of the Cognitive Complexity of the Task ............. 160
6.3.1 Anxiety/Perceived Difficulty Questionnaire ........................................ 160
6.3.2 Time Judgments .................................................................................. 161
6.3.3 Tabulation of Qualitative Comments in the During- and Post-Task Phase .................................................................................. 161
6.4 Measures of L2 Development: Productive and Receptive Assessments .... 162
6.4.1 Tailor-made Component of Assessments in Complex Groups ............... 162
6.5 Operationalizing +/- Intentional Reasoning: the History and Development of the Task ............................................................... 163
6.5.1 Pilot Study ............................................................................................ 164
6.5.1.1 First Task ....................................................................................... 165
6.5.1.2 Second Task .................................................................................. 166
6.5.1.3 Third Task ..................................................................................... 167
6.5.1.4 Fourth Task ................................................................................... 168
6.5.1.5 Fifth Task ....................................................................................... 170
6.5.1.6 Sixth Task ...................................................................................... 170
6.5.1.7 Seventh Task ................................................................................ 172
6.5.1.8 Eighth Task .................................................................................... 176
6.5.1.9 Ninth Task ..................................................................................... 178
6.6 The Current Study: Research Design and Methodology ......................... 181
6.6.1 Participants ......................................................................................... 181
6.6.1.1 Group Comparability ...................................................................... 183
6.6.2 Design ................................................................................................. 183
6.6.3 Treatment Task ................................................................................... 184
6.6.3.1 Recasts During the Treatment ....................................................... 186
6.6.4 Assessments ........................................................................................ 187
6.6.4.1 Critical Items Used for Treatment and Assessments ....................... 188
6.6.4.2 Multiple-Choice Receptive Test .................................................... 189
6.6.4.3 FTF and CMC Production Tasks .................................................. 191
6.6.5 Background and Bio-Data Questionnaire ........................................... 194
6.6.6 Anxiety and Perceived Difficulty Questionnaire ................................ 195
6.6.7 Working Memory Spans .................................................................... 196
6.6.7.1 OSPAN ......................................................................................... 196
6.6.7.2 RSPAN ....................................................................................... 198
6.6.7.3 CSPAN ......................................................................................... 199
6.6.8 Exit Questionnaire ............................................................................. 200
6.6.9 Equipment .......................................................................................... 201
6.7 Procedure ............................................................................................... 202
6.7.1 Transcription and Preparation of the Data .......................................... 210
6.8 Coding and Scoring Procedures ............................................................. 210
6.8.1 Bio-Data ............................................................................................ 210
6.8.2 FTF and CMC Production Tasks ....................................................... 210
7.4.2
7.4.1
7.4
7.3.4
7.3.2
7.3.1
7.2.1
7.2
7.1.4.3
7.1.4.2
7.1.4.1
7.1.4
7.1.3
7.1.3.1
7.1.3
7.1.2.5
7.1.2.4
7.1.2.3
7.1.2.2
7.1.2.1
7.1.2
7.1.1.5
7.1.1.4
7.1.1.2
7.1.1
7.1
6.9.4
6.9.3
6.9.2
6.9.1
6.9
6.8.4
6.8.3
6.8.2.1
1.3
.3
108
VII
RESULTS
.................................................................220
7.1
RQs 1-4: The Effects of Task Complexity and Modality on L2 Development...220
7.1.1
FTF Production Task ...........................................221
7.1.1.1
Descriptive Statistics...........................................221
7.1.1.2
Inferential Statistics ...........................................223
7.1.1.3
Descriptive Statistics per Group ................................225
7.1.1.4
Inferential Statistics per Group .................................226
7.1.1.5
+/- Complex: Variance in Performance..........................227
7.1.2
CMC Production Task ...........................................229
7.1.2.1
Descriptive Statistics...........................................229
7.1.2.2
Inferential Statistics ...........................................231
7.1.2.3
Descriptive Statistics per Group ................................233
7.1.2.4
Inferential Statistics per Group .................................234
7.1.2.5
+/- Complex: Variance in Performance..........................236
7.1.3
Multiple-Choice Receptive Test ..................................237
7.1.3.1
Descriptive Statistics...........................................237
7.1.3.2
Inferential Statistics ...........................................239
7.1.3.3
Descriptive Statistics per Group ................................241
7.1.3.4
Inferential Statistics per Group .................................242
7.1.3.5
+/- Complex: Variance in Performance..........................244
7.1.4
+Complex Performance on Tailor-Made Items ......................244
7.1.4.1
Tailor-Made Items in FTF Production ............................245
7.1.4.2
Tailor-Made Items in CMC Production ............................245
7.1.4.3
Tailor-Made Items in Receptive Development ....................246
7.2
RQ 5: Uptake as a Mediator of L2 Development ......................247
7.2.1
Uptake as a Mediator for FTF Production Development ............248
7.2.2
Uptake as a Mediator for CMC Production Development ..........249
7.2.3
Uptake as a Mediator for Receptive Development ................249
7.3
RQ 6: WMC as a Moderator of L2 Development ......................250
7.3.1
WMC as a Moderator for FTF Production Development ............251
7.3.2
WMC as a Moderator for CMC Production Development ..........252
7.3.4
WMC as a Moderator for Receptive Development ................253
7.4
RQ 7: Independent Measures of Cognitive Complexity ...............255
7.4.1
Time Judgments ..................................................255
7.4.2
Anxiety and Perceived Difficulty Questionnaire ..................258
7.5  Summary of the Results ........................................................... 260
7.6  Justification for Further Analyses ............................................ 265
7.6.1  Recasts, Uptake, and Production-of-Form During the Treatment ... 265
7.6.2  Analysis of Concurrent Processes During the Treatments ........... 269
7.6.2.1  Evidence of Noticing ....................................................... 270
7.6.2.1.1  FTF Mode ........................................................... 270
7.6.2.1.2  CMC Mode ........................................................... 270
7.6.2.2  Evidence of Awareness at the Level of Understanding .......... 271
7.6.2.2.1  FTF Mode ........................................................... 272
7.6.2.2.2  CMC Mode ........................................................... 273
7.6.3  Coding for Noticing and Awareness ........................................ 274
7.6.4  Noticing at the Level of Awareness as Predictive of L2 Development... 278

VIII  DISCUSSION AND CONCLUSION ............................................. 279
8.1  Research Questions 1-4 .......................................................... 279
8.1.1  Example of Interaction in FTF-C Condition ............................ 281
8.1.2  Example of Interaction in FTF+C Condition ........................... 283
8.1.3  Example of Interaction in CMC-C Condition ........................... 286
8.1.4  Example of Interaction in CMC+C Condition ........................... 288
8.1.5  Customizing Assessments to Measure Learning with Cognitive Complexity 291
8.2  Research Question 5 .............................................................. 292
8.3  Research Question 6 .............................................................. 294
8.4  Research Question 7 .............................................................. 296
8.5  Noticing and Awareness: Exploring Learners’ Concurrent Processes ... 298
8.6  Summary of Findings ............................................................. 301
8.7  Conclusion and Implications .................................................... 302
8.7.1  Conclusion .......................................................................... 302
8.7.2  Implications ......................................................................... 304
8.7.2.1  Implications for Feedback and Uptake ................................. 304
8.7.2.2  Theoretical Implications ................................................... 305
8.7.2.3  Methodological Implications ............................................. 307
8.7.2.4  Pedagogical Implications .................................................. 308
8.8  Limitations and Future Research .............................................. 308
8.8.1  Limitations ......................................................................... 308
8.8.2  Future Research ................................................................. 310
8.9  Summary ................................................................................. 312

BIBLIOGRAPHY .......................................................... 313
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>A Functional Model of the Phonological Loop (Baddeley, 2003)</td>
<td>99</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>The Multi-Component Working Memory Revision (Baddeley, 2003)</td>
<td>103</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Images from Homemade Videos for Fifth Pilot Task</td>
<td></td>
</tr>
<tr>
<td>Figure 4.</td>
<td>Example of Images from Video Used in Sixth Pilot Task, -Complex</td>
<td>170</td>
</tr>
<tr>
<td>Figure 5.</td>
<td>Chat Excerpt from Seventh Pilot Task</td>
<td>171</td>
</tr>
<tr>
<td>Figure 6.</td>
<td>Ninth Pilot Task: Final Task Version for the Dissertation</td>
<td>179</td>
</tr>
<tr>
<td>Figure 7.</td>
<td>Blackboard© Test Design Features for Multiple-Choice Receptive Test</td>
<td>188</td>
</tr>
<tr>
<td>Figure 8.</td>
<td>Example of OSPAN Stimuli</td>
<td>197</td>
</tr>
<tr>
<td>Figure 9.</td>
<td>Example of RSPAN Stimuli</td>
<td>199</td>
</tr>
<tr>
<td>Figure 10.</td>
<td>Example of CSPAN Stimuli</td>
<td>200</td>
</tr>
<tr>
<td>Figure 11.</td>
<td>Experimental Procedure</td>
<td>209</td>
</tr>
<tr>
<td>Figure 12.</td>
<td>FTF Production: Mean Scores for Task Complexity x Time</td>
<td>222</td>
</tr>
<tr>
<td>Figure 13.</td>
<td>FTF Production: Mean Scores for Modality x Time</td>
<td>223</td>
</tr>
<tr>
<td>Figure 14.</td>
<td>FTF Production: Mean Scores per Group</td>
<td>225</td>
</tr>
<tr>
<td>Figure 15.</td>
<td>FTF Production: Box Plot Showing Performance for +/- Complex Groups</td>
<td>228</td>
</tr>
<tr>
<td>Figure 16.</td>
<td>CMC Production: Mean Scores for Task Complexity x Time</td>
<td>230</td>
</tr>
<tr>
<td>Figure 17.</td>
<td>CMC Production: Mean Scores for Modality x Time</td>
<td>231</td>
</tr>
<tr>
<td>Figure 18.</td>
<td>CMC Production: Mean scores per Group</td>
<td>233</td>
</tr>
<tr>
<td>Figure 19.</td>
<td>CMC Production: Box Plot Showing Performance for +/- Complex Groups</td>
<td>236</td>
</tr>
<tr>
<td>Figure 20.</td>
<td>M-C Reception: Mean Scores for Task Complexity x Time</td>
<td>238</td>
</tr>
<tr>
<td>Figure 21.</td>
<td>M-C Reception: Mean Scores for Modality x Time</td>
<td>239</td>
</tr>
<tr>
<td>Figure 22.</td>
<td>M-C Reception: Mean scores per Group</td>
<td>241</td>
</tr>
<tr>
<td>Figure 23.</td>
<td>M-C Reception: Box Plot Showing Performance for +/- Complex Groups</td>
<td>244</td>
</tr>
<tr>
<td>Figure 24.</td>
<td>Mean Production of Uptake According to Group</td>
<td>267</td>
</tr>
<tr>
<td>Figure 25.</td>
<td>Reported Levels of Noticing per Group</td>
<td>276</td>
</tr>
<tr>
<td>Figure 26.</td>
<td>Reported Levels of Awareness per Group</td>
<td>277</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

Table 1. Summary of the Findings for Fluency, Accuracy and Complexity on Easy and Hard Tasks (Robinson, Ting, & Urwin, 1995) ........................................... 28

Table 2. Three Dimensions of Task Complexity and their Potential Interactions at Different Levels of a Syllabus (Robinson, Ting, & Urwin, 1995) .................. 30

Table 3. A Triad of Task Complexity, Condition and Difficulty Factors (Robinson, 2001b) ............................................................................................................. 33

Table 4. Pedagogic L2 Task Classification - Categories, Criteria, Analytic Procedures and Characteristics (Robinson, 2007b) .................................................. 37

Table 5. The Triadic Componential Framework for Task Classification (Robinson & Gilabert, 2007) ......................................................................................... 39

Table 6. Studies Examining the Effect of Task Complexity on Interaction .......... 68

Table 7. Studies Examining Task Complexity and Learning Outcomes ............... 86

Table 8. Studies Examining WMC in FTF and CMC: Tests, Methods, and Findings 123

Table 9. Studies Examining Discourse and Interaction in the CMC Mode .......... 132

Table 10. Studies Comparing CMC versus FTF Interaction and Discourse ....... 134

Table 11. Studies Examining Interaction and Learning Outcomes in CMC ........ 143

Table 12. Rule-Based System for Forming the Past Subjunctive in Spanish ....... 157

Table 13. Biota Information per Group ................................................................. 182

Table 14. Critical Items used for Treatment and Assessments ......................... 189

Table 15. Item Types Presented to Learners on the Multiple-Choice Receptive Test ................................................................. 190

Table 16. Items for the Production FTF and CMC Assessment Tasks ............... 193

Table 17. Mean Times for the Productive Assessment Tasks .............................. 208

Table 18. Mean Times for the Treatment Sessions by Group ......................... 208

Table 19. Scoring Procedure Options for WM Span Tasks ................................ 216

Table 20. FTF Production: Mean Scores and Standard Deviations for Task Complexity ................................................................. 221

Table 21. FTF Production: Mean Scores and Standard Deviations for Modality ................................................................. 222

Table 22. FTF Production: Between-Subject Effects .............................................. 224

Table 23. FTF Production: Within-Subject Effects .............................................. 224

Table 24. FTF Production: Mean Scores and Standard Deviations per Group ................................................................. 225

Table 25. FTF Production: Group Differences at Immediate Posttest ............. 226

Table 26. FTF Production: Group Differences at Delayed Posttest .................. 226

Table 27. CMC Production: Mean Scores and Standard Deviations for Task Complexity ................................................................. 229

Table 28. CMC Production: Mean Scores and Standard Deviations for Modality ................................................................. 229

Table 29. CMC Production: Between-Subject Effects ........................................ 232

Table 30. CMC Production: Within-Subject Effects ........................................... 232

Table 31. CMC Production: Mean Scores and Standard Deviations per Group ................................................................. 233

Table 32. CMC Production: Group Differences at Immediate Posttest ........... 234

Table 33. CMC Production: Group Differences at Delayed Posttest ............... 235

Table 34. M-C Reception: Mean Scores and Standard Deviations for Task Complexity ................................................................. 237

Table 35. M-C Reception: Mean Scores and Standard Deviations for Modality ................................................................. 237

Table 36. M-C Reception: Between-Subject Effects ........................................... 240

Table 37. M-C Reception: Within-Subject Effects ........................................... 240

Table 38. M-C Reception: Mean Scores and Standard Deviations per Group ................................................................. 241

Table 39. M-C Reception: Group Differences at Immediate Posttest .............. 242
| Table 40. | M-C Reception: Group Differences at Delayed Posttest | 243 |
| Table 41. | Uptake as a Mediator for L2 Development in FTF Production | 248 |
| Table 42. | Uptake as a Mediator for L2 Development in CMC Production | 249 |
| Table 43. | Uptake as a Mediator for L2 Receptive Development | 250 |
| Table 44. | WMC as a Moderator of L2 Development in FTF Production | 252 |
| Table 45. | WMC as a Moderator of L2 Development in CMC Production | 253 |
| Table 46. | Uptake as a Mediator for L2 Receptive Development | 254 |
| Table 47. | Descriptive Statistics for Time Judgments | 256 |
| Table 48. | Mean Judgment Time Differences per Group | 257 |
| Table 49. | Reported Anxiety and Perceived Difficulty Level by Group | 258 |
| Table 50. | Recasts, Uptake, and Production-of-Form During the Treatment | 266 |
| Table 51. | During-Task-Production as a Predictor of L2 Development | 268 |
| Table 52. | Means and Standard Deviations of Noticing and Awareness by Group | 275 |
| Table 53. | Reported Noticing and Reported Awareness at the Level of Understanding as Predictive of L2 Development | 275 |
| Table 54. | Production Compared to Reported Awareness as a Predictor of L2 Development | 277 |
CHAPTER 1: STATEMENT OF THE PROBLEM

Task-Based Language Teaching (TBLT) has received considerable attention in recent years as a theory of language teaching and learning, much in part due to its groundings in both theory and findings from psycholinguistic research. Advancing conversational interaction as a facilitator of second language acquisition (e.g., the Interaction Hypothesis; Long, 1996; the Pushed Output Hypothesis, Swain, 1995, 2005; Swain & Lampkin, 1995), as well as meaningful opportunities to notice and become aware (e.g., the noticing hypothesis, Schmidt, 1990), the very principles of TBLT promote the use of meaningful, real-world tasks as the ideal promoter of incidental learning. When learners engage in conversational interaction, empirical research has repeatedly shown that negotiation of meaning and second language acquisition (SLA) result, and that critically, engaging learners in interaction “is effectively obtained through the use of tasks” (Mackey, 1999, p. 584, my emphasis).

How to best design tasks for the second and foreign language classroom has consequently become an area of great importance in the field. In fact, the acronym TBLT is also coming to be known as the strand of research associated with the design of tasks, focusing on how different design features of tasks impact language production, cognitive processes, and learning outcomes in unique ways. Variables of interest have included task familiarity (Plough & Gass, 1993; Skehan, 1998); giving learners planning time (or not) before carrying out a task (Crookes, 1989; Ellis, 1987, 2005; Foster, 2001; Foster & Skehan, 1996; Skehan & Foster, 1997, 2005; Yuan & Ellis, 2003); learners’ perceptions of the benefits of pre-task planning (Ortega, 2005); task repetition (Bygate, 1996; 1999, 2001; Bygate & Samuda, 2005; Gass, Mackey, Alvarez-Torres & Fernández-García, 1999; Lynch & Maclean, 2000, 2001; Nemeth & Kormos, 2001); processing load of the task (Skehan & Foster, 1999), and distribution of roles in a task (Yule & MacDonald,
1990). So far, the findings from these studies are seemingly quite consistent, especially in regards to the complexity, accuracy and fluency of oral production that is generated. As Samuda and Bygate (2008) point out,

“Most would agree that ... different types of planning, repetition, familiarity, task structuring and task type are each likely to have a consistent effect on speakers whoever they are and whatever their learning context” (p. 95).

What is still so far under-researched is how different task design features differentially modulate learning. Complicating this endeavor even more are the numerous exogenous and endogenous factors that can play a role in task conceptualization. As a consequence, numerous proposals of task design taxonomies with which researchers and teachers can design, implement, and study tasks have been proposed. Frameworks for the study of tasks have included task ‘operations’ (e.g., listening, speaking, classifying, etc.; Willis, 1996); the different ‘mental activities’ that tasks promote (Anderson & Lynch, 1988); task genre (Arnaudet, 1984; Byrnes, 1990; Brynes & Sprang, 2004; Swales, 1990); task cognitive operations (Prabhu, 1987); interactant relationships, requirements, and task goals (Pica, Kanagy, & Falodun, 1993); input, conditions, processes, and outcomes (Ellis, 2003); and the cognitive complexity of tasks (Robinson, 2001a). This last task component, the classification of tasks according to the cognitive demands they place on the learner, has engendered a great deal of research, largely due to Robison’s Cognition Hypothesis of Task-Based Learning (Robinson 2001a, 2003, 2005a, 2007b, 2010; Robinson & Gilabert, 2007).

To researchers, an advantage of the Cognition Hypothesis of Task-Based Learning is, as recently highlighted by Gilabert, Barón, & Llanes (2009), the way in which it unites the cognitive and interactive research agendas. The main premise of the Cognition Hypothesis is that
the cognitive complexity of tasks should serve as the theoretical basis for task design. Robinson makes specific predictions in regards to increases of the cognitive complexity of tasks on interaction (e.g., Long, 1996; Mackey, 1999), noticing (e.g., Schmidt, 1990), and incorporation of forms (e.g., uptake, see Lyster & Ranta, 1997), and how individual differences in learners’ cognitive abilities affect task-based interaction. Two of his hypotheses that have potentially great implication for task design are that increases in the cognitive complexity of tasks will:

- “promote interaction and negotiation work, and heightened attention to, noticing of, and incorporation of forms made salient to the input and

- individual differences in cognitive abilities (e.g., working memory) and affective factors (e.g., anxiety) will increasingly affect task-based performance and learning as tasks increase in complexity” (Robinson, 2001, p. 56).

So far, six studies have looked at the effects of task complexity on interaction (Gilabert, Barón, & Llanes, 2009; Kim, 2009a; Nuevo, 2006; Révész, in press, a; Robinson, 2001b, 2007a) and six have looked at the effects on learning (Kim, 2009b; Kim & Tracy-Ventura, in press; Nuevo, 2006; Révész, 2009; Révész & Han, 2006; Révész, Sachs, & Mackey, in press). It seems to be the case that more cognitively complex tasks generate more LRE resolutions (Kim, 2009a), more hypothesis testing (Nuevo, 2006), more metalinguistic talk (Révész, in press, a); but that task type (Gilabert, Barón, & Llanes, 2009; Nuevo, 2006) and the proficiency level of the learner (Kim, 2009a) may differentially affect these outcomes. In terms of learning, three studies have found that more cognitively complex tasks result in more L2 development (Révész, 2009; Kim, 2009b; Kim & Tracy-Ventura, in press), while two have not (Nuevo, 2006; Révész & Han, 2006). One study has found that more cognitively complex tasks generate more uptake of forms (Robinson, 2007b), while one study has not (Révész, Sachs, & Mackey, in press). In terms of
affective variables, only learner anxiety has been explored in relation to its effect on more complex tasks and subsequent learning: Kim & Tracy-Ventura found that low anxiety learners performed better than high anxiety learners. No study has ever examined learners’ cognitive abilities, such as working memory capacity (WMC), and how they moderate learning with more or less complex tasks. There is no doubt that learner cognitive abilities (i.e., WMC, aptitude, reasoning, field independence) play a major role in SLA, and yet, the study of these cognitive factors in relation to different types of tasks and the learning they promote is minimal. That said, a major end goal of the Cognition Hypothesis is learner-task matching, pairing learners with tasks that best work for them.

In addition, interaction environment, or modality, is one design feature that deserves further investigation when researching tasks. Given the overwhelming increase in the use of technology for language teaching and learning, it is surprising that research on TBLT has so far been tangential to computer assisted language learning (CALL). The theoretical reasons for engaging learners in one-on-one, conversational interaction in computer-mediated communication (CMC) have been positively attested in empirical research, showing that CMC is an environment that supports and promotes negotiation of meaning (Darhower, 2002; Kötter, 2003; Iwasaki & Oliver, 2003; Lee, 2001; Pellettieri, 2000; Toyoda & Harrison, 2002; Tudini, 2003); the provision and noticing of recasts (Sachs & Suh, 2007); and L2 development (de la Fuente, 2003; Sachs & Suh, 2007; Shekary & Tahirian, 2006; Smith, 2004, 2005). Discourse mode has been argued to be a critical dimension in task design (Ellis, 2003), yet no study has ever examined how carrying out tasks of different levels of cognitive complexity is differentially mediated by mode (CMC versus face-to-face, or FTF). Even more surprising is the fact that no
research to date exists that examines how feedback and subsequent learning are mediated by mode, a gap pointed out in a recent meta-analysis on corrective feedback by Li (2010).

Related to this is recent research showing that WMC may affect whether or not learners can notice and subsequently benefit from implicit feedback provided during FTF interaction (Mackey, Philp, Egi, Fujii, & Tatsumi, 2002; Goo, 2009), and that interacting in the CMC mode may attenuate deficiencies in WMC (Payne & Whitney, 2002); yet, no one has explored this by comparing feedback in the two modes, looking at how WMC moderates the efficacy of the feedback. Another serious limitation of SLA research on the construct of WMC has been the way in which it has been measured, administered, and graded. For example, every SLA study so far exploring WMC has used working memory spans created by the researchers themselves, making generalization of findings to other populations difficult. The Cognition Hypothesis predicts that cognitive abilities such as WMC will increasingly affect task-based performance with more cognitively complex tasks. The pedagogical importance of this notion is that, when learners encounter real-world tasks that they must carry out in their L2 (i.e., having to listen and talk with two people at the same time), WMC plays a role; we must therefore design tasks that approximate these demands on the one hand, and further our understanding of how WMC affects the noticing of feedback with different types of tasks and modalities on the other.

Two other important factors also seem to play a role in the research that has operationalized cognitive complexity alongside SLA. First, when it comes to L2 development of a specific form, it may be the case that learners need to be paired with an expert interlocutor so that feedback on that form is potentially seen as such (e.g., Swain, 2005). Nuevo (2006) paired intermediate-level learners of ESL with other learners of the same proficiency level, and found no effects on learning. Révézsz (2009) paired learners with an expert (the researcher), and
provided them with recasts during the treatment. Contrary to Nuevo, she found that the combination of increased cognitive complexity and recasts led to the most learning. Second, the way in which learning is measured when operationalizing Robinson’s Cognition Hypothesis may also play a role. Nuevo (2006) has called for a combination of customized and standardized assessments to best grasp what the learner herself contributed to the task. No study has measured the effects of cognitive complexity on interaction and learning with customized assessments; doing so could afford important insights into the how to assess learning as derived from task-based interaction with tasks of more or less complexity. Thus, if a pedagogic goal is to create tasks that best prepare learners for what they will encounter in the real world, more research is needed on task design, different interactional environments for carrying out tasks, and importantly, on how learner cognitive variables, such as WMC, differentially mediate task performance and subsequent learning. This dissertation aims to fill some of these critical gaps in the literature.

The present study explored the task complexity differential (the Cognition Hypothesis) and its effect on L2 development. Examining the effects of modality, all learners engaged in conversational, one-on-one interaction with the researcher in either the FTF or CMC mode. In light of the findings by Nuevo (2006), Révész (2009) and Révész, Sachs, and Mackey (in press), learners in this study were provided with implicit feedback in the form of recasts by an expert interlocutor (the researcher) during interaction, with uptake of feedback explored as a mediating variable on learning. The study also examined the moderating effect of WMC on learning in the FTF versus CMC environments, and addressed limitations in the research by using WM spans and administrative procedures that are attested in the field of psychology (Conway et al., 2005). Following Nuevo’s (2006) advice, learning in this study was measured via a combination of
customized and standardized assessments. Independent measures of cognitive complexity were also collected (see Norris & Ortega’s 2009 call for more independent and sensitive measures of the construct).

This dissertation is organized into eight chapters. The next chapter outlines the history of Robinson’s Cognition Hypothesis, reviewing the empirical studies that led him to devise his taxonomy for the study of tasks. The third chapter reviews and critiques those studies that have operationalized the Cognition Hypothesis, looking at the effects of task complexity on interaction and on learning. Chapter four outlines theoretical approaches to the construct of WMC, reviewing findings from psychology and how WMC relates to human cognition. To follow is a critical review of those studies that have explored WMC in relation to SLA, both in the FTF mode and CMC. The fifth chapter discusses the environment of computer-mediated communication (CMC), theoretical implications for its use, and empirical findings on the effects of CMC on interaction, uptake and learning. The chapter concludes with the research questions and hypotheses that guided the study. Chapter six explains the design of the present study, operationalization decisions, the pilot study for the finalization of the task, and also describes the participants, materials, method, coding scheme, and statistical analyses employed. The seventh chapter presents the results of the study. To conclude is the eighth chapter, which discusses the quantitative findings. Chapter eight highlights the implications of the findings for the field, theory, task design and pedagogy; reviews the study’s limitations; and addresses areas for future research.
CHAPTER 2: HISTORY AND DEVELOPMENT OF THE COGNITION HYPOTHESIS

This chapter reviews Robinson’s early research agenda and highlights the path and findings that led him to develop the Cognition Hypothesis. It reviews earlier forms of the theory and task design components that were investigated up through the current Triadic Componential Framework (Robinson & Gilabert, 2007). The specific predictions of the Cognition Hypothesis are presented, as well as the most recent version of the theory, the SSARC Model (Robinson, 2010). To conclude is a presentation of other proposed task schemas for empirical research and an explanation for why Robinson’s model was chosen for the present study.

2.1 History of the Cognition Hypothesis: Robinson’s Early Research Agenda

It is interesting to consider research that led to the development of the Cognition Hypothesis. In looking at his earlier research, it is clear that Robinson was unsatisfied with reported findings, especially those from the field of psychology (i.e., Reber 1989, 1993), where claims were made on learning (often ‘unconscious learning’) without taking into account the nature of the task demands. Since Robinson published his dissertation in 1996 (Consciousness, Rules, and Instructed Second Language Acquisition), the empirical investigations he carried out led him to develop the theory based on what he saw as a clear need for an approach that took the cognitive resources of learners into account. Robinson’s empirical studies throughout the 1990’s show how and why this development unfolded. A review of those studies, highlighting his repeated call for the need of such a theory, and trace the development of his Cognition Hypothesis, is provided below.

Robinson’s dissertation explored the interactions of aptitude (specifically the measures of Grammatical Sensitivity and Memory, two subtests of the MLAT) and learning outcomes in four
experimental groups: incidental, instructed, rule-search and implicit conditions. He found that the measures of aptitude were not related to learning in the incidental (process for meaning) group. Since then, he became interested in person versus cognitive variables (i.e., affective variables versus cognitive variables such as aptitude and working memory capacity), and how these differentially affected processing with different types of L2 input and instructional settings. He thus began his research agenda with the examination of different learning conditions and how it is that aptitude complexes relied on different information processing and strategy techniques.

Robinson and Ha (1993) and Robinson (1997a) were two studies that examined these same participant conditions but within different theoretical frameworks taken from the field of psychology.

In Robinson and Ha (1993), the generalizability of Logan’s instance theory of automaticity in adult alphabet arithmetic problems was extended to adult SLA contexts. Robinson and Ha investigated whether adult learners can go from an “attentionally controlled processing” to one that is memory-based, where memory traces are developed because of practice (p. 414). Essentially, Robinson and Ha set out to test the external validity of Logan’s 1988 study (where learners were asked to verify equations such as $a + 2 = c$, $b + 3 = e$, $c + 4 = g$, etc., and came to be able to perform the task by memory instead of by calculations) to adults SLA contexts, or in other words, with natural languages. Their research questions asked if a) Logan’s theory would hold for SLA contexts, and b) the theory would hold for a more simple rule (monosyllabic verbs taking both double object and to-object forms) versus a more complex rule (disyllabic verbs, which only take to-object constructions). The rule employed in this study is one of Robinson’s earliest operationalizations of the +/- complex condition. They predicted that, in accordance with Logan’s theory, reaction times would be longer in the “early, algorithm-
based phase of responding,” and that reaction times would be shorter for monosyllabic verbs than for disyllabic verbs. This was due to the fact that “development of automaticity reflects a transition from algorithm-based processing to memory-based processing” (p. 415). 15 right-handed learners of intermediate-level ESL were first presented with an explanation of the rule, and then, during the training phase, were presented with 36 sentences that reflected both the less and more complex rules eight different times. Only high-frequency words were used for the lexical items presented in the sentences. After the training, learners were then presented with the transfer set of sentences, which consisted of both sentences previously seen and new sentences containing artificial verbs in mono- and disyllabic form. During the training phase, learners had to press the “c” key if the sentence conformed to the rule just presented, or “m” if it did not. Robinson and Ha found that reaction times were indeed longer when participants went from processing old to new instances in the transfer set. Therefore, mean reaction times and numbers of presentation times were inversely related when learners were shown novel verbs. This supported Logan’s theory that learners do go from an “algorithm-based retrieval of stored instances” to one that is based on memory (1993, p. 427). Surprisingly, the hypothesis that reaction times would be greater for disyllabic verbs than for monosyllabic verbs (and their relationship to object roles)—hence, greater reaction times for the complex rule versus the simple rule—was not confirmed. The authors suggest that this could be due to parsing as not being verb-first, a strategy they assumed learners would do. Rather, they suggest, it could be the case that learners pay attention to the entire frame during treatment. Robinson and Ha concluded by suggesting that it might be the case that learners, at least in L2 contexts, can and do rely on both algorithm-based and memory-based knowledge at the same time when judging the grammaticality of sentences. They called for more research into the development of memory-
based processing, especially in regards to whether different linguistic structures and developmental levels might differentially affect automaticity in SLA.

Robinson’s 1997a study also operationalized Logan’s instance theory of automaticity, this time with the four experimental groups he had designed for his dissertation: focus on form, i.e., the enhanced group and the instructed group, and non-focus on form, i.e., the implicit and the incidental group. 60 ESL learners, all native speakers of Japanese, were used in this study. Robinson assumed learning for all four conditions but suggested that the enhanced and instructed group would employ a transfer of rules (or a veridical memory) to the grammaticality judgment task. Importantly, for the implicit and incidental groups, he predicted that learning would be memory-based, item-specific, and nongeneralizable, “because training task demands in these conditions, although leading to obligatory encoding of instances in memory, are unlikely to lead learners to notice the structural coordinates or necessary features of rules” (p. 229, my emphasis). Robinson explained that, unlike the enhanced or instructed groups, the implicit and incidental groups do not have an algorithm on which to base their assessment of the grammatical examples present during training. All participants read 55 stimulus sentences (presented for 10 seconds each), and were asked a follow-up question depending on the condition they were in: “Did the words ______ and ______ occur next to each other in the sentence you just saw?” Learners were given feedback in the form of “correct” versus “incorrect.” The grammatical rule employed in the study was the same as in Robinson and Ha (1993): verb syllabicity and dative alternation. However, this time Robinson used both real and novel artificial verbs (i.e., donked, menided). Dependent variables were accuracy and reaction times, while independent variables were groups and grammatical rules (either old or novel, monosyllabic or disyllabic verbs). Learners in the instructed group outperformed all other groups in the ability to generalize...
knowledge developed during the training session to novel grammatical and ungrammatical sentences. All participants had significant faster reaction times to old grammatical sentences than they did to novel ones. In interpreting the results, Robinson (1997) made continued reference to the demands of the tasks as explanatory for his results:

…task demands constitutive of two training conditions, that is, the enhanced and instructed conditions, attempt to lead learners to allocate focal attention to critical aspects of the form of input necessary to noticing and learning, whereas the task demands in the implicit and incidental conditions do not aim to do this (p. 243, footnote 1).

Robinson suggested that outcomes were because of a “transfer of processing demands encouraged during training” (p. 241). Therefore, he explained, it is not an unconscious learning or deduction or rules, or a reliance on two different memory systems that differentially determine learning outcomes, but rather what the demands of the task are that cause learners to rely on different cognitive processes during the task. These findings were some of the first that prompted Robinson to consider how task demands led to different processes and outcomes.

2.2 The Construct of Cognitive Complexity

Robinson’s 1995a study was one of his first major empirical investigations in which he explored cognitive task complexity as an isolated independent variable (more versus less complex), examining how the different cognitive demands of the tasks affected L2 production. Reviewing literature from both L1 acquisition and psychology, Robinson provided a very detailed and convincing explanation about how and why the conditions of speaking in the “here-and-now” versus “there-and-then” elicit cognitively different processes, such as memory and attention. In this study, Robinson referenced L1 literature explaining how young children
develop the capability to communicate displaced reference.¹ Children acquire this ability, “a complex event [that is] acquired interactionally in the L1” (p. 102), late in their language development. A child’s ability to go from the ‘me’ and ‘present’ to referencing others and in different time frameworks goes in parallel with cognitive development. Citing Givón’s terminology of “syntacticization,” Robinson reviews how such development involves the functional and processing demands of language and communicative need, given that cognitive development corresponds with linguistic development in children. He explained that as babies develop in the L1, they move from the “pragmatic” mode (a iconic mode based in real-time context) to the “syntactic” mode. Therefore, the functional demand of being able to manage and communicate more complex ideas (such as “there and then”) is a reason for which children—and the L2 learner—need to develop greater syntactic resources. When no contextual support is available, syntacticization results: “greater structural complexity tends to accompany greater functional complexity in syntax” (Givón, cited in Robinson, 1995a, p. 102). Learners must then be able to retrieve from declarative memory in order to communicate successfully. With six hypotheses (no research questions were given in this study), Robinson predicted greater L2 production complexity in the more complex (there-and-then) task. Twelve intermediate-level learners of ESL participated in the study. They were given three strips of a Mr. Brown² comic with no wording present. For the here-and-now condition, participants were asked to describe the comic in the present tense and were allowed to look at it at the same time, hence real-time and based in the present context. All learners were given prompts that helped them start the narrative and thus be guided in terms of what tense to use. For the more complex, there-and-then condition, learners looked at the comic strips, then put them down and had to retell what

¹ Examples of displaced reference include the past tense, deictics, etc.
² This comic strip follows a character named Mr. Brown—see Robinson for details.
happened in the past tense (like the less complex group, they were also given prompts with which to begin their narratives). Both tasks were open and one-way; no interjections or questions were made by the researcher during the narrative. Robinson audio-recorded all of the participants and analyzed their production for measures of accuracy, fluency and lexical complexity. Results yielded significant effects for group in two of the hypotheses: there were significantly more lexical content words in the there-and-then condition than here-and-now, and there was near significance in more target-like use of articles in the there-and-then condition (the remaining four hypotheses were not supported in this study).

In citing the limitations and providing explanations for why the other hypotheses were not met (and acknowledging a potential Type II error\(^3\) in his results), Robinson highlighted the low number of participants, but more importantly, certain features of the task design. First, Robinson suggested that allowing the here-and-now group to keep the picture (contextual support) while narrating the task might not have been sufficient to truly distinguish their performance from the there-and-then group. Second, the task employed was an open, one-way task, hence not taking the hearer’s needs into account. Robinson explained:

*There-and-then tasks differ from here-and-now tasks to the extent that the speaker is forced to code presuppositions that cannot be assumed to be available from a shared context. Because the hearer’s needs were not sufficiently specified, this speaker obligation may have been minimized, and the expected differences under the two conditions reduced. Specifying the hearer’s needs more precisely and setting closed narrative tasks, in which the speaker has to deliver information necessary for the hearer,*

\(^3\) A type II error occurs if the researcher accepts the null hypothesis when it should have been rejected. This is contrary to a type I error, where the null hypothesis is rejected when it should not have been.
with predetermined correct solutions—as in the information-gap activities described by Pica et al. (1993)—may have overcome this deficiency (1995a, p. 127).

The reader thus sees how Robinson reflects on the different design features of a task (open, closed, one-way, two-way, etc), and how it may be the case that these components differentially modulate learner production. In fact, in a footnote on page 131, he distinguishes between task condition, task type, and task complexity, but at this time, these parameters had still not been developed in a detailed fashion. Importantly, this is one of Robinson’s first studies in which he pronounces the need for a theory, drawing on other fields and on existing empirical evidence, to explain why tasks can impose “greater cognitive load” than other tasks (p. 101). He suggests that such a theory should draw on findings from other fields, L1 research, and empirical evidence, and explains:

Tasks requiring accurate or precise use of language, and requiring the expression of multiple propositions drawn from memory, are more likely to “stretch” the interlanguage resources of second language users (Long, 1989) than are tasks not requiring them, and so will lead to greater communicative resource expansion (1995a, p. 101).

Robinson therefore looks to literature on first language acquisition, empirical findings from second language acquisition, and functional linguistic theory to explain what he calls the “task complexity differential” (1995a, p. 102).

2.3 The Importance of Investigating Individual Differences in Task Design

Many of Robison’s early 1990’s studies also explored the putative distinction of explicit/implicit knowledge and learning (e.g., Robison, 1994; 1997a), and it is worthy of note to examine how these empirical investigations also contributed to his growing interest in the design
characteristics of tasks as being a moderating factor with the ability to affect L2 performance. Often citing literature from the field of psychology, Robinson’s research findings caused him to dispute claims that the view of human memory as two dichotomous memory systems (implicit and explicit) are evidence for Krashen’s distinction of acquisition as an unconscious process, and learning as a conscious one. Specifically, Robinson referenced the different cognitive demands of tasks to counter these claims. One researcher Robinson referred to often was Reber (1989, 1993), whose studies deal with artificial grammars (i.e. strings of letters organized by a finite grammar), and experimental conditions in which participants are asked to either a) memorize letter strings or b) to explicitly look for rules that govern the strings. According to Reber, learners in both conditions typically “learn;” when asked to verbalize the rules, however, the implicit group is unable to do so. Reber argues that this is evidence for unconscious learning.

Robinson challenged this view, highlighting the methodological shortcomings in psychology studies that made claims for the acquisition of implicit knowledge as an unconscious process. He did so by referencing the task used in psychology studies and how they modulate outcomes. For example, several psychological studies claimed two separate forms of memory as based on their results of participants’ performance on indirect versus direct tests (Robinson, 1997b, p. 51). Robinson suggested that what was probably happening in these studies was a transfer of processing strategies encouraged by the different task instructions employed during the experiment. He suggests that Reber’s results, in which Reber claimed that participants learned unconsciously, could also be explained as a transfer of frequently presented bigrams, or veridical memory (this term comes from Perruchet & Pacteau, 1990, cited in Robinson, 1997b, p. 227).4 Robinson also referred to neurophysiological evidence for memory, citing studies that

---

4 Robinson also cites Perruchet (1994) in his 2001a study, p. 52.
showed that memory is affected when the hippocampus and neocortex surrounding the amygdala are impaired, but not lesions to amygdala. Note that Reber and Krashen never look to neurophysiology to support their findings.

In his 1997b study, Robinson examined the effect of task as well as individual differences and how these correlate with learners’ subsequent levels of awareness. His theoretical basis was Schmidt’s (1990) noticing hypothesis, which contends that conscious noticing is necessary for L2 development and that consciousness at the level of awareness (operationalized by being able to verbalize a rule) facilitates and is most related to L2 development. Similar to his 1997a study, Robinson looked at the effects of aptitude on learning under four conditions: unconscious implicit and incidental groups, and conscious rule-search and instructed groups. He also explored the relationship between aptitude and awareness, and the effects of aptitude on learning in these different conscious vs. non-conscious focus on form instructions. In this study, 104 intermediate-level learners of ESL were first explained the demands of their respective task and were then exposed to 40 sentences that employed an easy and a complex rule (the same monosyllabic and disyllabic verb relationship rule with object constructions as in the 1997a and 1993 studies). To measure aptitude, Robinson employed two subtests of the MLAT: the Grammatical Sensitivity and Memory tests. Robinson also examined their awareness of the rule, operationalized as learner responses to a debriefing questionnaire asking if they had (1) noticed rules, (2) were looking for rules, and (3) could verbalize the rules. The tasks differed in the following ways: the implicit group was simply told that they were going to perform a memory task. The incidental

5 Robinson’s operationalization of measures of awareness, where a binary option of yes/no in answering three questions on a questionnaire was employed (did you notice, did you look for rules, can you verbalize rules) is a limitation that he himself acknowledges. More introspective methods, such as verbal think-aloud protocols or stimulated recall would have been a more robust way to measure the construct. See Leow (2000; 2001; 2006) for methodological issues in measuring learner awareness in SLA.
group was informed that they were going to perform a comprehension activity, and were then going to read sentences and be asked to answer yes/no comprehension questions. The rule-search group was told to find rules that explained the sentences they were about to see. Finally, the instructed group was given both written and oral explanations of the easy and hard rules before the training task.

Results showed that awareness at the level of Noticing was not associated with learning in any of the four conditions, but awareness at the level of Ability to Verbalize was associated with learning in the implicit and rule-search groups. In other words, the demands of the task moderated outcomes. Interestingly, Robinson found that aptitude was related to learning under the two explicit conditions (rule-search and instructed), but not to learning under implicit conditions (implicit and incidental groups). Aptitude was not related to awareness at the level of Noticing or rule-search, however it was significantly related to awareness at the level of Ability to Verbalize. Therefore, participants able to verbalize a rule had higher aptitude scores. Finally, in regards to learning of both the easy and hard rules, the instructed condition performed significantly better than the implicit, incidental and rule-search conditions. As with past studies, Robinson once again references the design of the task to explicate the results, saying that the fact that Grammatical Sensitivity was positively related to noticing for participants in the rule-search condition was “to be expected, given the nature of the task demands in each condition” (p. 80). Importantly, Robinson concluded by saying that his data showed that learning in all conditions was “fundamentally similar” (p. 81) in that learning for all groups occurred as a result of the task demands of each experimental condition (my emphasis). Therefore, learning outcomes were found to be dependent on the demands of tasks. This was also one of the first studies in which he
showed empirically that learning can be moderated by individual differences (i.e., aptitude) in different ways, depending on the cognitive demand of the task:

To this extent, task demands also determined the role individual differences in Memory and Grammatical Sensitivity played in learning under different conditions. To … summarize, the tasks presented to learners in the 4 different conditions differed in their demands on those cognitive resources measured by the Memory and Grammatical Sensitivity tests; where an individual learner’s strength in those resources matched the demands of the task performed in any condition, aptitude often led to awareness, which often was associated with superior levels of learning. This interpretation counters the claim that learning under Implicit versus Rule-search, or Incidental versus Instructed, conditions reflects the processing operations of distinct unconsciously- and consciously-accessed systems (Robinson, 1997b, p. 81).

In other words, Robinson explains that the difference in what learners noticed was a result of the different task demands they were given, and not due to different memory systems as suggested by Reber (1989; 1993) or (unconscious) acquisition versus (conscious) learning, as suggested by Krashen (1981; 1985). What they noticed was furthermore modulated by their aptitude measures.

Robinson (2005b) is another study in which the effects of three individual differences—this time, aptitude, working memory and IQ—were explored, and how these affect implicit and incidental learning. In reporting three separate replication experiments (all done with the same group of participants over three sessions), the goals of this study were threefold: (1) to see if findings from the field of psychology on unaware learning of complex rule-systems (where artificial grammars were used) were generalizable to L2 settings with natural languages; (2) to
explore the relationship between individual differences and learning; and (3) to replicate Reber, Walkenfeld, and Herstadt’s (1991) study on implicit learning via the extension of their design to natural language settings.

In the first experiment, Robinson replicated Reber et al. (1991), where 20 learners had been placed into two different groups: implicit and explicit (the latter group was explicitly told to look for a rule). Both groups were shown a series of letter strings in a finite artificial grammar. In explaining that implicit learning and memory development earlier in children are not associated with the explicit, cognitive abilities of problem-solving and decision making, Reber et al. had predicted that implicit learning would show little population variance in scores and would not be affected by individual differences in IQ. Their hypothesis was confirmed. (This led Reber to conclude that unconscious learning is not sensitive to individual cognitive measures, such as IQ or aptitude). In his replication, Robinson changed the input by using a natural language, Samoan, and asked learners to process it for meaning. 54 Japanese learners of ESL were asked to memorize 27 vocabulary items in Samoan. They were then given 10 training trials during which they were shown 45 Samoan sentences. During the testing phase, learners were asked ‘yes/no’ questions on the meaning of the sentences; they were also given feedback in the form of correct/incorrect. In this way, their assessment of any acquisition of Samoan grammar rules (based on the use of ergative e, locative markers and noun-verb insertion) was therefore deduced. Robinson found that, like the Reber et al. results, IQ and aptitude were not related to incidental learning of Samoan. However, working memory was related to the listening grammaticality judgment test used.

In the second experiment (carried out by the same group of learners), Robinson used the same artificial grammar design as Reber et al., carrying out a full replication of their study. The
purpose of this replication was to prove his hypothesis that unlike the learning of Samoan (a natural language that requires semantic processing), individual cognitive abilities (IQ, working memory, aptitude) would not be related to learning of artificial grammar. The same learners were shown 12 series of letter sequences and were asked to choose the letter that best completed the series. The “problem-types” were either alphabetic (i.e., ABC, BCD, DEF) orderings or reverse pairs (where the letter string was mirrored). Participants were shown six of each type on the computer and were not provided with feedback. Robinson also provided learners with grammatical high and low chunk-strength strings and ungrammatical high and low chunk-strength (see description of Knowlton and Squire below on chunk-strength).

The third experiment, taking place congruently with the first and second, sought to replicate Knowlton and Squire (1996)’s concept of chunking and frequency in the input, for both the artificial grammar of letter strings and the learning of Samoan. Knowlton and Squire operationalized chunks as bigrams (pairs of letters) and trigrams (sets of three letters) that occurred together in the training set, and calculated chunk-strength as the weight of chunks determined by their frequency of occurrence during exposure. In the transfer set after training, learners were shown more letter series, some of which contained high-frequency chunks that had been shown during training. Knowlton and Squire found that both amnesic learners (with impaired explicit, declarative memory) and control learners were not affected by chunk-strength in their ability to determine grammatical strings as grammatical in the transfer set. The authors concluded that participants had acquired a rule-knowledge that was independent of memory. Participants also often marked ungrammatical items as grammatical if they contained high-frequency chunks; therefore, high-frequency items were affecting grammatical judgments. Robinson aimed to generalize these findings to L2 settings by using chunks in Samoan
(mentioned in experiment 1) and to the artificial grammar settings (experiment 2). Also, after carrying out experiment 1, learners were given a listening grammaticality judgment test (with high and low chunk-strength items) and a guided sentence completion test in the Samoan language.

Results were the following: (1) confirming the findings of Reber et al., there was little population variance in the incidental learning conditions but high variance in scores for the explicit learning conditions; (2) implicit learning of the artificial grammar was unrelated to performance on the posttests of incidental Samoan learning; (3) individual differences of working memory and aptitude did not affect incidental learning in the artificial grammar session; (4) individual differences in aptitude and IQ also did not affect incidental learning of Samoan, however, working memory was correlated to performance on the listening grammaticality judgment task;⁶ 5) the findings of Knowlton and Squire (1996) were replicated: high chunk-strength in both the artificial grammar and the Samoan language learning settings influenced learners to wrongly accept ungrammatical items in both languages. Chunk-strength in the input therefore seemed to affect grammaticality judgments independent of explicit memory.

Robinson discussed findings from the field of psychology where incidental learning of artificial grammars had been operationalized, and articulated how these findings might be extended to implicit learning conditions in L2 settings. First, chunk-learning is different in L2 settings because natural languages contain semantic content features that bigram and trigram chunks of an artificial language do not have. In an L2, “this additional information can influence the subsequent memorability of chunks, independent of frequency alone” (2005, p. 259). Second, it might be the case that individual differences of IQ, working memory and aptitude interact with

⁶ Robinson suggests that this may have been due to the aural mode of the GJT.
artificial languages differently than they do with natural language learning conditions. While Robinson found minimal evidence to support this, he suggested that it might be because of the type of assessment tasks used (listening grammaticality judgment task, sentence completion task, etc.). Robinson went on to explain that while this seemed to provide evidence for Krashen’s claim that incidental SLA works independently of aptitude measures, it may very well be the case that traditional aptitude measures (perhaps subcomponents of the MLAT) “are not well suited to measuring the abilities drawn on in incidental learning” (2005, p. 260), but instead are better suited to measure those abilities applicable to explicit learning or instructed conditions. He furthermore discussed measures of aptitude and in particular, working memory. Robinson drew three major conclusions from this detailed study: a) there is significantly greater variance in performance of explicit learning than implicit learning; b) implicit learning may indeed be different for artificial languages as opposed to natural languages, and this is because of the semantic content inherent to natural language input and subsequent semantic processing; and c) working memory capacity, “the capacity to rehearse some focally attended information and maintain it in working memory while processing other information” can affect incidental learning but not implicit learning of an artificial grammar.

Robinson concluded by calling for more replications of psychology studies in SLA settings and more recognition of the critical importance of individual differences and how natural language settings interact with these cognitive abilities to lead to different outcomes. One can see

---

7 The working memory test Robinson employed in the 2005b study was a reading span test, and not “a measure of aural phonological working memory” (p. 260). While he chose it to match the modality of the treatment sessions, it might have been the case that the test of working memory he used did not draw on the mode of input (Robinson suggests this as both a possible limitation and a cause for his results). Still, this study contributed to a better understanding of individual differences and their effect on incidental learning. Robinson was the first to explore measures of aptitude and working memory and the modalities under which they are delivered in natural language versus artificial language settings.
how, alongside task settings, individual differences came to be such a critical component for task
design and task research in Robinson’s Cognition Hypothesis and the Triadic Componental
Framework.

2.4 Operationalizing the Task Complexity Differential and Development of a Taxonomy

The first study in which Robinson operationalized different task demands and their
effects on both oral and written production was published in 1995 with two of his students, Ting
and Urwin. In this study, the authors describe four small-scale studies, followed by, as based on
their findings, three dimensions of task demands argued to affect the difficulty level of the task:
(1) amount of imposed cognitive load, (2) amount of planning time allowed, and (3) amount of
prior information given. While the studies were for the purpose of pedagogical implications
(with the goal being to inform teachers so that teachers can sequence and grade tasks in a task-
based syllabus), they also informed Robinson’s theory. In reviewing task-based approaches to
pedagogy, Robinson, Ting, and Urwin highlighted the fact that no clear research-based criteria
existed from which teachers and researchers can design, grade, and sequence tasks. A problem at
that time was no task type taxonomy existed that could be “pedagogically useful” (Robinson et
al., 1995, p. 63). The authors called for more action research studies that inform task sequencing
and also that “ensure the ecological validity of the tasks studies as viable units of classroom
activity” (p. 63).

Units of analysis for production for each of the studies were measured in terms of
accuracy (target like use of articles), complexity (S-nodes in each utterance and lexical density)
and fluency (number of pauses and number of words per utterance). Robinson, Ting, and Urwin
hypothesized that on the harder tasks, production would be less fluent, but more accurate and
complex. Once again citing Givón’s claim that “structural complexity tends to accompany functional complexity in syntax” (1985, p. 1021, cited in Robinson et. al. p. 65), the authors hypothesized that “greater attention and difficulty of the task would force the learner to try harder and complexify the speech she produced” (p. 65).

The first study, similar to Robinson 1995a, operationalized “Cognitive Load” by employing a task in the Here-and-Now condition (what they called low cognitive load) versus the There-and-Then condition (high cognitive load). 12 intermediate-level learners of ESL were given a series of the Mr. Brown comic strip and asked to narrate the events in the strip (where pictures only, without words, were provided). For the Here-and-Now condition, learners had to narrate the comic in the present tense while being allowed to view it. For the There-and-Then condition, they were asked to first view the comic and then narrate in the past tense after the strip had been removed. (Citing L1 acquisition literature, the authors explain that they were trying to simulate displaced reference, an ability that is “more effortful and later developed” in children than are activities that are in the present tense and that are referenced in contextualized events, p. 66). Their hypotheses were partially confirmed: the There-and-Then group had significantly more accurate production and significantly greater lexical density. However, in terms of complexity, no difference between groups was found for measures of S-nodes. The hypothesis was not met for measures of fluency: there was no difference in number of pauses, while there was an almost significant difference in number of words per utterance for the Here-and-Now group.

The second study also investigated Cognitive Load, but this time operationalized it via a map task with a pre-marked route from A to B (low cognitive load) versus a map without a pre-marked route (high cognitive load). Ten intermediate-level learners of ESL had to describe the
route from A to B, with those in the high cognitive load group having to simultaneously think up the route themselves. In this way, the authors attempted to operationalize single versus dual task performance as described by Wickens (1989), who “argues that dual task performance is more attention demanding and difficult than single task performance” (Robinson et al., 1995, p. 67). For this experiment, hypotheses were confirmed for measures of fluency (significantly more pausing on the harder unmarked route task), but were not confirmed for either measures of accuracy or complexity.

The third experiment investigated the second proposed dimension of task demand: amount of planning time allowed. This time, participants were native English-speaking learners of Chinese at the high-intermediate level. All learners carried out a difficult and a less difficult writing task in a counter-balanced design (where treatment days were separated by a few days). For the more difficult task, learners had to write a description of a picture sequence with no planning time allowed. For the easier task, learners were shown the picture and then given three minutes to plan what they would write before starting the task. Units of analysis measured were verb morphology, use of measure words and use of tense markers in Mandarin Chinese. Fluency measures include length of production (via words and in t-units). The hypothesis that having planning time would lead to more accurate written descriptions was not confirmed for this task; in fact, those in the unplanned condition have a higher mean of accurate tense markers (though the difference was not significant). The task was also carried out in the oral mode and once again, the hypothesis for greater fluency and accuracy was not confirmed.

Finally, the fourth experiment operationalized the final proposed dimension: Prior Information Supplied. Learners were beginning-level students of Mandarin Chinese, and were divided into three groups: the prior knowledge group, the -prior knowledge group, and the
control group. The prior knowledge group carried out a pre-task activity (focusing on form) where the content was tourism in Beijing. The -prior knowledge group carried out a pre-task activity on the organization of lectures in Chinese. The control group had no pre-task activity. The treatment for all groups was a lecture on tourism in China. Following the lecture, all participants were given a multiple-choice test for comprehension, focusing on facts covered during the lecture and inference questions. The authors hypothesized that the prior knowledge group would have better comprehension of the topic than the -prior knowledge group and the control. Their hypothesis was supported, but only for the inference questions on the test, not the recall items. (Surprisingly, the -prior knowledge and control groups performed similarly on the recall questions.) While mean differences seemingly supported their hypothesis, they did not reach significance.

In line with the findings from the four experiments described above, Robinson, Ting, and Urwin used their findings to create a classification for tasks along an easy to hard continuum:


<table>
<thead>
<tr>
<th>Easy Task</th>
<th>Dimension of Task Complexity</th>
<th>Hard Task</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>&lt;Easy Task</strong></td>
<td><strong>Dimension of Task Complexity</strong></td>
<td><strong>Hard Task</strong></td>
</tr>
<tr>
<td><strong>&lt;Here-and-Now</strong></td>
<td><strong>Cognitive Load Imposed</strong></td>
<td><strong>There-and-Then</strong></td>
</tr>
<tr>
<td>greater fluency</td>
<td>greater complexity</td>
<td>greater fluency</td>
</tr>
<tr>
<td>(more words per utterance)</td>
<td>(lexical density)</td>
<td>(target-like use of articles)</td>
</tr>
<tr>
<td><strong>&lt;Single (Map) Task</strong></td>
<td></td>
<td><strong>Dual (Map) Task</strong></td>
</tr>
<tr>
<td>greater fluency</td>
<td>equivalent accuracy (target-like use of articles)</td>
<td>equivalent complexity (lexical density)</td>
</tr>
<tr>
<td>(more words per utterance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>less pausing)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>&lt;Planned Writing</strong></td>
<td></td>
<td><strong>Unplanned Writing</strong></td>
</tr>
<tr>
<td>equivalent accuracy (target-like use of verb morphology, tense markers and measure words)</td>
<td></td>
<td>equivalent accuracy (target-like use of verb morphology, tense markers and measure words)</td>
</tr>
<tr>
<td>equivalent fluency (number of words &amp; sentences)</td>
<td></td>
<td>equivalent fluency (number of words &amp; sentences)</td>
</tr>
<tr>
<td><strong>&lt;Planned Speaking</strong></td>
<td></td>
<td><strong>Unplanned Speaking</strong></td>
</tr>
<tr>
<td>greater fluency</td>
<td>equivalent accuracy (verb morphology, measure words and tense markers)</td>
<td>equivalent fluency (number of utterances)</td>
</tr>
<tr>
<td>(number of words)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>&lt;Prior Knowledge</strong></td>
<td></td>
<td><strong>No Prior Knowledge</strong></td>
</tr>
<tr>
<td>(Form &amp; Content) greater inference</td>
<td>equivalent recall</td>
<td></td>
</tr>
</tbody>
</table>


In general, the researchers found that harder tasks (be it along the dimension of imposed cognitive load, planning time or prior knowledge) tended to yield more accurate production, while easier tasks led to more fluent production. In explaining why, in the third experiment, significantly greater accuracy was achieved in the planned oral production task but not in the
planned written task, the authors posited that in the oral mode, learners must focus their attention on form as well as meaning and the “propositional coherence of their oral descriptions” (p. 70). In the written mode, it might be the case that learners have more of an opportunity to plan out the form of their writing and possible oversupply a feature. This lead them to suggest that in There-and-Then tasks, a greater cognitive load is necessary “to create the functional ‘need’ for a particular form, and it may be that this, rather than a focus on the form itself, … leads to more target like use” (p. 70). This may have been what inspired Robinson to eventually propose resource-directing dimensions of task design, or variables that promote attention to form-function mappings, because they push the learner to recognize where he/she has linguistic deficiencies—very similar to creating what in 1995 was called a functional ‘need.’

To conclude, the authors proposed a schema of task design comprised of three dimensions for teachers to inform their syllabus designs. This schema (reproduced below) is specifically designed out of consideration for ways in which it might interact at three different levels of a syllabus. In doing so, Robinson, Ting, and Urwin provided suggestions for how an instructor might make a task more difficult (for example, one could make a listening task ‘dual’ by also asking learners to create a diagram or a flowchart during the listening activity) in Table 2 below:

---

8 This is one of the first studies by Robinson in which the important issue of modality in task design is raised. The authors questioned whether writing gives learners more time to plan online.
### 2.4.1 First Proposal for Task Classification

**Table 2. Three Dimensions of Task Complexity and their Potential Interactions at Different Levels of a Syllabus** (from Robinson, Ting & Urwin, 1995, p. 71)

<table>
<thead>
<tr>
<th>Level 1 Tasks</th>
<th>Level 2 Tasks</th>
<th>Level 3 Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive Load</strong></td>
<td><strong>Planning Time</strong></td>
<td><strong>Prior Information</strong></td>
</tr>
<tr>
<td>low</td>
<td>given</td>
<td>given</td>
</tr>
<tr>
<td>high</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>Easy</td>
<td>Easy</td>
<td>Easy</td>
</tr>
<tr>
<td>Hard</td>
<td>Hard</td>
<td>Hard</td>
</tr>
<tr>
<td><strong>High Cognitive Load</strong></td>
<td><strong>No Prior Information</strong></td>
<td></td>
</tr>
<tr>
<td>with planning time</td>
<td>low cognitive load</td>
<td>low</td>
</tr>
<tr>
<td>without planning time</td>
<td>high cognitive load</td>
<td>high</td>
</tr>
<tr>
<td>Not So Easy</td>
<td>Not So Easy</td>
<td>Very Hard</td>
</tr>
<tr>
<td>Much Harder</td>
<td>Much Harder</td>
<td></td>
</tr>
</tbody>
</table>


This classification for tasks was primitive along the course of Robinson’s development of a task design taxonomy. Indeed, Robinson, Ting, and Urwin acknowledged that other dimensions independent of task complexity features could also be incorporated into their schema, such as task condition factors, defined as “those which describe the interactive demands of tasks, such as whether the information flows in one direction, or in two directions, and whether the task requires participants to agree, and converge in their solutions, or whether it requires them to differ and diverge” (p. 72). (Indeed, Robison does end up including interactive conditions in his taxonomy later on). The researchers called for more research on tasks, task complexity, and also...
with languages other than ESL. Lastly, they suggested that more dimensions of task complexity be motivated by literature in other fields, such as ergonomics and task performance (p. 73). These suggestions comprised the first proposed taxonomy of operationalizable task design features for teachers and researchers by Robinson. Minimal in its design at the time, it gave concrete suggestions for how a task might be sequenced to inform syllabus design. The three dimensions they proposed (cognitive load, planning time and prior knowledge) later ended up being dimensions along the cognitive complexity parameter in Robinson’s Triadic Framework. Also not present in this model were individual difference factors; however, other research carried out by Robinson at the time was indicative of his interest in cognitive and affective variables, leading him to eventually incorporate those dimensions in the taxonomy for task design as well.

Indeed, in a book review three years earlier, Robinson (1993) called for the consideration of individual abilities that learners bring to tasks, as well as memory and attentional demands:

What designers need are objective criteria for deciding on the relative difficulty of task types within which they can operate, and which they can use to make recommendations about grading. Research is needed here to establish what these objective parameters of task difficulty are. What seems clear to me is that the notion of task difficulty, if it is to be operationalized as a criterion for task selection on a task-based syllabus, must make reference to shared, universal cognitive constraints that task types impose on learner’s abilities to perform them. This will involve, for example, close examination of the relative memory demands and attentional demands of tasks in order to establish invariant parameters of task difficulty within which task designers can operate (p. 444).

Later, Robinson (1996c) states: “It is not clear yet what the parameters or dimensions of
L2 task complexity are that syllabus designers will have to operate with, a fact widely acknowledged by proponents of task-based approaches themselves (Brindley, 1987; Candlin, 1987, 1993; Long, 1985, 1989; Long & Crookes, 1992, 1993; Nunan, 1989, 1991; Skehan, 1996)" (all cited in Robinson, 1996c, p. 2). He explained in the study that task complexity is a multidimensional construct, and that the field would benefit practically and theoretically by developing a research program on how differences in task complexity differentially affect learning outcomes in second language acquisition.

This call, as well as all of the research reviewed thus far, show why Robinson came to suggest the parameters of task design for research on increasing task complexity, as based on empirical findings and theoretical reasoning.

2.4.2 Second Version: A Triad of Task Complexity, Condition and Difficulty Factors

After his study with Urwin and Ting (1995), Robinson’s first theoretically motivated classification for the purpose of studying the effects of increasing task complexity on the production, comprehension and learning was published in 2001 (Robinson, 2001b). The specific goal of this paper was to provide researchers with an operationalizable framework so that task design, and how it differentially mediates production, comprehension and learning outcomes, could be researched. Showing his psycholinguistic approach to SLA, Robinson designed the framework so that work done in cognitive psychology, such as the psycholinguistic constructs of memory, attention, and automaticity, was also accounted for.

Recall that so far, two proposals had been set forth in terms of how to operationalize and research the increasing of task complexity for research and syllabus design decision-making: the three parameters for task design, No Planning Time, High Cognitive Load, and No Prior Information (Robinson, Ting, & Urwin, 1995) and then the three parameters of Task Condition,
Task Type, and Task Complexity (Robinson, 1995a, footnote). The second option had not been
developed at the time, however.

In the 2001b model, Robinson continued with the three parameters he suggested in
1995a. There, he provided his first detailed triadic framework for examining task complexity. He
took into account three distinct parameters: task complexity (“the task dependent and proactively
manipulable cognitive demands of tasks”), task difficulty (“dependent on learner factors such as
aptitude, confidence, motivation, etc.”) and task conditions (“the interactive demands of tasks”)
(2001b, p. 287). The dimensions in this triadic framework proposed by Robinson were:

Table 3. A Triad of Task Complexity, Condition and Difficulty Factors (from Robinson,
2001b, p. 294)

<table>
<thead>
<tr>
<th>Task complexity (Cognitive factors)</th>
<th>Task conditions (interactive factors)</th>
<th>Task difficulty (learner factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) resource-directing</td>
<td>a) participation variables</td>
<td>a) affective variables</td>
</tr>
<tr>
<td>e.g., +/- few elements</td>
<td>e.g., one-way/two-way</td>
<td>e.g., motivation</td>
</tr>
<tr>
<td>+/- here-and-now</td>
<td>convergent/divergent</td>
<td>anxiety</td>
</tr>
<tr>
<td>+/- no reasoning demands</td>
<td>open/closed</td>
<td>confidence</td>
</tr>
<tr>
<td>b) resource-depleting</td>
<td>b) participant variables</td>
<td>b) ability variables</td>
</tr>
<tr>
<td>e.g., +/- planning</td>
<td>e.g., gender</td>
<td>e.g., aptitude</td>
</tr>
<tr>
<td>+/- single task</td>
<td>familiarity</td>
<td>proficiency</td>
</tr>
<tr>
<td>+/- prior knowledge</td>
<td>power/solidarity</td>
<td>intelligence</td>
</tr>
</tbody>
</table>

Sequencing criteria___________________________________________Methodological criteria
Prospective decisions about task units Online decisions about pairs and groups

Key to his model was that all of the criteria with which researchers could make decisions
about how to increase the cognitive demands of a task were empirically researched and
supported. One important observation: this was the first time that learner factors, vis-à-vis the
task difficulty parameter, were taken into account. At the same time, Robinson said in the paper
that his goal would be to focus on task complexity, because learner factors often “cannot be
anticipated in advance of implementation of a syllabus and therefore can be of no use to the prospective materials and syllabus designer” (p. 287).

In this triadic framework, Robinson suggested that task complexity be sequentially increased to eventually meet real world tasks that the learner will encounter. The teacher or researcher has options to choose from in this model, so that task complexity “can be manipulated to progressively increase the cognitive demands of pedagogic tasks, so they approach the full complexity of the target task” (p. 292). In other words, so that tasks can be sequenced in terms of increasing of complexity, Robinson explained that certain dimensions can be chosen from with either a + or a - feature. The first feature, +/- few elements, serves as an example whereby task complexity can be manipulated to increase the cognitive demands of a task. (Recall that this same feature was operationalized in an earlier study by Robinson, Ting, & Urwin in 1995). This dimension is then broken down into two more dimensions: resource-directing dimension of task complexity and the resource-depleting dimension. The resource-directing dimension refers to how increased demands might be made on the learner’s cognitive resources, “while at the same time potentially directing their resources to aspects of language code that can be utilized in completing the task” (p. 295). An example is the +/- here-and-now feature, where having to communicate in the -here-and-now would require that learners refer to displaced reference such as the past tense. The resource-depleting feature places greater resource demands on the task, so that the task can get closer to approximating real world tasks. A way to operationalize +/- here-and-now is exemplified with the +/- single task, where a -single-task requires that the learner carry out multiple actions at the same time as having to communicate, something that all humans

9 However, Robinson did suggest that if they are addressed in situ, learner factors may in the future be able to inform “on-line methodological choice of options, such as how to pair and group learners under different task conditions” (2001b, p. 287). He pointed out though that there is little research in this area.
have to do in the real world (i.e., answer the phone in English while cooking and looking after a child at the same time).

The next dimension proposed in the Triadic Framework is Task Conditions. This refers to interaction features as based on the participation that a task might require. Examples include having to converge on a decision-making task together or asking learners to maintain different perspectives and to defend their differing viewpoints. The two features on this dimension are Participation variables, which can include +/-one-way, versus Participant variables, which refers to power roles (i.e. +/- power/solidarity) or gender, as an example.

Lastly, Robinson set forth the Task Difficulty parameter, which references how it is that learners might perceive the demands of the task, and is motivated by both affective and cognitive ability factors. Affective factors include variables such as learner motivation and anxiety that might be present while carrying out the task. Ability factors refer to a learner’s aptitude, or say intelligence, that s/he brings to the task. While these factors are out of control of the task designer, they will still play an important part in regards to the outcome of the task, and so should be taken into consideration for task design. Robinson pointed out that it is this dimension, task difficulty, which will explain variation in task performance “between any two learners performing the same task (simple or complex)” (p. 295). Task complexity, on the other hand, should explain “within learner variation in performance on any two tasks (simple and complex) (p. 295).

Robinson went on to say that the Participation/Participant features of the Task Condition dimension, as well as Task Difficulty features, will be hard, if not impossible, to operationalize when designing tasks or making sequencing decisions. This is because many times, the factors along this dimension are out of the control of the researcher and/or teacher. However,
theoretically they could make a difference in terms of production/comprehension/learning outcomes, and future research could look into these variables. Robinson concluded by suggesting that they be specified during, for example, a needs analysis for a course. This rendered the taxonomy as one that is ecologically applicable to learners and instructors too, because it gives instructors options on how to best accommodate their learners.

2.4.3 Third Version: Pedagogic L2 Task Classification

An updated version of the Triadic Framework is provided in Robinson (2007b) (shown below in Table 4). While the three parameters for task design are the same (Task Complexity, Task Conditions, Task Difficulty), there is a significant amount of features that have been added (nine total per parameter), as well as further fine-tuning of subcategorization and a specification of classification criteria and procedures. When asked what led him to incorporate these extra features, Robinson said that it became obvious that more features needed to be included to provide a more holistic taxonomy for task design. In gradually increasing the complexity of any task, learners “approach the full complexity of authentic, real world performance, enabling them to achieve the ‘ends’ of learning … i.e., the ability to complete a range of performance objectives,” which is what we want them to achieve (personal correspondence, 2008). In providing a detailed taxonomy of task features that include learner-related internal variables, sequence design decisions are more fine-tuned. Robinson’s research paradigm therefore has the end goal of task-learner matching, or in other words, the combining of task and instructional conditions that can best suit learner needs and strengths. In line with meeting learners’ needs, this is the pedagogic goal of his Cognition Hypothesis.
Table 4. Pedagogic L2 Task Classification - Categories, Criteria, Analytic Procedures and Characteristics (from Robinson, 2007b, pp. 15–16)

<table>
<thead>
<tr>
<th><strong>Task Complexity</strong> (Cognitive factors)</th>
<th><strong>Task Condition</strong> (Interactive factors)</th>
<th><strong>Task Difficulty</strong> (Learner factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification criteria: cognitive demands</td>
<td>Classification criteria: interactional demands</td>
<td>Classification criteria: ability requirements</td>
</tr>
<tr>
<td>Classification procedure: information-theoretic analyses</td>
<td>Classification procedure: behavior descriptive analyses</td>
<td>Classification procedure: ability assessment analyses</td>
</tr>
</tbody>
</table>

**Sub-categories:**
(a) Resource-directing variables making cognitive/conceptual demands
- +/- here and now (Robinson, 1995)
- +/- few elements (Kuiken et al., 2005)
- +/- spatial reasoning (Becker & Carroll, 1997)
- +/- causal reasoning (Robinson, 2005a)
- +/- intentional reasoning (Baron-Cohen, 1995)
- +/- perspective-taking (MacWhinney, 1999)

(b) Resource-dispersing variables making performative/procedural demands
- +/- planning time (Skehan, 1998)
- +/- prior knowledge (Urwin, 1999)
- +/- single task (Robinson et al., 1995)
- +/- task structure (Skehan & Foster, 1999)
- +/- few steps (Fleishman & Quaintance, 1984)
- +/- independency of steps (Romiszowski, 1988)

Sub-categories:
(a) Participation variables making interactional demands
- +/- open solution (Long, 1989)
- +/- one way flow (Pica et al., 1993)
- +/- convergent solution (Duff, 1986)
- +/- few participants (Crookes, 1986)
- +/- few contributions needed (McGrath, 1984)
- +/- negotiation not needed (Long, 1983)

(b) Participant variables making interactant demands
- +/- same proficiency (Yule & MacDonald, 1990)
- +/- same gender (Pica et al., 1991)
- +/- familiar (Plough & Gass, 1993)
- +/- shared content knowledge (Pica et al., 1993)
- +/- equal status and role (Yule & MacDonald, 1990)
- +/- shared cultural knowledge (Brindley, 1987)

Sub-categories:
(a) Ability variables and task relevant resource differentials
- h/l working memory (Mackey et al., 2002)
- h/l reasoning (Stanovich, 1999)
- h/l task-switching (Monsell, 2003)
- h/l aptitude (Robinson, 2005b)
- h/l field independence (Skehan, 1998)
- h/l mind-reading (Langdon et al., 2002)

(b) Affective variables and task relevant state-trait differentials
- h/l openness (Costa & Macrae, 1985)
- h/l control of emotion (Mayer et al., 2000)
- h/l task motivation (Dörnyei, 2002)
- h/l processing anxiety (MacIntyre & Gardner, 1994)
- h/l willingness to communicate (MacIntyre, 2002)
- h/l self-efficacy (Bandura, 1997)
2.5 Robinson’s Cognition Hypothesis and SSARC Model of Pedagogic Sequencing

The most recent version of the Cognition Hypothesis is the 2007 publication of the Triadic Componential Framework with Gilabert and the 2010 explanation of his SSARC model (SSARC refers to the steps taken in his Simple, Stable; Automatization; Restructuring; and then maximum Complexity sequence; see explanation below). The Triadic Componential Framework is essentially the same as the Pedagogic L2 Task Classification published in the same year, with very minimal changes. The main premise of the Cognition Hypothesis and SSARC model is that a sequencing of tasks be delivered to learners in an order of gradually increasing complexity, with the most simple version—on all available task demand parameters—being done first, and then eventually doing versions of the task that are more complex. (For example, starting learners out with tasks that allow planning time would precede a task that does not allow planning time).

With his colleague Gilabert (2007), Robinson provides a taxonomy of task features for teachers and researchers to use when designing tasks (reproduced in Table 5 below):
2.5.1 The Triadic Componential Framework

Table 5. The Triadic Componential Framework for Task Classification (from Robinson and Gilabert, 2007, p. 164):

<table>
<thead>
<tr>
<th>Task Complexity (Cognitive factors)</th>
<th>Task Condition (Interactive factors)</th>
<th>Task Difficulty (Learner factors)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Classification criteria: cognitive demands)</td>
<td>(Classification criteria: interactional demands)</td>
<td>(Classification criteria: ability requirements)</td>
</tr>
<tr>
<td>(Classification procedure: information-theoretic analyses)</td>
<td>(Classification procedure: behavior-descriptive analyses)</td>
<td>(Classification procedure: ability assessment analyses)</td>
</tr>
<tr>
<td><strong>a) Resource-directing variables</strong> making cognitive/conceptual demands</td>
<td><strong>a) Participation variables</strong> making interactional demands</td>
<td><strong>a) Ability variables</strong> and task-relevant resource differentials</td>
</tr>
<tr>
<td>+/- here and now</td>
<td>+/- open solution</td>
<td>h/l working memory</td>
</tr>
<tr>
<td>+/- few elements</td>
<td>+/- one-way flow</td>
<td>h/l reasoning</td>
</tr>
<tr>
<td>+/- spatial reasoning</td>
<td>+/- convergent solution</td>
<td>h/l task-switching</td>
</tr>
<tr>
<td>+/- causal reasoning</td>
<td>+/- few participants</td>
<td>h/l aptitude</td>
</tr>
<tr>
<td>+/- intentional reasoning</td>
<td>+/- few contributions needed</td>
<td>h/l field independence</td>
</tr>
<tr>
<td>+/- perspective-taking</td>
<td>+/- negotiation not needed</td>
<td>h/l mind/intention-reading</td>
</tr>
<tr>
<td><strong>b) Resource-dispersing variables</strong> making performative/procedural demands</td>
<td><strong>b) Participant variables</strong> making interactant demands</td>
<td><strong>b) Affective variables</strong> and task-relevant state-trait differentials</td>
</tr>
<tr>
<td>+/- planning time</td>
<td>+/- same proficiency</td>
<td>h/l openness to experience</td>
</tr>
<tr>
<td>+/- single task</td>
<td>+/- same gender</td>
<td>h/l control of emotion</td>
</tr>
<tr>
<td>+/- task structure</td>
<td>+/- familiar</td>
<td>h/l task motivation</td>
</tr>
<tr>
<td>+/- few steps</td>
<td>+/- shared content knowledge</td>
<td>h/l processing anxiety</td>
</tr>
<tr>
<td>+/- independency of steps</td>
<td>+/- equal status and role</td>
<td>h/l willingness to communicate</td>
</tr>
<tr>
<td>+/- prior knowledge</td>
<td>+/- shared cultural knowledge</td>
<td>h/l self-efficacy</td>
</tr>
</tbody>
</table>


There are two minor differences between this version and the L2 Pedagogic Task Classification framework Robinson provided in 2007b. First, empirical references following each variable have been removed for simplicity’s sake. Robinson also organized the subcategories more so as to better distinguish differences between classification criteria and procedures. Second, two
variables are more fine-tuned. What was previously “high/low mind reading” (e.g., Langdon, Coltheart, Ward, & Catts, 2002) under Ability Variables has been changed to “high/low mind/intention reading.” Under affective variables, what was previously “high/low openness” (e.g., Costa & Macrae, 1985) has been changed to “high/low openness to experience.”

In this Triadic Componential Framework, Robinson distinguishes between resource-dispersing and resource-directing variables in Task Complexity, or cognitive factors. Robinson argues that advancement in the sequencing of task design features along the resource-dispersing variables, such as +/- planning time, allows for increased automatization of a learner’s linguistic resources. Tasks along these lines will therefore promote a learner’s ability to perform the task, simulating the processing conditions that learners may encounter in the real world. Increases in complexity along resource-directing variables, on the other hand, such as +/- here and now (where a learner tells of an event in the here and now, versus having to tell of the event in the past, use deictic references, etc.) can promote attention to form-function mappings in the FL, because the learner may recognize where he/she has linguistic deficiencies, and so be more receptive to noticing feedback so that he/she can perform the demands of a cognitively complex task. Therefore, increases in complexity along these lines are argued by Robinson to promote interlanguage development. Robinson makes predictions between task complexity, task condition and task difficulty, as well as the effect of increasing task complexity on both general and specific measures of performance. He integrates both information-processing and interactionist accounts of L2 task effects (Long, 1996; Schmidt, 2001), and predicts that increasing the cognitive demands of tasks will push learners to a) greater accuracy and complexity, and b) promote interaction and heightened attention to and memory for input. Robinson’s taxonomy also takes into account important variables such as affective variables
(motivation, processing anxiety for example), ability variables (working memory and aptitude) and participant variables (same gender, power relationships, etc.).

2.5.2 Predictions of the Cognition Hypothesis

The main predictions of the Cognition Hypothesis are that increasing the cognitive complexity of tasks along certain dimensions will:

*In terms of L2 production:*

1) push learners to greater accuracy and complexity of L2 production in order to meet the greater functional/communicative demands they place on the learner;

*In terms of interaction, noticing, and uptake production:*

2) promote interaction and negotiation work, and heightened attention to, noticing of, and incorporation of forms made salient to the input; and

*In terms of individual differences:*

3) individual differences in cognitive abilities (e.g. working memory) and affective factors (e.g. anxiety) will increasingly affect task-based performance and learning as tasks increase in complexity” (Robinson, 2001b, p. 56)

2.5.3 SSARC Model

In his 2010 publication, Robinson provides three steps on how tasks might be ordered, recommending that tasks be sequenced along the cognitive demands they make and that increases in the complexity of resource-dispersing dimensions of a task be carried out before increases in the complexity of resource-directing dimensions. Therefore, as he explains, a “SSARC” model can be followed, where SS is ‘simple and stable’ task demands (first along resource-dispersing and then along resource-directing dimensions); A is automatization of the current interlanguage system; R is restructuring of this system, and C is the introduction of
“maximum complexity” to continue the restructuring and new form-function mappings, of the interlanguage so that development takes place:

SS (simple, stable task demands) + A (automatization) + R (restructuring) + C (max complexity)

In following the SSARC model as a guide for the ordering of tasks, alongside the task design dimensions proposed in the Triadic Componential Framework (Robinson & Gilabert, 2007), researchers can take into account learner cognitive and affective variables as well as findings from SLA, psychology and other fields. In addition, Robinson’s SSARC model and taxonomy for task design is to date the only proposal on task design research that considers internal and external factors in such a detailed way, making it perhaps the most comprehensive when compared to other proposed task design taxonomies.

2.6 Other Proposed Schemas for Designing and Investigating Tasks

Numerous taxonomies for the implementation and investigation of tasks have been proposed by other researchers in the field (see, e.g., Anderson & Lynch, 1988; Brindley, 1987; Brown, Anderson, Shillcock, & Yule, 1984; Byrnes, 1990; Candlin, 1987; Ellis, 2003; Honeyfield, 1993; Long, 1985; Nunan, 1988, 1989, 1994; Pica, Kanagy, & Falodun, 1993; Prabhu, 1987; Samuda & Bygate, 2008; Skehan, 1998; Willis, 1996). What, therefore, should teachers and researchers consider when deciding upon a taxonomy with which to design tasks? Different foci and scopes exist, and while ideally a taxonomy selection would be based on the teaching context in which the researcher/teacher is working and on the needs of the learning community, most schemas are based on the designer’s empirical approach to SLA. Willis (1996) based his pedagogic classification on the operations learners are required to engage in when carrying out a task. He distinguishes between listing, ordering and sorting, comparing, problem-
solving, sharing personal experiences, and creative tasks. Brynes (see Swales, 1990, p. 58) suggested that tasks be distinguished via a rhetorical classification, basing task classification on theories of rhetoric that distinguish different types of discourses. She looked at narratives, instruction-giving, descriptions, and reports. She also discusses genre-based tasks, where examples of tasks would be the writing of recipes, political speeches, job application letters, and leaving phone messages. Another example is that by Prabhu (1987), who proposed a cognitive classification, basing his task differences on the types of cognitive operations that tasks elicit. He distinguished between information-gap tasks, reasoning gap activities, and opinion gap activities.

Motivated by the Interaction Hypothesis, Pica, Kanagy, and Falodun (1993) proposed a psycholinguistic classification for tasks. Essentially, their task typology is based on the interactant role that the participant takes on during task completion. Pica et al. distinguish between a jigsaw task, an information-gap task, a problem solving task, a decision-making task, and an opinion exchange task. Distinguishing between *interactional activity* and *communication goal*, they also look at how much negotiation different task types can elicit. The authors highlighted the following:

- **Required exchange tasks:** tasks in which information has to be passed from one speaker to the other in order for the task to be completed, and
  - One-way tasks: one speaker has information to give to the other
  - Two-way tasks: both/all speakers have information to give and hence also have information to receive
- **Optional exchange tasks:** tasks in which information can be passed from one speaker to the other, although this is not required, and the actual information to be exchanged may not be provided by the task
• Convergent tasks: tasks in which all speakers are working to a joint agreed outcome
• Divergent tasks: tasks in which speakers can all come to different conclusions or outcomes

Samuda and Bygate (2008) suggested a task design taxonomy of variables that could arise in a classroom context. While certain variables could be refined more for operationalization purposes (i.e., “prototypically of task structure”), they are so far the only ones to incorporate teachers’ decision-making into the schema. Their schema is provided below:

**Potential Elements of Focus in Researching Second Language Pedagogic Tasks** (from Samuda & Bygate, 2008, p. 84)

<table>
<thead>
<tr>
<th><strong>Teachers’ procedures and processes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers’ pre-task preparation</td>
</tr>
<tr>
<td>Teachers’ post-task follow-up</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Learners’ procedures and processes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners’ talk; phases of talk; outcomes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Impact of elements of task design</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity of the task</td>
</tr>
<tr>
<td>Demands of the discourse type</td>
</tr>
<tr>
<td>Prototypically of task structure</td>
</tr>
<tr>
<td>Types of roles required by the task</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Impact of conditions of implementations</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarity with the task</td>
</tr>
<tr>
<td>Planning time</td>
</tr>
<tr>
<td>Repetition of the task</td>
</tr>
<tr>
<td>Distribution of roles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Dynamics around the task</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Negotiation for meaning</td>
</tr>
<tr>
<td>Processes of talk construction</td>
</tr>
<tr>
<td>Kinds of learner generated through task involvement</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Construals</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners’ interpretations of the task</td>
</tr>
<tr>
<td>Learners’ evaluations of the task</td>
</tr>
</tbody>
</table>
Teachers’ perceptions of the task

**Individual differences**
Learner differences: motivation, style, cultural background, age proficiency
Teacher differences: e.g. style, background, experience

**Task, lesson and curriculum**
Relationship between tasks and schemes of work/course plan
Relationship between tasks and target tests
Relationship between tasks and overall curriculum goals

Skehan (1996), similar to Robison, also proposed that task classification be based upon cognitive complexity, and highlights three main areas for the study of tasks: language, cognition, and performance condition. These are explicated with what he referred to as code complexity, cognitive complexity, and communicative stress. Skehan provided the following schema for task investigation:

(Skehan’s task complexity analysis, from Skehan & Foster, 2001, p. 194):

1. **Code complexity**
   - linguistic complexity and variety
   - vocabulary load and variety

2. **Cognitive complexity**
   - *cognitive familiarity*
     - familiarity of topic
     - familiarity of discourse genre
     - familiarity of task
   - *cognitive processing*
     - information organization
     - amount of ‘computation’
     - clarity of information given
     - sufficiency of information given

3. **Communicative stress**
   - time pressure
   - scale
     - number of participants
     - length of text used
   - modality
   - stakes
opportunities for control

As a final demonstration, Ellis (2003) distinguished between four design variables of tasks: input, condition, processes and outcomes. Coming from a pedagogical approach, he gives easy to hard versions of each of these four criteria:

**Criteria for Grading Tasks** (from Ellis, 2003, p. 228)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Easy</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A Input</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Medium</td>
<td>pictorial → written</td>
<td>oral</td>
</tr>
<tr>
<td>2. Code complexity</td>
<td>high frequency</td>
<td>low frequency</td>
</tr>
<tr>
<td></td>
<td>vocabulary; short and simple sentences</td>
<td>vocabulary; complex sentence structure</td>
</tr>
<tr>
<td>3. Cognitive complexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. information type</td>
<td>static → dynamic</td>
<td>abstract</td>
</tr>
<tr>
<td>b. amount of information</td>
<td>few elements/</td>
<td>many elements/</td>
</tr>
<tr>
<td></td>
<td>relationships</td>
<td>relationships</td>
</tr>
<tr>
<td>c. degree of structure</td>
<td>well-defined structure</td>
<td>little structure</td>
</tr>
<tr>
<td>d. context</td>
<td>here-and-now</td>
<td>there-and-then</td>
</tr>
<tr>
<td>dependency</td>
<td>orientation</td>
<td>orientation</td>
</tr>
<tr>
<td>4. Familiarity of information</td>
<td>familiar</td>
<td>unfamiliar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>B Conditions</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interactant relationship</td>
<td>two-way</td>
<td>one-way (negotiation of meaning)</td>
</tr>
<tr>
<td>2. Task demands</td>
<td>single task</td>
<td>dual task</td>
</tr>
<tr>
<td>3. Discourse mode required to perform the task</td>
<td>dialogic</td>
<td>monologic</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>C Processes</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cognition operations:</td>
<td>exchanging information → reasoning</td>
<td>exchanging opinions</td>
</tr>
<tr>
<td>a. type</td>
<td>exchanging information</td>
<td>exchanging opinions</td>
</tr>
<tr>
<td>b. reasoning need</td>
<td>few steps involved</td>
<td>many steps involved</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>D Outcomes</strong></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Medium</td>
<td>pictorial</td>
<td>written → oral</td>
</tr>
<tr>
<td>2. Scope</td>
<td>closed?</td>
<td>open?</td>
</tr>
<tr>
<td>3. Discourse mode of task outcome</td>
<td>lists, descriptions, narratives, classifications</td>
<td>instructions, arguments</td>
</tr>
</tbody>
</table>
2.7 Benefits of Operationalizing the Cognition Hypothesis

In briefly reviewing some of the other proposed task design taxonomies, one can see how exceptionally detailed the Triadic Componential Framework is. Robinson distinguishes between \textit{Task Complexity}, \textit{Task Condition}, and \textit{Task Difficulty}, and across each of these three categories, he gives thirty-six different features, basing them on ways in which they can be classified (i.e., information-theoretic analyses, behavior-descriptive analyses, ability assessment analyses).

Within the task complexity, or cognitive, factors, Robinson highlights both resource-directing variables that place cognitive/conceptual demands on the learner, and resource-dispersing variables, which place performative/procedural demands. Within the Task Condition category, participant variables that make interactional demands are listed, and include features such as +/- open solution and even +/- shared cultural knowledge. Lastly, the Task Difficulty category suggests ways in which both task-relevant resource differentials and task-relevant state-trait differentials may affect task outcomes. Robinson’s Triadic Componential Framework is advanced when compared to other proposed taxonomies not only in the amount of features it incorporates, but the fact that it considers so many aspects that can play a part in how pedagogic tasks might best lead SLA, and how different learners with unique abilities will carry tasks out.

The model is very ecologically valid as well, because its end goal is research that promotes the best learner-task matching so as to maximize SLA. Robinson bases his criteria on both theory, to include the Interaction Hypothesis, the Pushed Output Hypothesis, and the Noticing Hypothesis, and empirical findings from multiple fields, to include SLA, psychology, L1 studies, and even ergonomics. With nearly two decades leading to its development, it is all encompassing, superior in scope, and easily operationalizable by both teachers and researchers.
2.8 Summary of the Chapter

The aim of this chapter was to review Robinson’s research agenda during the past two decades and to provide a history of the development of the Cognition Hypothesis. A comparison of other well-known task design taxonomies was also given, as was an explanation for why the Cognition Hypothesis and Triadic Componential Framework were chosen for operationalization in the present study. In the next chapter, research that has specifically operationalized Robinson’s Cognition Hypothesis, looking at the effects of task complexity on interaction and on learning, will be reviewed, followed by a critique of the literature on four constructs that are operationalized in this dissertation: the individual difference working memory capacity and environment for interaction, specifically, face-to-face (FTF) versus computer-mediated communication (CMC). That chapter will also review how uptake, or the incorporation of forms from the input, and learning outcomes have been examined in both the FTF and CMC modes.
CHAPTER 3: RESEARCH ON THE TASK COMPLEXITY DIFFERENTIAL

The relationship between task design and L2 outcomes is currently an intense area of research with two main strands being explored: (1) tasks and how their design features differentially affect L2 oral language production; and (2) how tasks differentially mediate interaction and acquisition. Ellis (2003, p. 69) pointed out that both empirical strands are important: “we need to explore the relationship between task and language use on the one hand and language use and language acquisition on the other,” hence highlighting their indirect relationship. So far, the majority of research has focused on the first strand, typically investigating monologic tasks and how task design features elicit language production. Task-design variables of interest have included learner’s familiarity with the task (Plough & Gass, 1993); availability of planning time (Crookes, 1989; Ellis, 1987, 2005; Ellis & Yuan, 2003; Foster, 2001; Foster & Skehan, 1996; Ortega, 1999; Skehan & Foster, 1997, 1999, 2005); task rehearsal/ task repetition (Bygate, 1996, 2001; Bygate & Samuda, 2005; Gass, Mackey, Fernández & Álvarez-Torres, 1999; Lynch & Maclean, 2000, 2001; Sheppard, 2006; Tajima, 2003), split versus shared information (Newton & Kennedy, 1996); and dual task processing load (Skehan & Foster, 1999). Typically, dependent variables are measures of the complexity, accuracy and fluency of learners’ language production. While illusory to assume that complexity, accuracy and fluency are universal constructs (such as length of T-unit, c-unit, clauses per AS-unit, pauses per clause, etc.),\(^{10}\) results of these studies seem to indicate a high level of generalizability: “most would agree that ... different types of planning, repetition, familiarity, task structuring and task type are each likely to have a consistent effect on speakers whoever they are and whatever their learning context” (Samuda & Bygate, 2008, p. 95). The construct of

---

\(^{10}\) Other language phenomena are also in need of empirical investigation.
task complexity is also being explored in this arena (e.g., Gilabert, 2007; Gilabert, Barón, & Llanes, 2009; Kuiken et al., 2005; Kuiken & Vedder, 2008; Robinson 1995a; 2001a), with the theoretical premise that more cognitively complex tasks yield more complex and accurate language. This is because greater cognitive demands push learners to recognize where their linguistic deficiencies are, as posited by the Cognition Hypothesis (Robinson 2001b, 2005a, 2007b, 2010; Robinson & Gilabert, 2007). This is an important area of investigation given that one of the pedagogical goals of TBLT is to create tasks and sequence them to best help learners produce and acquire an L2/FL.

Needless to say, in the real world, language production is not typically monologic. Humans interact with other humans, and the very basis of communication is to relay information to others. Yet, very few studies have examined tasks in relation to dialogic interaction and increases in task complexity, looking at L2 development. Indeed, according to Ellis, “[t]asks that stimulate negotiation and through this provide comprehensible input and feedback and push learners to reformulate are the ones that will work best for acquisition” (2003, p. 80).

3.1 Studies Operationalizing the Cognition Hypothesis

According to Robinson’s Cognition Hypothesis, more cognitively complex tasks will result in more interaction work that can lead to SLA. In the current literature, twelve studies have pursued this strand of research: six have researched task complexity and its effect on conversational interaction within the framework of the Cognition Hypothesis, and six have looked at task complexity, interaction, and learning outcomes. All have been in either ESL or EFL contexts. The focus of this chapter will be to critically examine these studies, first reviewing
those that explored task complexity and its effect on interaction, and then the strand of research exploring task complexity, interaction, and opportunities for L2 development.

3.2 Strand 1: The Effect of Increases in Task Complexity on Interaction

Robinson (2001a) was one of the first studies to look at increases in cognitive complexity and how they affected interaction. The dependent variables in his study were measures of learner production, measures of interaction, and learner perceptions of task difficulty. Robinson also looked at different orders of task sequencing. 44 Japanese L1 learners of English were paired into dyads and were assigned the role of either speaker (information-giver) or hearer (information-receiver). Task complexity was operationalized via two dimensions: amount of information (more/less) and +/- prior knowledge. The task involved two different map tasks. The simple map version required that the speaker give directions from Point A to B with a map of a small area of their university campus, an area that participants were familiar with (hence, + prior knowledge). The complex map version covered a much larger version of the Nihombashi area of the city of Tokyo, and was an authentic street map. Participants were expected not to have prior knowledge of this area. The hearer role in both tasks had only point A on his/her map, and had to listen and draw the routes on the map as described by the speaker. Hearers were told that they could ask questions if they wanted to. All dyads performed both the simple and complex task versions in a reverse-ordered fashion: half did the simple task first, followed by the complex task, while the other half did the complex task first and then the simple task. This was so that Robinson could examine whether or not task sequencing (in a simple to complex order) would affect production and interaction, as his hypothesis predicts.

All speakers’ oral production was analyzed for measures of complexity, accuracy and fluency by examining their clauses per C-unit, error-free C-units, and words per C-unit.
All hearers’ oral production was analyzed for measures of interactional units: clarification requests and confirmation checks. After carrying out the task, participants filled out the difficulty perception questionnaire, which had a total of five items with 9-point Likert scale ratings (i.e., “I thought this task was easy (0) to “I thought this task was hard” (9)). Each item represented the following constructs: difficulty (sentence 1), stress (sentence 2), ability (sentence 3), interest (sentence 4) and motivation (sentence 5).

Robinson found an effect for task complexity on fluency and on lexical variety, with the speakers in the +complex task producing less fluent language and fewer words than during the -complex task. There was a trend towards greater accuracy (error-free C-units) in the more complex task, but it was not significant. For the hearer role, learners used significantly more confirmation checks during the more complex task, but not clarification requests. In looking at the task difficulty perception questionnaire, a repeated-measure ANOVA revealed that learners reported the more complex task as more difficult and stressful at the level of significance. They did not rate their level of ability differently in either task however, nor were there differences in their interest and motivation between the two tasks. This confirmed the hypothesis that the more cognitively complex caused less fluent and more accurate language, and also yielded more hearer interaction in one of the measures. Learners also perceived the more complex task as more difficult and stressful.

In looking at the effect of task sequencing, Robinson found that ordering the two map tasks from simple to complex did not significantly affect ratings of the difficulty of the complex version, but that it did affect the accuracy and fluency of speaker production. Robinson also examined the effect of role assignment, given that learners were assigned to be either speakers or
hearers. The hearers judged both the simple and complex tasks to be less difficult than the speakers, and were more confident in their ability than the speakers were.

This study supports the claims of the Cognition Hypothesis during task-based interaction, and was the to examine the effect of complexity and interaction in relation to perceptions on task difficulty. However, one potential limitation could be the fact that Robison used only three assessment measures, namely, one each for complexity, accuracy and fluency. One wonders if the findings could have been more robust had more measures of oral production been used. Also, he did not explain why interactional measures were only examined for the hearer role, and production measures only for the speaker role. While both the simple and complex map task inherently involved role assignment, one assumes that both dyad partners engaged in both interaction and production. It is not indicated in the study whether or not this was strictly controlled for. Indeed, not assigning roles may have been more natural, letting learners take on both roles. Another limitation to this study is the questionnaire instrument. Though measuring learner perception of task difficulty was a novel component of this study, concepts such as ‘difficulty,’ ‘stress,’ and even ‘motivation’ are heavily investigated constructs that typically involve entire questionnaires to gain insight into said construct. Robinson did not provide statistics that demonstrate the validity of the questionnaire or provide item-analysis. With only five items, the questionnaire is restricted in scope. Nevertheless, this study was the first attempt to measure the learner’s perception of task complexity as opposed to assuming it during interaction, an empirical component that is overall lacking when it comes to the construct of cognitive complexity.

A second study that investigated task complexity and learner-learner dyadic interaction was Nuevo (2006). In her unpublished dissertation (and in fact the first dissertation to
operationalize the Cognition Hypothesis for SLA), Nuevo measured both interaction features and L2 development among dialogic interaction with two different types of tasks. (For the purpose of this discussion, only results on the effects on interaction in her study will be discussed here).

Task complexity was operationalized by examining +/- reasoning demands, and Nuevo did so for two different tasks: a narrative task and a decision-making task. Each task naturally elicited two different linguistic items, the English past tense (narrative task) and locative prepositions (decision-making task). 113 learners of ESL were assigned to one of three groups: high-complexity, low-complexity, or no treatment (control group). A total of 10 measures of interaction, or L2 “learning opportunities,” were coded for: recasts and uptake, clarification requests, confirmation checks, comprehension checks, hypothesis formulation, metalinguistic talk, self-repair, noticing an interlanguage deficiency, and other repetition (such as imitation).

For the narrative task, learners had to collaboratively tell and write a story (past tense stories adapted from Sandra Heyer’s Easy True Stories) in three different treatment sessions. Each story had a different plot, and was designed to target the English past tense. During the treatment, learner pairs were given two worksheets. The first contained eight pictures, while the second was a worksheet with eight spaces, the time of each picture, and a space where the dyads were to write one sentence about the event. The -reasoning, or low-complexity, version of this task provided learners with numbered, or already-sequenced, ordering of the pictures. The + reasoning, or high-complexity version of this task included pictures that were not numbered. Each dyadic pair therefore had to work together to first put the pictures in order to ensure that they had the correct plot.

For the decision-making task, learners had to decide together where to seat people at a table for a dinner party (task 1), a work meeting (task 2), and a debate class (task 3) (adapted
from Penny Ur’s *Discussions That Work*). This task was designed to elicit locative prepositions. Materials included a picture of the tables with empty seats, onto which participants had to write their assignments, a sheet with an explanation of each character (and consequently any conflicts that characters might have if seated by a certain person), and a third sheet onto which they had to write their seating decisions. The -reasoning (low complexity) version of the task had six characters to seat and one obvious outcome for the seating, based on the descriptions of each character. The +reasoning (high complexity) version had ten characters to seat. Multiple outcomes were possible, and no single outcome could successfully meet the task requirement (seating everyone perfectly so that no conflicts would arise). This naturally elicited more reasoning.

For both the narrative and decision-making tasks, Nuevo found that the low-complexity group (-reasoning) has significantly more recasts, comprehension checks, and other-repetitions. For the narrative tasks, the high-complexity group (+reasoning) formulated hypotheses significantly more, but there was no effect for any of the other dependent measures. Nuevo explained that in general, “the low complexity group used significantly more learning opportunities that focused on syntax than the high complexity group” (p. 122). In sum, no conclusive findings were achieved in regards to the effects of task complexity on measures of interaction, and Nuevo even found evidence that appears contrary to the predictions of the Cognition Hypothesis. Some of these findings may be due to some inherent limitations as discussed below.

First, an overwhelming majority of participants in this study were native Spanish speakers and, as Nuevo reported, those L1 Spanish speakers who were paired with other Spanish speakers engaged in a significant amount of metalinguistic talk in their L1 together, something
she was not able to control for. Other dyadic pairs who did not share L1s were obviously not able to do so. It is indisputable that being able to engage in L1-meta-talk impacts the nature of the task’s interaction and outcome, so the fact that some learners were able to engage in their L1 task talk while others could have impacted the nature of the task. Second, Nuevo’s study underscored the importance of verifying group comparability before assigning them to study cells. ANOVAs in her study revealed significant differences between groups on the pretest, with high-complexity group having significantly higher pretest scores than the low complexity group for the past tense, and the control group having significantly higher scores on locative preposition knowledge than the low-complexity group. While it is a standard practice in the field of SLA to counterbalance tests and then run follow-up analyses when the experiment is over, this study showed the importance of ensuring test and group parallelism prior to the study. This is also a limitation that could have also affected results. (In addition, 30% of the participants in Nuevo’s study were missing for the delayed posttests.) Third, it is interesting that Nuevo blended oral interaction with writing in the narrative task. Learners had to agree on what they were going to write down together, and one questions if the use of two modes for conversation influenced the task outcome. A multiplicity of research shows that writing is a different kind of modality than speaking, with greater online-planning going during the process of writing (e.g., Abrams, 2003; Adams, 2006; Payne & Whitney, 2002; Pellettieri, 2000). Other researchers have reported high amounts of negotiation for meaning in tasks that involve collaborative writing as Nuevo’s participants did (Adams, 2003; Storch, 2001; Swain & Lapkin, 1998, 2001). Given that Nuevo only employed oral productive posttests and grammaticality judgments, keeping everything in the same modality might have yielded different results. (In other words, keeping the task as one that was strictly in the oral mode would have ensured modality as a single, monolithic feature.)
Nuevo’s study therefore poses some unique empirical questions about modality and collaborative writing in a dialogic task. Finally, dyads were comprised of learners paired with other learners in this study. As part of Robinson’s Cognition Hypothesis premises, increases in task complexity will allow for more forms made salient in the task-input through proactive (e.g., premodified input flood) and reactive (e.g., recasts) focus on form techniques (Robinson, 2007). Given that it has been empirically demonstrated that interaction between learners and other learners differs from interaction between learners and native speakers and/or expert interlocutors, (Gass & Varonis, 1989; Mackey, Oliver, & Leeman, 2003), it may be the case that learners of the same proficiency level were incapable of providing corrective feedback or were not seen as beneficial or credible providers of feedback by their partners. Pairing learners with an expert as opposed to someone of their same proficiency level may be the necessary task implementation feature when learning of some linguistic form is the desired outcome; this may also have affected measures of interaction differently. Limitations aside, Nuevo’s study paved the way for more research on cognitive complexity and interaction because it was the first to operationalize +/- reasoning embedded in conversational interaction.

Robinson (2007a) also paired learners with other learners (specifically, Japanese L1 learners of EFL) and gave them three interactive tasks that increased in the complexity of resource-directing reasoning demands. Robinson set out to test the claims of his hypothesis that tasks of increasing cognitive complexity would a) result in more accurate and complex, but less fluent, speech production and b) result in more interaction and uptake of linguistic forms necessary to be able to carry out the demands of the task “when these are made salient in the input” (p. 194). For the purpose of the present discussion, only results on interaction measures, and not production, will be discussed here. (Further detail on this study is provided in Chapter 2).
The task complexity variable operationalized in this study was +/- intentional reasoning, or having to reason about the motives, beliefs and thoughts that cause people to do actions. 42 Japanese L1 learners of EFL were assigned to the roles of speaker/storyteller or listener/sequencer, and were paired together to perform three narrative tasks at simple, medium, and complex levels of intentional reasoning demand. This study was therefore the first to explore task complexity along a continuum instead of a dichotomy. The ordering of the tasks was counterbalanced across all dyads. The task required the speaker/storyteller to decide on the correct ordering of pictures that depicted an event, and then relay the story to the partner, the listener/sequencer, so that he/she could arrange his/her pictures in the correct order based on what the partner said. The pictures were from the Wechsler Adult Intelligence Scale, and Robinson used the first, fifth and eighth levels of complexity to serve as the simple, middle and most complex versions of the task. Measures of interaction were confirmation checks, clarification requests, and number of turns. Uptake was also explored and coded for all instances of exact uptake, i.e., if a learner used the exact same form given to him/her by the speaker and partial uptake, i.e., if the learner used the phrase but omitted or changed elements in the phrase. Robinson also looked at amount of exact uptake per turn (UEPT) and amount of partial uptake per turn (UPPT). As with his 2001a study, learners’ perception of task difficulty was measures with the same 5-item instrument employed in his 2001 study (with a 9-point scale), as well as with a separate questionnaire to measure learners’ anxiety during three different phases of task performance: input, processing and output (the scale was created by MacIntyre & Gardner, 1994). Robinson correlated learners’ anxiety on the three different processing stages to the dependent variables of speech production, interaction and uptake. Lastly, Robinson also examined some specific measures of cognitive complexity, such as reference to psychological
state terms, cognitive states, and emotional terms (following Lee & Rescorla, 2002), as well as the ratio of these measures per turn.

In examining the general measures of oral production, Robinson did not find any significant differences between the task types (simple, medium or complex). However, he did find that in examining specific measures (psychological, cognitive and emotional state terminology and reference), the more complex tasks resulted in more reference to such terms, with the complex task yielding the most reference to cognitive state terms. For measures of interaction, Robinson’s hypotheses were confirmed: there were significantly more turns during the complex task, as well as significantly more clarification requests and confirmation checks. There was also significantly more partial uptake and partial uptake per turn across tasks (complex > medium > simple), but not exact uptake. In regards to learner perceptions of the task, the most complex was seen as significantly more difficult and stressful than the medium and simple task versions. Participants also indicated that their confidence was the lowest on the more complex task. An interesting component of this study was the association of anxiety (via MacIntyre & Gardner’s 1994 instrument) at different levels of production. Robinson found that anxiety levels affect output but not input or processing. Robinson even divided learners into low/high output anxiety groups, and found that the low anxiety group, contrary to the high anxiety group, was induced to produce more complex speech with the more complex task demands.

So far, this study provides the most direct evidence for Robinson’s Cognition Hypothesis, and makes an interesting call to further explore general versus specific measures of complexity. In this study, learners viewed the most complex as most difficult, individual differences in anxiety most affected their performance in the most cognitively demanding task, and there was
more interaction and partial uptake as tasks increased in complexity. Also, the fact that Robinson used MacIntyre and Garnder’s validated instrument to measure anxiety levels at the three different stages of L2 production (input, processing and output), and correlated these measures with production and interaction was a novel addition to this area of study.

What would have been interesting is if Robinson had also explored learning, for example, of some of the lexical items provided in the task (“holding a plank,” for example), and whether or not the hearer/sequencer learned these items from his/her partner best in the more complex task. As with the 2001(a) study, allowing participants to carry out the task naturally, instead of assigning roles, might have yielded different results. Also, the 5-item questionnaire attesting the learners’ perception of task difficulty is once again a limitation. As explained above, no instrument reliability nor item analysis was reported for this questionnaire. While the MacIntyre and Gardner scale is highly reliable (and highly attested in the literature), one questions whether the 5-item scale by Robinson does indeed measure what he set out to measure in a robust way. Also, the questionnaire is not provided in the appendices. Instrument aside, the design, task complexity continuum, and dependent variables explored in this study were insightful.

Kim (2009a) was the first to explore the effects of task complexity alongside proficiency level with learner-learner interaction in a natural classroom environment. Kim explored the task complexity differential in relation to the occurrence and resolution of language-related episodes\(^1\) (LREs), or learning opportunities that arise during interaction. LREs are a type of negotiation for meaning, induced by learners themselves, and can provide a context for focusing on form. The task complexity features explored in this study were +/- reasoning and +/- few elements. 34

\(^1\)Swain and Lampkin (2001, p. 104) define LREs as parts of discourse “where students talk about language they are producing, question their language use, or other- or self-correct their language production.”
learners of ESL were paired into dyads of either low-low proficiency or high-high proficiency, and carried out four interactive tasks: two picture narration tasks, reflecting the +/- reasoning demands, and two picture difference tasks, which reflected the +/- few elements. The picture narration tasks differed in that participants either had to decide on the correct sequence of 8 pictures versus simply coming up with a story based on pre-sequenced pictures. For the picture difference tasks, participants had to agree on the differences in a beach scene, with one scene having more elements than the other, making it more difficult. Tasks were conducted in class, and learners had 15 minutes to do each task. Following Robinson 2001a and 2007a, Kim gave learners a questionnaire once they had finished all tasks to gain insight on their perception of the complexity and learning opportunities of the four tasks. The questionnaire had a total of 6 Likert-scale items with 9 scales each, measuring perceptions on task difficulty, stress, perceived ability, interest, and motivation, as well as questions as to the helpfulness of the task and learners’ view of working with a partner.

All interactions were transcribed and analyzed for the occurrence of lexical and grammatical LREs, to include requests for assistance, negotiation sequences, and explicit/implicit feedback. Then, each LRE was analyzed for a resolution: correctly resolved, unresolved, and incorrectly resolved. Kim found that the low proficiency dyads produced significantly more LREs during the simple picture narration task, but the opposite occurred for the high proficiency dyads; they produced significantly more LREs during the complex version. For the picture difference task (+/- few elements), significance was found only for the low proficiency-level learners; they produced significantly more LREs during the complex task version. All dyads types correctly resolved more LREs during the complex tasks than the simple tasks. Interestingly, the total amount of LREs was positively correlated with learners’ perception
of the task as providing learning opportunities, measured via item 6 on the questionnaire. In terms of the questionnaire data, learners’ perception of the task difficulty was not discussed in this study. Kim suggests that the fact that low proficiency level-learners produced more LREs in the less complex picture narration task (recall that this finding is similar to Nuevo, 2006) is because less complex tasks permit the use of more attentional resources with which learners can monitor their partners’ output and provide feedback. She therefore found partial support for the Cognition Hypothesis in the picture narration task, at least with regard for the high proficiency learners (not for those of lower proficiency). For higher proficiency learners, the more complex task led to more interaction and negotiation for meaning. The Cognition Hypothesis was supported, however, for lower proficiency learners in the picture difference task.

Kim’s study was the first to look at task complexity, interaction, and how learner proficiency level moderates these variables. She found that task complexity affected the way that learners interact with one another and respond to LREs. Her study also showed that learners are capable of providing each other with feedback, and together, can resolve LREs. It would have been interesting too had Kim looked at learning outcomes. One limitation to the study is the fact that the questionnaire was given to learners after all tasks had been completed as opposed to being disbursed immediately after each individual task. Giving learners each questionnaire immediately after carrying out each task would have more robustly captured their perception of that task’s difficulty level. In addition, the researcher did not report all of the results of the difficulty questionnaire, such as level of perceived difficulty (item 4 on the instrument). Still, the study was the first of its kind to associate task complexity with learner-learner derived LREs alongside proficiency levels that arise naturally inside the classroom, making it ecologically valid.
Révész (in press, a) also examined the effect of task complexity on LREs during learner-learner dialogic production in the classroom. In addition, she explored the potentiality of three individual differences (IDs) to moderate interaction: linguistic self-confidence, anxiety, and self-perceived communicative competence. 43 high-intermediate and advanced-level learners of ESL worked in self-selected groups (3-4 people per group) during English class. Like Kim (2009), Révész operationalized the +/- reasoning and +/- few elements resource-directing dimensions to increase task complexity. Like Robinson (2007b), she used both general and specific measures of L2 oral speech. For speech production measures, the dependent variables were AS-units, values of D, error-free AS-units, error repairs (both as ratios), serving as the general measures; the specific measure was participants’ use of conjoined clauses. For interactional measures, the dependent variables were LREs, confirmation checks, clarification requests, recasts, and metalinguistic talk. The task that learners had to do was to decide on community recipients for a NYC trust fund. One fund was valued at $10,000,000 dollars and could be distributed to six city programs. The other was worth half of that sum, $500,000 and could be given to three programs. Both the complex and simple tasks were decision-making tasks and were convergent in nature. Each task was carried out in three phrases in the classroom. For the first five minutes, learners individually considered their own priorities for the task. Then, they negotiated in groups for 15 or 20 minutes (simple or complex, respectively) to make a final decision. The last phase was a presentation to the class by each group (10-15 minutes). Learners did the simple task first and then the complex task, so that everything was done in one class session. At the end of the lesson, both the learners and the teachers filled out a second questionnaire that gauged their perception of the difficulty of the task.
Révész found that participants produced more lexically diverse and accurate language when carrying out the more cognitively complex task. However, their L2 oral production was significantly less complex. Learners also produced more conjoined clauses (the specific measure) during the more complex task. In terms of interaction, the more complex task resulted in significantly more production of LREs and metalinguistic task. There was no significant difference in number of clarification requests, confirmation checks, or recasts. When analyzing the questionnaire data, it was found that none of the correlations between IDs and measures of speech production or of interaction-driven language learning opportunities were significant, counter to the predictions of the Cognition Hypothesis. The task perception questionnaire however did reveal that both students and teachers rated the more complex task as more difficult. They also rated the more complex task as more useful for learning, and as more interesting. Learners further reported that they felt more stressed during the cognitively complex task compared to the simple task; this was in contrast to the teachers’ questionnaires, which reported that students were “equally relaxed during the two tasks.” Learners also reported on the questionnaire that the more complex led them to pay more attention to their linguistic output.

One major advantage of Révész’s study, as with Kim’s (2009) study, is its high ecological validity, given that it was conducted in the classroom. While the pre-paired learners and dyads stayed the same in Kim’s study, Révész’s learners worked in self-selected groups in the classroom, something that often happens inside a classroom environment. Révész also explains the pedagogical relevance of investigating task complexity and its effect on LREs:

“As several researchers have noted, errors in communicatively redundant, perceptually nonsalient L2 constructions, which are notoriously difficult for learners to acquire on their own … rarely cause communications failure… Hence, it would appear that task
features with a capacity to promote the incidence of LREs unaccompanied by comprehension problems may have an especially great value in providing opportunities for improving grammatical competence; because such LREs, being solely triggered by problems in the code, might be more likely to focus attention on L2 constructions which are non-meaning bearing and non-salient and thus less noticeable."

Révész is the first researcher exploring IDs in relation to interaction with tasks of varying degrees of complexity. She used instruments that were created by other researchers and had been attested as valid measures of linguistic self-confidence, anxiety, and self-perceived communicative competence. Révész also provided Cronbach’s alpha correlates for the questionnaire items, improving the construct validity of the measures. As she herself said, the fact that learner IDs were not correlated with their task performance is “puzzling;” indeed, more research into the effect of IDs on task interaction and learning is necessary. The results of this study thus do not fully support the Cognition Hypothesis.

Gilabert, Barón, and Llanes (2009) were the first to examine increases in cognitive complexity across three different types of tasks, looking at how the different task and complexity types affect interaction. They examined +/- here-and-now, +/- few elements, and +/- causal reasoning, hence, six different tasks total. 54 learners of EFL were paired into dyads and carried out all six tasks in a Latin square\textsuperscript{12} design. The three tasks were a narrative task, an instruction-giving task, and a decision-making task. For the narrative task, learners were given a comic strip and had to narrate the story shown either in the present tense while looking at the comic, or in the past tense while not being able to look at the comic. For the instruction-giving task, learners had to tell a partner

\textsuperscript{12} A Latin square design is used to control for variation. For example, an experimental treatment is assigned randomly across rows and columns (hence the ‘squares’); each experimental treatment is repeated once per column and per row. (I.e., ABCD [first row], BCDA [second row], and on).
how to follow a map. The simple version required that learner 1 tell his partner where/how to go on a route-marked city map. Movement involved only the left, right, and straight directions. The complex version involved more complex stops on the map and multiple axes of direction, such as left, right, straight, up or down the stairs, front to the back of a store, etc. Finally, the decision-making task was the “Fire Chief” task, in which learners must make decisions on who will be rescued from a burning building, how, and in what order. The complex version of this task involved unique circumstances, such as dealing with an elderly person and an injured person. All tasks (both the simple and complex version of all three task types) were carried out in one single session. All interactions were transcribed and coded for the following measures of interaction: confirmation checks, clarification requests, comprehension checks, recasts, LREs, and repairs. Learners also filled out a questionnaire in which they indicated their perception of the task (difficulty, stress, confidence, interest, and motivation—the same questionnaire used in Robinson 2001a and 2007a). The questionnaire was administered after each set of two tasks had been completed (i.e., both the simple and complex versions of each task).

The researchers found that the more complex versions tasks generated more interactional moves and opportunities for negotiation for meaning. However, there were slight differences according to task type. For example, the complex version of the narrative task resulted in significantly more clarification requests, but significantly lower confirmation checks. There were significantly more LREs and repairs, but not recasts, and these same findings held for the map task as well. The map task yielded significantly more confirmation and comprehension checks. In fact, comprehension checks were largely exclusive to the map task. The decision-making task however only showed an effect for task complexity in terms of number of repairs. Number of repairs was the only dependent variable that was significantly higher on all three complex
versions of the tasks when compared to simple versions. Other variables were differentially affected by task type. Also, learners rated the complex versions of each task as more difficult, stressful, and affecting confidence levels when compared to the simple versions. There was no effect for interest and motivation—identical to the results found by Robinson (2007a).

This study provides support for the Cognition Hypothesis in that, for the most part, more cognitively complex tasks results in more interactional moves. However, the researchers found that this may be dependent on task type. As the researchers explain, one reason for these results, and for the fact that the dependent variables did not hold across task types, was the open versus closed nature of the tasks. Both the map and narrative tasks were closed in nature, while the decision-making task had no real ‘correct’ solution. Had the researchers kept this task condition constant, they might have had more comparable results. At the same time, this finding poses interesting questions into how goals of a task mediate interaction (i.e., open vs. closed, convergent vs. divergent, etc.). Also, it was reported that the decision-making task generated other types of pragmatic moves, such as suggestions, agreeing vs. not agreeing. The researchers call for more research that explores the effects of task complexity on other interactional features besides those examined in their study.

To sum up this section on studies that have operationalized the Cognition Hypothesis and its effect on dialogic interaction, Table 6 below is provided that highlights the main points of each study:
<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>TASK COMPLEXITY FACTOR</th>
<th>DYAD TYPE</th>
<th>TASK</th>
<th>DEPENDENT VARIABLES OF INTERACTION</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robinson (2001a)</td>
<td>+/- few elements (Resource-directing)</td>
<td>EFL Learners paired with learners</td>
<td>Map tasks: get partner from A to B on map of area of campus or of area in Nihombashi section of Tokyo</td>
<td>Confirmation checks, Clarification requests</td>
<td>Sig. more confirmation checks, but not clarification requests. Sequencing of tasks (simple to complex) sig. affected accuracy and fluency</td>
</tr>
<tr>
<td>Nuevo (2006)</td>
<td>+/- causal reasoning (Resource-directing)</td>
<td>ESL Learners paired with learners</td>
<td>1) Narrative task: ordering of pictures 2) Decision-making task: seating arrangements of people w/ unique personalities</td>
<td>Recasts, uptake, clarification requests, confirmation checks, comprehension checks, hypothesis formulation, metalinguistic talk, self-repair, noticing interlanguage deficiency, other repetition</td>
<td>-Causal reasoning resulted in sig. more recasts, compr. checks, and other-repetitions in both tasks. +Causal reasoning resulted in sig. more hypothesis formation in narrative task only</td>
</tr>
<tr>
<td>Robinson (2007a)</td>
<td>+/- intentional reasoning (Resource-directing)</td>
<td>EFL Learners paired with learners</td>
<td>Simple, Medium or Complex picture sequencing and telling (person 1); arranging of pictures based on what partner says (person 2).</td>
<td>Confirmation checks, clarification requests, number of turns. Exact and Partial Uptake</td>
<td>Sig. more turns, clarification requests, confirmation checks</td>
</tr>
</tbody>
</table>
(Table 6 continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Methodology</th>
<th>Tasks</th>
<th>Measures</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim (2009a)</td>
<td>+/- reasoning, +/- few elements (Resource-directing)</td>
<td>ESL Learners with learners during class; L vs. H proficiency</td>
<td>2 picture narration tasks (order vs. tell a story) and 2 picture difference tasks (beach scene with more or less elements)</td>
<td>More LREs for H prof. learners in + reasoning picture narration; more LREs for L prof. learners in +more elements picture diff. task. More LREs resolved in more complex tasks</td>
</tr>
<tr>
<td>Gilabert, Barón, Llanes (2009)</td>
<td>+/- here-and-now, +/- few elements, +/- causal reasoning (Resource-directing)</td>
<td>EFL Learners paired with learners</td>
<td>Narrative task, instruction-giving task, and decision-making task (6 versions total)</td>
<td>Confirmation checks, clarification requests, comprehension checks, recasts, LREs, repairs</td>
</tr>
<tr>
<td>Révész (in press, a)</td>
<td>+/- reasoning, +/- few elements (Resource-directing)</td>
<td>ESL Learners with learners. Self-selected groups of 3-4 in classroom.</td>
<td>Deciding upon community recipients for a NYC trust fund (10 million dollars vs. half a million, 3 recipients vs. 6)</td>
<td>LREs, confirmation checks, clarification requests, recasts, and metalinguistic talk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IDs of Lx self-confidence, anxiety, self-perceived comm. competence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Significantly more LREs and metalinguistic talk in +complex task; no sig. difference for cc’s, cr’s, or recasts. No conclusive correlations between IDs and L2 speech production or interaction</td>
</tr>
</tbody>
</table>
3.2.1 Summary of Cognitive Complexity and the Effect on Interaction

All of these six studies examined the effects of increases in task complexity on interaction. All dealt with either EFL or ESL contexts. Four were laboratory studies (Gilabert et al., 2009; Nuevo, 2006; Robinson, 2001a, 2007b) while two were classroom-based (Kim, 2009; Révész, in press, a). The task complexity features explored were +/-causal reasoning, +/-few elements, +/-here-and-now, and +/-prior knowledge. So far, and with the exception of Nuevo (2006), this group of studies provides support for the effects of increases in task complexity on dialogic interaction, showing greater number of LREs produced by learners and more interactional moves. However, task type (i.e., closed. vs. open) may differentially mediate the types of interactional moves, as Gilabert et al. found. All of these studies show that learners reported the more cognitive complex tasks as more difficult and stressful; interestingly, Révész’s study showed that teachers may view tasks differently than learners. How individual differences mediate speech production on tasks of different complexity levels is inclusive, as Révész found no correlation between variables. Clearly, more research is needed in this area. Also, Nuevo (2006) and Révész (in press, a) found that certain interactional measures, to include recasts, do not make a difference depending on task complexity. Whether the learner sees his/her partner as a beneficial provider of feedback seems to be a critical factor in this outcome. Indeed, as will be seen in the next session, providing learners with expert interlocutors may make a difference. The next six studies explore the effects of task complexity on learning outcomes, and are described below.
3.3 Strand 2: The Effect of Increases in Task Complexity on L2 Development

So far, only six studies have examined the effects of task complexity on L2 learning outcomes (two of which are currently in press). Nuevo (2006), in addition to measuring interactional moves (described above), also measured learning gains in two different task types (narrative and decision-making) by looking at +/- causal reasoning. 113 learners of ESL were assigned to one of three groups: high-complexity, low-complexity, or no treatment (control group). To measure L2 development, an oral production test and a grammaticality judgment test were used in a pretest, posttest, and delayed posttest design. The linguistic items of focus were the English past tense and locative prepositions. Nuevo found no difference among groups in development of the past tense (in fact, all groups achieved similar gains on the delayed posttests). The same was found for learning gains made on locative prepositions—no group learned statistically more than the other, with everyone making gains on the posttests. The only significant finding was that the complex group performed better than the control; there was no statistically significant difference between the complex and -complex group. Nuevo therefore did not find evidence for the Cognition Hypothesis in terms of L2 development, similar to her lack of findings in regards to the effects of task complexity on interaction. However, and in addition to some of the other possible limitations outlined above, this may have been due to the fact that learners interacted with learners, and so did not view their partners as legitimate providers of feedback. It may be the case that pairing learners with expert interlocutors for linguistic items that are less salient is required to achieve effects for SLA (e.g., Swain, 2005). Also, Nuevo reported that pre-tests scores showed that groups were not comparable at the onset of the study, a finding she was unable to control until after the experiment. The lack of group comparability may have also affected learning outcomes.
A study by Révész and Han (2006) explored the variables of task content familiarity and task type on the efficacy of recasts. Unlike Nuevo, Révész and Han paired learners with an expert interlocutor, the researcher, who attempted to draw their attention to form by providing learners with implicit corrective feedback in the form of recasts during the dialogic interaction. Révész and Han sought to operationalize the component of Robinson’s theory claiming that increasing the task complexity along the resource-directing dimension will “increase the likelihood of detecting and selecting seeded aspects of the input …” (Robinson, 2003, p. 650, taken from Révész & Han, 2006, p. 163). 36 adult learners of ESL were randomly assigned to one of four groups: (1) same video group, (2) different video group, (3) same notes group, and (4) different notes group. The linguistic item being investigated was the past progressive form, i.e., Andrea was walking. In a pretest, posttest, and delayed posttest design, participants engaged in a total of three treatments, where they either saw the same video or a different video for all three sessions, or, received the same notes or novel notes during each session. The task was to assist the police in assessing a robbery scene from the day before. Each task was designed to naturally elicit the past progressive tense. In the video groups, learners were shown 15 silent video scenes each 15 seconds in duration, of the criminal engaging in activities, i.e., smoking or running. The participants had to describe the videos as they were watching them. In the notes groups, participants were shown notes (as opposed to video content) via PowerPoint that corresponded to 15 different times of the day. They were told that they had been taking notes while following the criminal and so had to report the actions to the police. Notes included “10:00 am” and “in the library,” hence priming the learner to tell what the criminal was doing at that time. The between-groups difference was that the same groups saw the same video or were given the same notes for each session, while the different groups had novel videos or notes for their
sessions. Recasts were provided by the researcher to the learner whenever there were errors made on the past progressive, and were kept constant across all four groups to explore how their efficacy was mediated by the task design variables of task familiarity and task type. There were also two “manipulation checks” done during the experiment, with one occurring right after the first session for the different notes/video groups, and one occurring right after the third session for the same video/notes groups. For this check, the researcher asked learners to recall as many scenes or notes from their session as possible. Learning gains were measured via a fill-in-the-blank task, a written description task, a video-primed oral task, and a notes-primed oral task in a split-block design.

Révész and Han found significant main effects for both of the independent variables. First, learners who received recasts with tasks of the same content performed better than tasks with new content. Second, learners who received recasts through video-primed tasks performed better than notes-primed. The effects were found for mostly oral, as opposed to written, production however. There was no significant effect for either task content familiarity or task type when looking at learning measured with the fill-in-the-blank task. The researchers interpret these findings as demonstrating that when learners are familiar with the content of a task, they will notice recasts more (i.e., recasts can be more effective). They conclude by suggesting that the notes group probably did poorer than the video group because they were engaging in a dual task. As Révész and Han explained, those participants had to read notes, think of a scene, and communicate actions in the past tense. The notes groups therefore engaged in a resource-depleting activity, and so therefore weren’t as able to attend to the recasts. The less complex task (video and familiar content) attended to recasts best. This runs contrary to what the Cognition Hypothesis predicts: that increases in the cognitive complexity of tasks will “increase the
likelihood of detecting and selecting seeded aspects of the input,” which indicates that feedback would be more effective in the more complex tasks.

Révész and Han was the first study that looked at the efficacy of recasts alongside task complexity. Their study was pioneer in that it raised some interesting questions as to how learners process feedback while engaging in tasks of different levels of complexity. However, the study does not go without limitations, as the authors themselves acknowledge. First, in trying to establish the lack of findings of the fill-in-the-blank task, post-hoc analyses revealed that “previous formal learning experience” differentiated some participants from others, with never having taken an English class before being strongly related to low performance on the assessment. This meant that previous formal learning experience was an affecting variable on gains in on the assessment, begging the question of group comparability. Also, the number of participants in each group was low, with an $n$ of only 9 per cell. In addition, according to the authors, a major limitation of their study was the fact that recasts were kept constant. They explained that since recasts served as a dependent variable instead of as a covariate, any interpretations of findings would not truly be able to measure recasts as a “catalyst of change” (p. 176), or in other words, assuming rather than confirming that recasts play a positive role. However, this may simply mean that all four of the treatment types worked, and arguably, the study revealed conditions under which recasts may be more effect. In fact, significant differences between each of the groups were not reported. While some gains made on three of the assessment measures were higher for some groups (e.g., the same video group), keeping recasts constant may have simply allowed the researchers to explore and isolate task variables that have the potential to contribute to the efficacy of recasts during interaction.
Instead, what might be argued to be a limitation of this study is the way they operationalize complexity. It seems to be the case that the notes versus video is +/- dual processing or here-and-now versus there-and-then. The fact that two different modalities were used (that of note reading versus video) seems to be under-acknowledged. In explaining why the notes group did worse than the video groups, Révész and Han state: “[t]o complete the notes-primed tasks, participants needed to undergo at least three major cognitive processes: decoding notes, formulating thoughts, and encoding them … Though this is also largely true for the participants undertaking the video-primed tasks, the process of decoding in the latter case could be more straightforward …” (p. 174). In her dissertation (2007) and 2009 publication, Révész improved this component of Révész and Han (2006) by specifically exploring recasts alongside the here-and-now vs. there-and-then dimension of resource-directing variables in task design.

Révész (2009) examined the relationship between task complexity and learning outcomes by pairing learners with an expert interlocutor (the researcher) and explored how task complexity modulates the effects of recasts. Using the Cognition Hypothesis as the theoretical foundation, Révész specifically sought to address the prediction made by Robinson and Gilabert (2007), where it is claimed that increases in task demands “will lead to more extensive noticing, a greater amount of uptake, and longer term retention of input made salient by recasts” (2009, p. 442). She operationalized task complexity as +/- contextual support by providing learners with pictures while describing a robbery scene or not (with the idea that not having a photo would be more complex). This also served as the +/- here-and-now dimension, because the -contextual support, or the there-and-then, was more complex given that learners had to remember the story. Révész hypothesized that the more complex task (-contextual support/ -here-and-now) alongside recasts would lead to greater efficacy of recasts and learning outcomes, as is predicted by the Cognition
Hypothesis. She explained: “to use Robinson’s (2001, 2005) terminology, the +/- here-and-now dimension of a task is resource-directing, because it has the capacity to draw learner’s attention to specific L2 forms” (Révész, 2009, p. 442).

90 learners of EFL (Hungarian L1) were divided into one of five groups: the control group or one of four experimental groups, differing in terms of +/- contextual support and whether or not they received corrective feedback (recasts) from the researcher. The targeted linguistic item was the English past progressive. Post-tasks consisted of an oral descriptive task with and without a photo, a written descriptive task, a fill-in-the-gap task, and a grammaticality judgment test. The task was to describe photos of scenes in New York City at the time that a robbery had happened. Naturally eliciting the past progressive, learners had to describe the photos to the researcher, who took on the role of a police officer (i.e., “he was running”). The task prompts were delivered via Microsoft PowerPoint. All learners were shown a photo with a title of the area in the city and the time, such as 11:00 am. After 10 seconds, the picture disappeared for those in the -photo group. They had to describe the scene as best as they could without being able to look at the photo. Those in the +photo group were able to look at the photo while describing it (all had 40 seconds to do so). Learners in the +recasts groups received recasts on all erroneous uses of the past progressive, while those in the -recasts group received no feedback. Recasts were of the simple, isolated type, without any emphatic stress to the linguistic form being investigated. All learners in the four experimental groups partook in a total of three sessions. Révész then issued a questionnaire after the treatment to gain insight into learners’ perception of the task.

Révész found that the learners who received recasts and did not have contextual support (i.e., -here-and-now) achieved the greatest L2 gains, outperforming those who received recasts
and were able to view the photos at the same time. Those learners who received no recasts did not develop much at all. The findings are important because they show that recasts are more efficient in the absence of contextual support (compare this to Révész and Han’s finding of recasts being more efficient with video as opposed to notes support). Révész contended that this is because, in accordance with the Cognition Hypothesis, the -photo task was more cognitively demanding, requiring dual task demands of the learner, inducing them to look for external assistance. When learners had the photo present, they had to engage in processing both visual and aural input, which, according to the researcher, probably competed for learners’ attentional capacities. Thus, the availability of the photo in a way was a ‘distractor,’ and did not permit full attention to be paid to the recasts (Révész, p. 462). Those without the photo subsequently had an advantage because they did not need to decode both the photos and the researcher’s recasts simultaneously. Also, learners reported that being able to see the pictures while describing them to the researcher was less difficult than not being able to view them simultaneously. The learners’ perception of the task was therefore similar to the researcher’s: that the -photo group was more complex. In discussing her findings, Révész herself begged the question of why participants in the +photo group paid more attention to the photos than the recasts. She explains that this was probably because the task was not truly dialogic in nature. Though the researcher provided recasts, she did not have an active role in the scene description tasks. If the hearer’s needs had been more emphasized, it might have been the case that learners would have paid greater attention to the recasts. Her 2009 study therefore supports the predictions of the Cognition Hypothesis.

Kim’s 2009 dissertation also explored the effect of task complexity on L2 learning, specifically, the development of English question formation and also the English past tense. This
study was the first to conduct an experiment in the EFL classroom over the course of a semester. Also unique to her design was the use of a continuous, as opposed to dichotomous, scale to operationalize task complexity: simple, +complex, and ++ complex, as Robinson (2007a) did with the simple, medium, and complex continuum. Four intact EFL classes participated in the study, three acting as the experimental groups and one class as the comparison control group. The comparison group only took the pretests, posttests, and delayed posttests, and had traditional class content. The researcher taught all four classes during the semester herself. Kim developed 12 different (4 simple, 4 +complex, 4 ++complex) tasks in relation to the content themes in the course: the world of work, vacationing, and university life in Korea. The task complexity differential explored in her study was +/- reasoning demand and +/- few elements, with the ++complex task being +reasoning demand and -few elements. To provide an example of one of the units (world of work), the simple task (class 1) required learners (in pairs) to complete a report of university students’ part time jobs, with both learners collecting information from their partners. The + complex version (class 2) required learners to suggest, based on each others’ backgrounds, appropriate part time jobs for each student, with two considerations to be met in order to make their decision (+few elements). The ++complex version (class 3) also required learners to suggest part time jobs, but this time they had to make their decision based on four considerations (-few elements). 191 Korean learners of EFL students paired themselves with other students in the classroom and kept the same partners for the majority of the semester. All interaction was recorded throughout the semester. For each unit, one class did all four of the simple tasks, one class did all four of the +complex tasks, and one class did all four of the ++complex tasks in accordance with each theme. Kim controlled for time-on-task (40 minutes) by requiring students in the simple class to come up with more sentences per event in the past
tense and by having more gaps on the task input for question formation. To measure L2 development, learners carried out both individual and paired oral production tests that were also task-based (i.e., predict interview questions/ describe what the person did in this picture (individual) and interview your partner on X topic/describe to your partner what you did (paired)). Learners also did a written metalinguistic test, in which they had to correct erroneous parts of a sentence in English and explain why they were not grammatical. All measures of development were delivered in a pretest-posttest-delayed posttest design.

Kim found a significant effect for task complexity, with the class that always did the ++complex tasks producing more occurrences of LREs for both question formation and the use of the past tense than the +complex and than the simple group (respectively). In terms of developing question formation, 64% of learners advanced to a higher stage of question formation in the simple group, 72% advanced in the +complex group, and 84% in the ++complex group, compared to only 46% in the control group (traditional class). Results were significant for the ++complex compared to the control group. Indeed, the most complex task encouraged learners to produce the most LREs, at the level of significance. In addition, the ++complex group showed more benefits of output during task interaction. The +complex group produced significantly more LREs than the simple group, but the outcome of their LREs did not include as many advanced question forms compared to the ++complex group.

This study provides support for the Interaction Hypothesis, and also for task-based language teaching in EFL classrooms. In terms of Robinson’s Cognition Hypothesis, it showed that the highest reasoning demands results in more opportunities for learning and interaction, thereby confirming the role of task complexity as a promoter of attention to linguistic forms during task-based interaction. This study was also insightful in the use of two different linguistic
items. (Though more LREs were produced for question formation as opposed to the past tense). At the same time, this study does not go without limitations. For example, in reading over some of the LRE transcriptions, it seems that the ++complex group had more practice than the other groups as well as more time. Also, only LREs were assessed. It would have been insightful—and beneficial for comparison to other empirical findings—to look at more local measures of interaction, such as clarification requests, recasts, and/or confirmation checks. Lastly, while the fact that this study took place in the classroom during the length of an entire study does give it high ecological validity, its quasiexperimental design opens up itself up to an inability to control for many variables that could have affect the results, such as outside exposure. Kim’s dedication however to exploring task complexity in classroom contexts is quite commendable.

A study by Kim and Tracy-Ventura (in press) is thus far the first to look at task complexity and how it, alongside learner individual differences, mediates interaction in the classroom. Specifically, they examined anxiety as a moderating factor that has the potential to affect task-based interaction. Recall that Robinson’s Cognition Hypothesis considers variables such as anxiety, motivation and working memory capacity (to name a few) as “Task Difficulty” factors. In accordance with the Cognition Hypothesis, the authors suggested that “when already anxious learners are asked to participate in a task that is complex, their anxiety might hinder whether they are able to benefit from the additional interaction of a more complex task, causing little or no linguistic development.” As with Kim’s dissertation, Kim and Tracy-Ventura operationalized task complexity on a continuum as opposed to a dichotomy: simple, + complex, and ++complex. The task complexity criteria explored was +/- causal reasoning and +/- few elements. 128 Korean learners of EFL were randomly assigned to one of three groups. Each group carried out four interactive tasks, all having to do with the theme of university life in their
country. For example, one of the four tasks was assisting in the preparation of an upcoming political race. The simple group carried out an information-gap task regarding the life of one of the candidates, and then summarized their information exchange. The +complex group had to do the same, but also had to decide who they would vote for, basing their decision on two criteria. The ++complex group did the same as the +complex learners, but had to base their final decision on four criteria instead of two. Learning was measured via individual and paired oral production tests in a pretest-posttest-delayed posttest design. After the treatment, participants filled out an anxiety questionnaire. The questionnaire used in this study had a total of six items and was adapted from Sheen (2008), with Cronbach’s alpha for the instrument at .80. The linguistic item being investigated was the English past tense. All interaction for all task complexity versions was carried out in three intact EFL classrooms in Korea over a period of two weeks.

Results showed that the ++complex group achieved the highest gains on the posttests, followed by the +complex group and then the simple group in that order. Statistical analyses showed that the ++complex and +complex performed significantly better than the simple group, however, differences between the two complex groups were not significant. In looking at the impact of anxiety, Low and high anxiety learners were established in using the 50th percentile of questionnaire scores. 62 learners were deemed low anxiety and 68 were deemed high anxiety. The researchers found that low anxiety learners performed better on posttests compared to high anxiety learners, though differences were not significant. Significance was found at the delayed posttests, however, with low anxiety learners once again performing significantly better than high anxiety learners. This study showed support for the Cognition Hypothesis in that more complex tasks promoted better development of the English past tense, even though there was not a large difference between the two complex groups. Also, it showed that overall, low anxiety
learners performed better than high anxiety learners, but only at the delayed posttests. The researchers did not find an interaction effect between task complexity and anxiety, which was not in line with the Cognition Hypothesis.

One point worth mentioning is that in this study, the researchers did not specify the type of anxiety that their instrument set out to measure, referring to the construct simply as “language anxiety.” The construct of anxiety is a highly attested one in the field of psychology and in SLA (Chamorro-Premuzic & Furnham, 2003; Dörnyei, 2005; Ellis, 2008), with multiple instruments that have been tested in order to properly measure it. Work done on anxiety in the field of psychology demonstrates three separate types of anxiety: trait anxiety, state anxiety, and situation-specific anxiety (Ellis, 2008, p. 691), and there have been recent calls in the literature for researchers to be more specific about what type of anxiety they are operationalizing and measuring (Baralt & Gurzynski-Weiss, in press). A larger instrument, perhaps one of those created by MacIntyre and Gardner (1989; 1991a; 1991b, 1994a; 1994b) and widely attested in the field, may have yielded different results; as well, a triangulation of methods to gain insight into learners’ perception of the task (and their anxiety as caused by the task) might have also been revealing. Clearly, more research is needed on anxiety and how it modulates learners’ interaction with different tasks types in the classroom environment; Kim & Tracy-Ventura are the first to do so thus far. The fact that this study, like Kim’s (2009) dissertation, took place entirely in the classroom also places an important emphasis on task-based interaction and task design within a pedagogical context. Also like Kim (2009), however, is the fact that certain variables were unable to be controlled for (i.e., outside exposure to forms), given the quasiexperimental nature of the study.
Révész, Sachs and Mackey (in press) explored the effect of task complexity on interaction, uptake, and L2 development. They specifically sought to test the claim of the Cognition Hypothesis that engaging in more cognitively complex tasks will result in more uptake of forms made salient in the input (i.e., via feedback) and will cause longer retention of those forms. Cognitive complexity was operationalized as +/- visual support (i.e., +/- here-and-now). They also aimed to explore whether learners’ production of uptake was a predictor of L2 development. 54 native Hungarian learners of EFL were randomly assigned to one of three groups: the +complex group (-visual support), the -complex group (+visual support), and the control group. The two experimental groups took part in three sessions in which they had to tell a robbery scene, differing only in that they were allowed to look at the photos while describing the scene versus not being able to view the photos. All groups (experimental and control) did two versions of pre- and posttests: an oral description task with photo support and an oral description task without this photo support. The targeted linguistic item was the past progressive tense in English. The task was the same as that described in Révész (2009), where learners had to describe a robbery scene in New York City. During the treatment sessions, participants received recasts from the researcher whenever they erroneously produced the past progressive form. As with Révész (2009), recasts were of the isolated, declarative type. After transcribing all of the data, the researchers coded all of participants’ responses to recasts for uptake. A learner response was considered as uptake regardless of whether it was an exact (full) repetition or a partial (half) repetition of what the researcher had said. However, if the learner did not respond or simply acknowledged the recast, it was not coded as uptake. Therefore, acknowledgements such as “yes” or “uhum” were not considered uptake.
The researchers found that learners in the +photo group produced slightly more uptake than the -photo group, however, the difference was not significant. In terms of L2 development, both experimental groups achieved gains after the three treatment sessions, with the -photo group achieving a moderately larger gain than the +photo group. Similar to Révész (2009), the control group did not accomplish hardly any gains. When the researchers ran regression analyses to address whether the production of uptake was a predictor in L2 development, they did not find significance. Therefore, there was no significant relationship between learners’ uptake and gains made on the past progressive form. The interaction of the independent variables revealed a more interesting trend: the more learners produced uptake as a response to recasts on the simple task (+photo), the more likely they were to show development. This did not hold for the complex version (-photo).

Contrary to what the Cognition Hypothesis predicts, task complexity was not a significant predictor in terms of the amount of uptake learners produced after receiving recasts on ill-produced forms in this study. Révész, Sachs, and Mackey go on to say that the fact that an effect for task complexity was found for the interaction of uptake and L2 development (for the +contextual support group) could indicate that learners’ modification of output “may mean something different under different task conditions.” This seems to provide support for Robinson’s claim that task complexity and task design can and does affect the way that focus-on-form techniques (such as recasts) play out in task-based interaction. The researchers suggest that the lack of an effect for task complexity on the development in the -contextual support group is because this group “experienced a heightened sense of problematicity.” They also acknowledge a ceiling effect with scores on the immediate posttests, which may have not permitted a comparable distinction between groups.
An explanatory reason for their results is the monologic nature of the task, a point made by Révész, 2009. While the researchers claim that learners did have the opportunity to ask for feedback, one must consider the nature of this task, which was not interactive in nature: the participants were looking at a PowerPoint screen, were re-telling a scene in the past with or without the presence of photos, and every time they made an error, they were corrected by the researcher. The fact that the sessions were not truly dialogic conversations may have heightened learners’ awareness of the recasts as a form of corrective feedback, because repeating recast in a monologic setting may have rendered it unnatural. More research is needed from both monologic and dialogic tasks to gain insight into the effects of task complexity and uptake when forms are made salient in the input via recasts. It might also have been prudent had the researchers coded for different types of uptake. Given the vast amount of literature on uptake (Ellis, Basturkmen & Loewen, 2001; Havranek, 2002; Loewen, 2005; 2007; McDonough & Mackey, 2006; Nassaji, 2007; Yoshida, 2010), and the fact that it has various forms, it is surprising that the researchers did not code for different types of uptake as is normally done in the literature. For example, as explained above, Robinson (2007a) coded for partial and exact uptake, which was not distinguished in this study. That said, the fact that they allowed both full and partial responses to count as uptake, and that there was no correlation between complexity and uptake, is a strong indicator that the two might not be relevant. Perhaps, as the researchers argue, focusing on production of the form during the task (hypothesis testing and trying) is more related to a delayed retention of the form (this was found by McDonough & Mackey, 2006 for example). Révész, Sachs, and Mackey conclude by calling for more research on task complexity, learning outcomes, and recasts, and also on individual differences that play a role on the efficacy of recasts, such as working memory capacity.
To summarize all of the studies that have explored the effects of task complexity on L2 development, please see Table 7 below:

Table 7. Studies Examining Task Complexity and Learning Outcomes

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>TASK COMPLEXITY FACTOR</th>
<th>DYAD TYPE</th>
<th>TASK</th>
<th>DEPENDENT MEASURES OF LEARNING</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Révész and Han (2006)</td>
<td>Task familiarity (+/- prior knowledge) task type (+/- there-and-then) (Resource-directing)</td>
<td>ESL Learners paired with researcher</td>
<td>Describe the actions of a criminal using video input or written notes input; same video/notes or different video/notes. All received recasts</td>
<td>Fill-in-the-blank task, written description task, video-primed oral task, a notes-primed oral task (prett-postt-delayed-postt)</td>
<td>All groups achieved gains in English past progressive; greater effect size for task familiarity than for task type</td>
</tr>
<tr>
<td>Nuevo (2006)</td>
<td>+/- causal reasoning (Resource-directing)</td>
<td>ESL Learners paired with learners</td>
<td>1) Narrative task: ordering of pictures 2) Decision-making task: seating arrangements of people w/ unique personalities</td>
<td>Oral description and narrative test, GJT (prett-postt-delayed-postt)</td>
<td>No difference in L2 development between groups: English past tense and locative prepositions</td>
</tr>
<tr>
<td>Révész (2009)</td>
<td>+/-contextual support (+/- here-and-now) (Resource-directing)</td>
<td>EFL Learners paired with researcher</td>
<td>Describe a robbery scene in NYC with or without pictures present. Half of learners received recasts</td>
<td>Oral descriptive tasks, written descriptive task, a fill-in-the-gap task, GJT (prett-postt-delayed-postt)</td>
<td>Those with recasts and without photos (+complex) achieved highest gains: English past progressive</td>
</tr>
</tbody>
</table>
(Table 7 continued)

| Kim (2009b) | +/- causal reasoning +/- few elements (Resource-directing) | EFL Learners paired with learners in class (simple, +complex, and ++ complex groups) | 4 different tasks in class over course of semester, all regarding work, vacation, or university life in Korea | Occurrence and resolution of LREs; individual and paired oral tests, metalinguistic written test (prett-postt-delayed-postt) | Sig. more LREs produced in ++complex task for q. form.; more LREs for + and ++complex groups for past tense. ++Complex group had greatest gains in question formation and past tense |
| Kim & Tracy-Ventura (in press) | +/- causal reasoning +/- few elements (Resource-directing) | EFL Learners paired with learners in class (simple, +complex, and ++ complex groups) | 4 different tasks in class over course of semester, all regarding university life in Korea | Individual and paired oral tests | ++complex group and +complex group performed sig. better than simple. No difference between + and ++ groups |
| Révész, Sachs, & Mackey (in press) | +/-contextual support (+/- here-and-now) (Resource-directing) | EFL Learners paired with researcher | Describe a robbery scene in NYC with or without pictures present. All learners received recasts | Two oral description tasks, one with photo support, one without | No effect for task complexity; not a significant predictor in production of uptake. A relationship between uptake and L2 development in -complex task only |
3.3.1 Summary of Cognitive Complexity and the Effect on L2 Development

As with the small group of studies that explored the effects of task complexity on interaction, all of these studies looking at L2 development were in the ESL or EFL contexts. Four were in the laboratory (Nuevo, 2006; Révész, 2009; Révész & Han, 2006; Révész, Sachs, & Mackey, in press) while two were classroom-based (Kim, 2009; Kim & Tracy-Ventura, in press). Unlike Nuevo (2006) and Révész and Han (2006), the remaining four studies showed more learning gains for those that carried out the more cognitively complex version of the task. Also, Révész (2009), Révész and Han (2006), and Révész, Sachs, and Mackey (in press) are the only ones exploring learning not as arising out of LREs but rather via recasts delivered by an expert interlocutor during the task interaction. How increases in the cognitive complexity of tasks affect learning, the production of uptake when forms are made salient during interaction, and how individual differences mediate all of this clearly warrants further investigation. Perhaps the factors of expert interlocutor (versus not), as well as true dialogic interaction (versus not) may be the cause of the conflicting findings so far. Also, it is so far inconclusive as to whether or not increases in the cognitive complexity of tasks lead to more uptake versus production of the form during treatment. It may be the case that production of the form during the task, and not production of uptake, is more predictive of delayed retention of the form as Révész, Sachs, and Mackey found. More research in this area is necessary and essential to understand the effects of task complexity on learning, and how this differs for different types of learners.

3.4 Measuring L2 Development with +/- Complex Tasks: Nuevo’s Call for Customized Assessments
Another issue of contention is how to best go about measuring learning with tasks of more or less cognitive complexity. Many of the criteria in Robinson’s Triadic Componential Framework are ones for which learners must come up with their own linguistic means to perform the demands of the task, such as communicating causal reasons, intentional reasons, events in the there-and-then frame, etc. Nuevo (2006) has raised the important point of using customized tests to best grasp what the participant came up with him/herself during the task. In explicating her study’s findings that more complex tasks did not lead to more L2 development, she pointed out that her assessments might not have properly captured what the learners did during the task. Nuevo called for future research to use custom-made posttests, especially if the end goal is learning from task-based interaction and an acquisition of items and/or structures that arise from that interaction.

3.4.1 Benefits of Customized Assessments and the Cognition Hypothesis

A large benefit of using customized or tailor-made assessments is that, as Nuevo explained, they capture what the learner came up with during the interaction. If the researcher is able to make customized assessments quickly and in a way so that there is not a time delay between the treatment and the assessment, customized tests can match what the learner came up with, thereby measuring development in a more meaningful way to the learner. Customized assessments address in a scientific way what the learner did—they are tailored to the learner. If claims want to be made in regards to the Cognition Hypothesis and learning, customized assessments may need to be used in order to best measure learning.

3.4.2 Cons

At the same time, there are disadvantages to using customized assessments. Besides being time consuming to create, customized assessments can reduce the generalizability of
results. Any results that are deduced from customized assessments will be germane only to that specific study, its participants, and the linguistic items that arose during their interaction. Another major disadvantage of customized assessments is their comparability; it is difficult to ensure that one is testing all participants in the same way.

3.4.3 A Mixed-Method Approach

Using a combination of standardized and customized assessments, or even the creation of one instrument that incorporates both elements, is a way to avoid these potential pitfalls. Say a researcher is operationalizing +/- causal reasoning, as Nuevo (2006) did. If a task is designed in such a way so that participants in the +complex group come up with many of the same reasons (deduced by extensive piloting) as those on the standard assessments, the researcher could establish an 80% overlap of items for a task and its subsequent assessment. The remaining 20% of items could then be tailored to each learner. In this way, statistical comparisons can be carried out, exploring the comparability of learning among participants as well as items particular to individual participants. This would be a more robust way of measuring L2 development with tasks of more or less cognitive complexity within the Cognition Hypothesis framework. More studies are needed that employ both standardized and customized assessments to measure learning that arises out of task-based interaction situated with tasks of different levels of complexity.

3.5 Measuring Cognitive Complexity: Norris and Ortega’s Call for Independent Measures

Another critical area when doing research on the construct of cognitive complexity is to robustly measure that construct, ensuring that it has been properly operationalized. What the researcher believes is an increase in cognitive complexity may not always be what learners
believe; Révész’s (in press, b) findings of mismatches in student and teacher perceptions of task difficulty is an excellent example of this. This is such an important issue that Norris and Ortega (2009a) published an article dedicated to the problems with “the case of complexity” and ways to measure it as a construct (see also Norris & Ortega, 2003). The researchers rightfully pointed out that evidentiary logic for measuring cognitive complexity has often been faulty in the literature, especially with those studies that have looked at the effects of increases in cognitive complexity on language production. In other words, claims have been made for cognitive complexity as having a positive effect on more complex language, which lead to positive results on cognitive complexity because of the complex language that was produced. Norris and Ortega highlight the problem of this circular logic, and call for evidence as to the effect of cognitive complexity that is independent and not linguistically based. The researchers point out ways in which this has been done in the field of psychology, such as (1) self-reported data in regards to how participants feel about the difficulty of the task; (2) subjective time estimations after completing a task; (3) reaction time data for dual tasks; (4) physiological data, to include heart rate variation, skin conductivity, papillary response, and even brain activation and blood flow patterns. While the last two options might require expensive equipment in order to collect data, the first two options, self-reported data and subjective time estimations are quite easy for researchers to collect.

3.5.1 Self-Reported Data: Quantitative and Qualitative

In terms of self-reported data, the researcher assumes that the participant is able to introspect on his/her cognitive processes and effort expenditure. This can be done quantitatively with questionnaires that have Likert-scale items, or qualitatively by noting comments said by participants during a task in which they are engaging in reasoning. According to Paas, Tuovinen, Tabbers, & Van Gerven (2003), “it has been demonstrated that people are quite capable of giving
a numerical indication of their perceived mental burden” (p. 66). Questionnaires can be very revealing and, as indicated before, are thus far the only way that cognitive complexity has been measured in the field. Qualitative comments said by participants during the task can be very revealing about how they are processing the task, feedback, and what is going through their minds. Such comments can be tabulated for further insight as to the challenges that the learner reports. Qualitative, online data like this can assist in showing cognitive overload, crystallizing what goes on during the task. At the same time, not all participants will think aloud during a task, make comments or engage in private speech\(^{13}\) to themselves, so qualitative comments will only provide part of the picture.

### 3.5.2 Time Judgment of Tasks

In terms of subjective time estimations, quite a bit of literature in the field of psychology has addressed this issue. Hicks, Miller and Kinsbourne (1976) explained that “judged time is usually an increasing function of the number of stimuli that occur during an interval.” Methodologically speaking, it seems to make a different if time estimation is done prospectively or retrospectively (in other words, if subjects are explicitly told or not that they will be required to estimate the time that it took them to take the test, see Fink & Neubauer, 2000). Several studies have reported that information processing leads to increased time judgments, and that time estimations decrease as more intervals pass (e.g., Hicks, 1976; Michon, 1965; Thomas & Weaver, 1975). Some studies have also found that judged time increases with the amount of information processed (Ornstein, 1969; Underwood & Swain, 1973;) and with how much is remembered about the events that took place during the task (Block, 1974; Ornstein, 1969). At the same time, many studies have shown a decrease of subjects’ subjective time judgments

---

\(^{13}\) Quoting Ohta (2001), Ellis (2003) provides a definition of private speech: “audible speech not adapted to an addressee” (p. 178).
alongside an increase in nontemporal information processing demands (Burnside, 1971; Casini & Macar, 1997; Hicks, Miller, & Kinsbourne, 1976; Macar, Grondin, & Casini, 1994). In other words, it may be the case that time is underestimated in relation to the “experienced time interval” (Dzaack, Trösterer, Pape, & Urbas, 2007). These findings subscribe to the attentional gate model (e.g., Block & Zakay, 1996), where it is posited that attentional demands take away from a person’s sense of time. At the same time, the selection of tasks in these studies may be a reason for the disproportionate findings: “One critical issue is that there has been little if any theoretical coherence in the choice of tasks used to manipulate cognitive load [in psychology]” (Block, Hancock, & Zakay, 2010, p. 331). Block et al. carried out a meta-analysis and reviewed some key features of study designs that can differentially affect how subjects judge time: (1) types of cognitive load, (2) attentional demands, (3) response demands, (4) +/- familiarity, (5) memory demands (such as knowing that they will be tested on the content afterwards), (6) processing changes, and (7) processing difficulty. In reviewing 117 different experiments, Block et al. found that human time judgments are affected by cognitive load but in opposite directions depending on the prospective (i.e., knowing in advance that they will have to judge the duration of the task) versus retrospective (not knowing until after the task has finished) paradigm. They found that in retrospective designs of tasks with high cognitive load, participants typically judge the task as taking longer than it really did; this may depend on task type however.

At the same time, no study in psychology has employed tasks that involve second or foreign language processing where semantic content and semantic processing is inherent to the task. Instead, tasks in psychology have involved, to provide some examples, anagram manipulation (i.e., Buchwald & Blatt, 1974); card sorting (i.e., Hicks, Miller, & Kinsbourne, 1976); word classification tasks (i.e., Hanley & Morris, 1984); and ignoring versus attending to a
second word list (Brown, 1985). It is unknown whether time judgments for L2 language tasks of increases in cognitive complexity within the Cognition Hypothesis Framework will yield the same results as these types of task. Clearly, more studies are needed that employ these kinds of independent measures of the construct, as are other methods employed in psychology. Asking participants to estimate the amount of time they believe took them to carry out a task of more or less cognitive complexity may be very revealing in tapping the construct, especially with the Cognition Hypothesis framework.

3.6 Summary of the Chapter

Up until now, the research on (1) how task complexity affects learning outcomes and (2) how the efficacy of recasts is mediated by the task complexity differential remains inconclusive. At the same time, the six studies reviewed in the second half of this chapter provide support for the Interaction Hypothesis in that SLA can be achieved through conversational interaction. They also demonstrate that the design of tasks, and specifically, the cognitive complexity of tasks, mediate these outcomes. Pairing learners with other learners, Nuevo (2006) found no effect for increases in task complexity on the development of locative prepositions and the past tense. Contrarily, Kim (2009) and Kim and Tracy-Ventura (in press) did find an effect on development of English question formation and the past tense. Révész and Han (2006) paired learners with a researcher and provided recasts during the interaction. They reported that all groups achieved gains; however, the Video group with familiar content (the -complex group) did best. Conflictingly, Révész (2009) also provided recasts to learners during a +/- complex task, and reported that those in the +complex group achieved the highest gains. Most recently, Révész, Sachs and Mackey (in press), while finding that more learning did occur in the more complex
task version, did not find an effect of task complexity on the production of uptake, and reported that uptake production was not a significant predictor of L2 development. All of these studies were in the EFL context except for Nuevo, which was ESL. All operationalized either +/- reasoning demands or +/- few elements or, in the case of Révész’ studies, +/- contextual support. Of all of these studies, only one looked at Task Difficulty variables as having the potentiality to affect interaction and learning outcomes in unique ways for tasks with differing levels of task complexity; Kim and Tracy-Ventura (in press) reported that low anxiety learners performed better than high anxiety learners for the most part, but that differences in learning were not significant. At the same time, the instrument used by the researchers to measure the construct of anxiety has scale limitations. No study looking at the effects of cognitive complexity on learning has ever measured development with customized assessments to best capture what the learners did during the interaction. Also, no study has measured the construct of task complexity with anything other than an anxiety and perceived difficulty questionnaire. The present study will attempt to address all of these issues in its design by (1) using a combination of standardized and tailor-made post assessments to measure L2 development, and (2) using independent measures of cognitive complexity via a) the tabulation of qualitative comments made by participants during the task in regards to its difficulty, b) time judgments of the duration of the task and c) an Anxiety and Perceived Difficulty Questionnaire for comparability purposes with other studies.

The next chapter will discuss the construct of working memory capacity and review literature on this construct from both the fields of psychology and SLA, explaining its relevance to SLA, task design, and feedback during dialogic interactive tasks.
CHAPTER 4: THE CONSTRUCT OF WORKING MEMORY CAPACITY

According to Conway, Kayne, Bunting, Hambrick, Wilhelm, and Engle (2005), working memory capacity (WMC) is a “multicomponent system responsible for active maintenance of information in the face of ongoing processing and/or distraction” (p. 770). Robinson makes specific claims on cognitive abilities such as WMC, theorizing that they will increasingly affect task-based performance as tasks increase in cognitive complexity. Only a handful of studies have explored WMC in the field of SLA. Before reviewing those studies, its original investigation in the field of psychology, upon which SLA studies are mainly premised, will be presented.

4.1 On the History of Memory

Historically, the construct of “memory” has been approached by two main camps: semantic and episodic memory (Tulving, 1972), and implicit and explicit memory (one of the most widely accepted dimensions). Past views on the construct of memory have ascribed a dichotomous short-term memory (STM) and long-term memory (LTM), or maintenance and long-term storage, with the latter account implicating a transfer to LTM as a result of rehearsal in STM (Atkinson & Shiffron, 1971). According to Atkinson and Shiffron, some in this camp see long-term storage as a function of how many times an item is rehearsed. Free recall tests were often used to explore STM. A common free recall test consists of presenting the participant with a list of random items, one at a time. Later, the participant is asked to recall as many of those items as s/he can, in any possible order. Free recall is affected by amount of items: words from shorter lists are recalled better than words in longer lists and words presented slowly are also better recalled (Atkinson & Shiffrin, p. 84).
Others have ascribed a dichotomous notion of STM and LTM, what Baddeley distinguished as “based on temporary electrical activation” (STM) and “based on neuronal growth” (LTM), (Baddeley, 2003, p. 830). Earlier models from the 1960s assumed that the short-term store also served as a human working memory system, critical for learning and retrieval. However, this idea did not seem to be supported by neurophsychological evidence (with patients who were unable to retain anything in the short term showing normal long-term storage). Further evidence in patients with neuropsychological damage suggested a separation of STM and LTM—it was shown that patients with damage to their STM capacity were still capable of long-term learning (Baddeley, 2003. p. 830). This led Baddley to develop his tripartite model of working memory (WM). (See Engle, Laughlin, Tuholski & Conway, 1999, for a detailed explanation on the differences between STM and working memory). WM is asserted to be a “storage component as well as an attention component,” and “the contents of STM plus the limited-capacity controlled-attention processes associated with the central executive that can be used to maintain some set of those STM units as the focus of attention” (Engle et al., 1999, p. 310).

4.2 Two Theories of Working Memory Capacity

There are two theories of working memory that are most operationalized in the field: Baddeley’s multi-component model (Baddeley & Hitch, 1974; Baddeley, 1986, 2000) and Just and Carpenter’s (1992) constrained capacity theory. These two theoretical approaches are presented below.

4.2.1 Baddeley’s Multi-Component Model

About thirty years ago, Baddeley and Hitch (1974) proposed a model of working memory that is comprised of three parts: the central executive, which is “the control system of limited
attentional capacity,” and two subordinate storage systems that assist the central executive: the phonological loop (for sound and language) and the visuospatial sketchpad (for imagery and dimension). Important to this model (the one most identified with) is the “emphasis [it places] on combined processing and storage, and the stress on its functional importance as a system that facilitates a range of cognitive activities, such as reasoning, learning and comprehension” (Baddeley, 2003, p. 829). Recently, Baddeley has added a fourth component to his model, the episodic buffer (Baddeley, 2000). Each of these components are explained below.

4.2.1.1 Phonological loop

The phonological loop has two subcomponents: a phonological store (where memory traces are held) and an articulatory rehearsal process. Studies that have employed tasks with increasing numbers of digits or word lengths provided evidence for the second component of articulatory rehearsal; for example it has been shown that immediate memory span decreases as word length increases (typically from one to five syllables, according to Baddeley, Thomson, & Buchanan, 1975). Neuropsychological evidence is also revealing: anarthric patients (patients unable to speak) have no problem processing visually presented words, while dyspraxic patients (who are unable to make speech-motor connections) are unable to do so. This has led Baddeley to conclude that “subvocal rehearsal does not seem to depend on the capacity for peripheral control of articulation” (p. 831). According to Baddeley, phonological loop capacity is a good predictor of an adult’s ability to acquire a second language. Specifically, he explains that language development is assisted by the phonological loop in two ways:

…the store should provide relatively unconstrained temporary representation for new phoneme sequences, and the articulatory system should facilitate learning through
rehearsal, provided that the new sounds can be represented using existing output processes (2003, p. 833).

Inhibitory factors that interfere with the phonological loop include articulatory suppression (to include anarthric patients), phonological similarity (examples would be b and p, f and v; but not m vs. w) and word length (with five syllabus being more difficult to retain than one). See Baddeley’s model of the Phonological Loop below in Figure 1:

**Figure 1. A Functional Model of the Phonological Loop** (from Baddeley, 2003, p. 831)

With this model, Baddeley demonstrates the components and processes that make up the phonological loop. Note that both auditory and visual stimuli can count as input, upon which they undergo either phonological or visual analysis. Auditory and/or visual input can then be processed in short-term storage (STS). Auditory input is directly accessed in the phonological store, after which it passes through the phonological output buffer. Here such stimuli can be either recalled or rehearsed. Visual input is recoded verbally and also gains access to the phonological store via a process of orthographic to phonological recoding, or rehearsal. Note that Baddeley also accounts for neurophysiological placement in his model—after rehearsal, both auditory and visual input can eventually be programmed for output. This takes place in the premotor cortex area of the brain, or Broca’s area. Baddeley proposes that this phonological loop evolved so that language acquisition could take place (p. 832). Evidence to support this theory comes from patients with “pure” phonological loop deficiencies who are unable to learn vocabulary of a new language, even though they have functioning verbal LTM. “Phonological loop capacity is a good predictor of the ability of children and adults to learn a second language” (2003, p. 832), and it does so in two ways: (1) the store allows for “unrestrained temporary representation for new phoneme sequences” and (2) the articulatory system allows for learning through rehearsal (p. 833).

### 4.2.1.2 Visuospatial sketchpad

The other subsidiary storage system central to WMC is the visuospatial sketchpad, which also has limited capacity. As with the phonological loop, retention of items (which in this storage system can consist of color, shapes, location) requires attention. Many researchers have argued for a division of the visual and spatial components, just as the phonological store and the articulatory rehearsal process are two subcomponents of the phonological loop. Spatial memory
is typically measured in studies of psychology using the Corsi block task (this task shows participants a 3D picture of blocks; the researcher taps blocks and then the participant repeats what the researcher did until they no longer can remember the sequence). Visual memory is often measured by using pattern spans. For these tasks, participants are shown a pattern (i.e., blocks filled on a grid), and then are asked to fill in the same blocks when the pattern is taken away. Though less research has been done on this component of working memory, it is possible that participants tap both components during certain tasks. Baddeley suggests, for example, that “a number of studies … have found that subjects will use coding in the phonological loop to store items while manipulating them visuospatially” (2003, p. 834). In terms of learning, he suggests that the visuospatial sketchpad may play a role in the acquisition of semantic information, such as how objects look and how to use them. There is an overall paucity of research thus far on this important component of WMC.

4.2.1.3 Central executive

Previously considered to be a “pool of general processing capacity,” the central executive is considered the “controller” of working memory, the regulator that allows humans to direct their attention to certain features of any input (or not), to change that allocation of attention, and to connect working memory to LTM (Baddeley, 2003, p. 835). The central executive is attentionally limited. Key to this component is its supervisory activating system, or SAS. Evidence for such a controller system derives from patients with damage to the frontal lobe area of the brain. Such patients typically have impaired SAS function and are characterized as having “excessive distractability” (p. 835). In other words, the central executive is key for humans to be able to engage in self-control and divide and switch their attention. Baddeley has postulated that
working memory capacity is crucial for conscious awareness. So far, this component of WMC is the least understood, even though, according to Baddeley, it might be the most important.

### 4.2.1.4 Episodic buffer

The last component, added in 2000, is the episodic buffer. Attentional controlled by the executive, the episodic buffer was proposed to allow the two components, the phonological loop and the visuospatial sketchpad, to interact, as well as to inform the ways in which working memory serves conscious awareness. Baddeley has suggested that it could be the storage section of the executive control. He further contends that the episodic buffer “binds together information to form integrated episodes” … and that the “multi-dimensional coding allows different systems to be integrated … conscious awareness provides [this] convenient binding and retrieval process” (2003, p. 836).

Evidence for the anatomical locations of the three subsections of working memory give further evidence to Baddeley’s model. The phonological loop is shown to be located in the left temporoparietal lobe (Valler & Papagno, 2002), with spatial coding associated with the inferior parietal cortex. The central executive has been shown to be highly associated with the frontal lobes of the brain. Processes of rehearsal are associated with Broca’s area. The model that Baddeley provides for his multi-component proposal of WMC is provided below:
An examination of Baddeley’s model shows that all components through which input is received are controlled by the central executive. Visual semantics, episodic LTM, and language comprise what Baddeley explains is long-term, or “crystallized” knowledge. Between the visuospatial sketchpad and the phonological loop is the episodic buffer, which serves as an interface between the two STM and LTM sub-systems. This is where the SAS can intervene with designating attention and having control over this designation.

4.2.2 Just and Carpenter’s Constrained Capacity Theory

Just and Carpenter (1992) instantiated a model of WMC more in accordance with specifically language comprehension, explaining that the differences in L2 language comprehension performance observed in individuals is a function of their WMC. “[L]anguage comprehension is an excellent example of a task that demands extensive storage of partial and final products in the service of complex information processing” (Just & Carpenter, 1992, p.
Central to their model is the concept of activation manipulation, which they contend fits accordingly with connectionist models of comprehension (e.g., Cottrell, 1989; St. John & McClelland, 1990; Waltz & Polack, 1985; cited in Just & Carpenter, 1992, p. 123). Just and Carpenter argue that the processes involved with being able to comprehend language (i.e., operating, retrieving, comparing), in addition to storage, are what make up a person’s WMC. Unlike Baddeley, Just and Carpenter do not include separate modality-specific components as part of their model (i.e., the phonological loop vs. the visuospatial sketchpad). Rather, they contend that WMC specifically has to do with how Baddeley’s central executive would deal with language comprehension. For Just and Carpenter, a person’s WMC is the maximum amount of activation permitted for simultaneous storage and processing. When a learner is processing information, all elements of input (e.g., words, grammatical structure, propositions, etc.) have a level of activation. During comprehension, these elements are encoded from either written or oral input. The encoding process for comprehension to take place is generated by a computation or is retrieved from LTM. Importantly, “as long as an element’s activation level is above some minimum threshold value, that element is considered part of working memory, and consequently, it is available to be operated on by various processes” (1992, p. 123). The notion of a threshold is also key to Just and Carpenter’s model; if the input being comprehended requires more activation than is available to a person’s WM system, some of the elements in the input will be ‘deallocated,’ or forgotten. At the same time, the WMC system allows the learner to balance accordingly; depending on the activation level necessary to process certain elements, other elements can be added or deleted via a change to the activation previously assigned to them.
4.2.2.1 Simultaneous activation and exceeding capacity

Just and Carpenter (1992) further explain that the processes involved in comprehension, i.e., syntactic, semantic, pragmatic calculations of oral and/or written input, can happen simultaneously and can also “generate partial products at the same time” (p. 123). If however the total amount of processes being activated exceeds a person’s capacity, the ‘propagation attempts’ are reduced so that activation is maintained within its ‘maximum bound’. To conclude, the authors highlight:

In sum, the time course and content of language processing within this system depends on the capacity for storage and computation. When the task demands exceed the available resources, both storage and computational functions are degraded. We call this theory *capacity constrained comprehension* (p. 124).

Just and Carpenter argue that the Reading Span task (created by Daneman & Carpenter, 1980; explained in further detail below) is best in order to be able to assess an individual’s WMC. The Reading Span task requires processing and storage simultaneously; participants are shown a set of unrelated sentences, and then must recall the final words from each sentence per set. A person’s reading span is consequently the maximum number of sentences per set for which he/she can remember the final words. Reading spans usually vary from 2 to 5 for university-aged students (p. 125); a high span would constitute spans of four words or more; a medium span would be three to three and a half, and a low span is one of less than three words. Reading Span task measures have been shown to correlate with verbal SAT scores and comprehension skills, showing that “language comprehension and simultaneous digit recall can draw on a shared resource” (p. 125). Just and Carpenter also report on research showing that listening versions of the Reading Span yield similar results (p. 125).
4.3 Similarities and Differences Between the Two Models

Currently, Baddeley’s multi-component model (Baddeley & Hitch, 1974; Baddeley, 1986, 2000) and Just and Carpenter’s (1992) constrained capacity theory are the two models most referred to when discussing working memory. Both models agree on a central ‘controller’ of attentive resources: the Central Executive (Baddeley’s model) versus the Executive Function (Just & Carpenter, 1992). However, they differ in how they view the putative modularity of working memory resources. Baddeley sees WM as having multiple components while Just and Carpenter see it as a monolithic entity. The majority of research investigating Baddeley’s model has focused on the functions and interactors of the phonological loop, while most research into Just and Carpenter’s theory has operationalized the executive function.

4.4 Research on the Construct of WMC

Both models have resulted in a decade of intense research into the construct of WMC (e.g., Ackerman, Baddeley, 2007; Beier, & Boyle, 2005; Conway, Cowan, & Bunting, 2001; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Jarrold, Kane, Miyake, & Towse, 2007; Conway, Kane, & Engle, 2003; Daneman & Merikle, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003; Kane, Hambrick, & Conway, 2005; Jarrold & Towse, 2006; Miyake, 2001; Miyake & Shah, 1999; Oberauer, Schulze, Wilhelm, & Süß, 2005; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002), showing that it plays an important role in complex cognitive capabilities such as comprehension and problem solving (Engle, 2002); general intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Engle, Tuholski, Laughlin & Conway, 1999); simple attention control, reasoning, and importantly, language comprehension.
(Ackerman, Beier, & Boyle, 2005; Conway, Cowan, & Bunting, 2001; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Daneman & Merikle, 1996; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003; Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Finally, due to neuroimaging and neuropsychological studies (Kane & Engle, 2002), it is now known that these abilities related to WM function are located in the prefrontal cortex of the brain. The following definitions are provided in order to highlight how the construct has been operationalized in the fields of both psychology and SLA. These definitions are insightful when considering how the construct is relevant for learning and noticing feedback during interactional tasks:

4.5 Different Definitions of WMC

1) WMC is a “multicomponent system responsible for active maintenance of information in the face of ongoing processing and/or distraction” (Conway, Kayne, Bunting, Hambrick Wilheim, & Engle, 2005, p. 770).

2) WMC is a “storage component as well as an attention component,” and “the contents of STM plus the limited-capacity controlled-attention processes associated with the central executive that can be used to maintain some set of those STM units as the focus of attention” (Engle et al., 1999, p. 310).

3) WMC “reflects the ability to maintain the activation of knowledge units in the focus of attention” (Engle et al. 1999, p. 312).
4) WM is “a mental construct that accounts for how the key processes of perception, attention and rehearsal take place. It is believed to play a central role in L2 acquisition” (Ellis, 2008, p. 983).

5) WMC is, “[i]n addition to its role in storage, … the pool of operational resources that perform the symbolic computations and thereby generate the intermediate and final products” (Daneman & Carpenter, 1992, p. 122).

6) Verbal WMC is “the system that stores information temporarily and allocates attention to the processing of this information” (Martin, 2005, p. 204).

In sum, irrespective of whether one ascribes to the Baddeley multicomponent model or the Just and Carpenter unitary model, the construct of WMC deals with simultaneous processing and storage, and is important when it comes to language learning (e.g., Harrington & Sawyer, 1992; Mackey et al., 2002; Miyake & Friedman, 1998; Robinson, 2002, 2005b; Sagarra, 2007; Tokowicz et al., 2004). Critically, WMC may be related to noticing feedback and modified output production (Mackey et al., 2002, 2008). This is precisely where the present dissertation aims to contribute in regards to WMC and noticing of feedback during task-based interaction. The next section of this chapter will review methods employed for measuring WMC and will then review the literature on WMC that is specific to the field of SLA.

4.6 Measuring Working Memory Capacity

While the notion of WM as a construct has had a great impact in multiple disciplines, one serious methodological issue is the type of task used to measure it and the procedure employed in order to do so. According to Conway et al. (2005), three specific verbal tasks stand out as most reliable in measuring the construct of WMC: the operation span task (OSPA N) (Turner & Engle,
1989; see Engle et al., 1992, for automated version); the counting span task (CSPAN) (Case, Kurland, & Goldberg, 1982); and the reading span task (RSPAN) (Daneman & Carpenter, 1980). In their methodological review, they review the history of these three tasks and provide the reliability and validity of them as empirical measures. For example, these three tasks in particular load on the same factor in factor analyses “as WM span tasks in which that tasks demand spatial processing and storage” (Conway et al., 2005, p. 771). Conway et al. provide detailed discussions on WM task reliability and issues with administration and scoring procedures that they recommend researchers to follow, avoiding inconsistent and problematic procedures that make findings nongeneralizable. In reviewing a history of the development of the three tasks, the researchers mandate that the following critical methodological points be employed when measuring the construct of WMC:

4.6.1 Recommendations of Conway et al. (2005)

(1) that each test is administered individually (as opposed to group presentation) to each learner. This ensures “immediate and vigilant stimulus presentation” (Conway et al., p. 773), considered critical so as to avoid rehearsal;

(2) that sufficient item size must be met, with at least a ‘range of 2-5 elements per item;’

(3) that participants be required to recall digits (i.e., isolated letters) instead of words, given that (a) “individual differences in reading ability leads to differences in the ability to generate the words at test on the basis of the gist of the sentence (rather than on the basis of episodic recall)” (p. 772); and (b) previous research indicates a “potential shared variance

---

14 The review of these three tasks is done for the purpose of providing research tools for researchers in a “theory neutral” way (p. 770).
15 All of the span tasks are available to other researchers by contacting Dr. Engle and asking for permission, for the purposes of replication and advancement in the field. He and his team have graciously offered access to programs to all researchers at http://psychology.gatech.edu/rengelelab/tasks.htm
between span tasks with words and a measure of higher order cognition, to include reading comprehension, and that this is due to word knowledge” (Engle, Nations, & Cantor, 1990; Engle et al, 2005, p. 501);

(4) that all set orders be randomized so that participants do not develop strategies that derive from their knowing the size of the memory set. The underlying structure is the same for all three: the OSPAN, CSPAN, and RSPAN, in that each span task must “[force] WM storage in the face of processing” (p. 773);

(5) that the order of the WM span tasks is delivered to all learners in a counterbalanced order;

(6) that populations of young, healthy adults be used, as this is the population that comprises the majority of other comparable populations in the literature—this renders findings comparable.

To follow is a brief description of each span.

4.6.1.1 OSPAN (Turner & Engle, 1989)

For the OSPAN task, participants are given arithmetic problems via the computer, and are asked to read and solve the problems aloud. After solving the problem aloud, participants are shown a second screen, on which a possible answer is given, as well as a prompt for them to say “true” or “false” if that answer is true or not. Immediately following this, participants are shown a letter, which they also say aloud. Upon hearing the letter, the researcher immediately presents the participant with next mathematical string, preventing the participant from engaging in rehearsal. All items are randomized throughout the task. After each set, participants are prompted to write down all of the digits that followed each mathematical string in the order that he/she saw. This task was created and has been adapted by Turner and Engle (1989) and Engle et al.
(1992); a more recent, computerized and automated version is described in Unsworth, Heitz, Schrock, and Engle (2005).

4.6.1.2 RSPAN (Daneman & Carpenter, 1980)

For RSPAN task, participants are shown sentences on the computer, half of which are semantically anomalous and half of which are not (original versions included syntactically erroneous sentences as well). Participants are prompted to read the sentence out loud, and then must judge the sentence aloud by saying “yes” or “no” when seeing the prompt screen. After making the oral judgment, participants are immediately shown the next slide, with a letter on it, which they also say aloud. Participants are immediately shown another sentence, and the sequence continues. After each set of sentences (ranging from 2-5, with three trials of each size, for a total of 42 sentences), the participant is given a screen with three question marks (“???”) which is their cue to write down the final words of each sentence they had just seen. The original version of this task was created by Daneman and Carpenter (1980); see Turner and Engle (1989), Engle, Tuholski, et al. (1999) and finally, Unsworth, Redick, Heitz, Broadway, and Engle (2009) for the latest automated version.

4.6.1.3 CSPAN (Engle, Tuholski et al., 1999)

The original CSPAN task was created by Case et al. (1982) for child populations; it has subsequently been adapted for adult populations. Like the OSPAN and the RSPAN, the CSPAN task requires that learners remember a final count amount for a subsequent recall test. In the most recent version created by Engle, Tuholski et al. (1999), learners have to count aloud dark blue circles against a backdrop where visually/colorfully similar items are placed, specifically, light blue circles and dark blue squares. For this task, after counting aloud the dark blue circles, participants must then immediately repeat the final count number. The researcher immediately
shows the participant the next slide, with another series of circles and squares, for which the participant must start counting again. No pausing is allowed, so as to prevent rehearsal of any kind. Upon seeing the cue screen, participants write down the letter sequences they saw in that set in the same order that they saw them.

Importantly, all three spans (OSPA, RSPAN, and CSPAN) must be administered on an individual basis to prevent rehearsal, are done aloud on behalf of the participant, and engage concurrent processing and storage.

4.6.2 Grading WMC Tests

Conway et al. (2005) also provide a detailed section on how to grade the three tests. In the past, most WM spans were scored in a quasi-absolute way (similar to the original Daneman and Carpenter, 1980, study; also employed in Just and Carpenter, 1992). Using that methodology, a participant’s WMC was essentially the maximum threshold that they reached during the test (i.e., the maximum number of sentences for which they had to recall words). Once threshold is reached, the testing then stops. For example, a participant’s WMC is 3 if a set size of three sentences is the maximum for which he could recall final words. Conway et al. (2005) explain that this absolutive method is problematic in that it threatens span reliability across different tests. Furthermore, “by simply estimating the item size at which a subject falls below a given threshold (and then ending the task), information on all other trials is discarded” (p. 774). With scores of between two and six, the sensitivity and value of the measure is limited. To compensate for these limitations, Conway et al. recommend that a point be assigned to all correctly recalled items within a set, taking into account that, for example, “an item with six elements [has] twice as many responses as for an item with two elements” (p. 774). This data is then aggregated, upon which the researcher can employ four different procedures for scoring:
partial-credit unit scoring, all-or-nothing unit scoring, partial-credit unit scoring, or all-or-nothing load scoring. The essential difference between these four procedures has to do with (1) full versus partial credit for correctly recalled items (i.e., recalled in their correct place or not); and (2) whether more weight is assigned to items with a higher memory load or not. In reviewing Kane et al. (2004), (for which data from 236 subjects were subjected to the four different scoring procedures) the researchers recommend partial-credit scoring be done, and in particular, partial-credit unit scoring.

4.6.2.1 Partial-credit unit scoring

Partial-credit unit scoring involves calculating the mean proportion of elements that were recalled correctly in an item. In other words, all items are counted as equal regardless of the item size (recalling two elements out of a four-element item is the same as recalling three elements out of a six-element item). These proportions are then averaged to calculate the score. (This is in contrast to calculating the average of all correctly recalled elements and giving a higher weight to items containing a higher load).

In sum, this review of the literature demonstrates that the OPSAN, RSPAN, and CSPAN are the most attested and reliable when measuring WMC. In following the strict methodological procedures as described by Conway et al. (2005), and in using the partial-credit unit scoring method for grading, researchers can examine the construct in the most robust way. One serious problem with how WMC has been studied in the field of SLA is the fact that researchers have all used different tests, limiting generalizability and putting into question instrument validity (see review below). Looking to the field of psychology and using the OSPAN, RSPAN, and CSPAN will bridge the fields of psychology and SLA and allow for replication studies. The present dissertation will use these three spans and methodology as suggested by Conway et al. (2005).
The remainder of this chapter will review WMC as it has been studied in SLA. To conclude is a comparison of the instruments and methods for measuring WMC in the field, which has thus far been incomparable in scope.

4.7 Exploring Working Memory in the Field of SLA

In the field of SLA, the construct of working memory is just beginning to be explored, mostly in the FTF (oral mode) context, with three studies to date exploring WMC in an electronic or computerized environment, namely, in the research strands of computer-mediated communication (CMC) or computer-assisted language learning (CALL). To start is a review of studies looking at WMC in the FTF mode.

4.7.1 Working Memory Capacity in the FTF Mode

A study by Mackey, Philp, Egi, Fujii and Tatsumi (2002) explored WMC and how it impacts to what and how learners pay attention during second language interaction. 30 Japanese L1 learners of ESL participated in three dyadic, task-based sessions during which they received recasts on their erroneous production of question formation in English. (Sessions included a picture drawing, picture difference, and story completion tasks.) The authors specifically measured phonological STM and verbal WMC. The tests they used were a nonword repetition test to measure phonological STM and two listening span tests, one in English and one in Japanese, for verbal WM capacity. For the nonword repetition test, learners had to repeat as best as they could 16 pairs of non-words after hearing a tone prompt. For the two listening span tests, participants had to answer “ok” or “nonsense” upon judging the sentences, and then had to recall all sentence initial words in the Japanese span test, and sentence final words in the English span test. From these three tests, the researchers created one composite z-score for each participant’s
WMC. Learners also participated in a stimulated recall and an exist questionnaires to measure noticing, as well as pre- and posttests to see how much they had improved in question formation from the sessions. Mackey et al. found that most of the learners who reported noticing the feedback also had higher working memory scores, though the relationship was not found to be significant. Still, Mackey et al. (2002) reported that learners with higher WMC noticed and were better able to make use of interactional feedback that contained grammatical corrections, showing that learners with high working memory capacity might be more efficient at processing input. They also found a potential role for developmental level: “those learners who noticed more may have been those for whom the feedback was most relevant or appropriate” (p. 203).

This study does have limitations, the first being the fact that learners had to recall words as opposed to digits. Recall that it has been argued (see Baddeley, 2003) that digits be used for recall instead of words, so that a “nonsemantic” memory is tapped and so that learners with higher reading abilities do not have an advantage when taking tests. Conway et al. (2005, p. 501), in reviewing previous research, indicate a “potential shared variance between span tasks with words and a measure of higher order cognition, to include reading comprehension, and that this is due to word knowledge.” Therefore, the best kind of test to measure WMC should avoid words for recall. The number of learners per cell in the Low and High WM groups was overall small, with, for example, 8 ‘high’ learners and 6 ‘low’ learners for the Japanese Listening Span test. Also, the researchers created their own WMC spans; had they used the OSPAN, RSPAN, and CSPAN tests, their findings might have been more robust and also generalizable. Limitations aside, this study was the first to pose the critical question of whether or not low WMC prevents learners from being able to successfully notice feedback as corrective of their own erroneous production (something which is argued to be necessary for learning).
Robinson (2005b) also explored working memory capacity, in addition to aptitude and IQ and how these affected implicit and incidental learning. To measure WMC, Robinson employed one reading span task, in which learners read sentences in Japanese and had to recall individual words in each sentence that had been underlined. Set sizes increased from two to five sentences. 54 Japanese learners of ESL were asked to memorize 27 vocabulary items in Samoan. They were then given 10 training trials during which they were shown 45 Samoan sentences. During the testing phase, learners were asked ‘yes/no’ questions on the meaning of the sentences; they were also given feedback in the form of correct/incorrect. In this way, their assessment of any acquisition of Samoan grammar rules (based on the use of ergative e, locative markers and noun-verb insertion) was therefore deduced. While Robinson did not find effects for IQ nor aptitude for incidental learning of Samoan, he did find that working memory was related to the listening grammaticality judgment test used (though, as he acknowledged, this might have been due to the aural mode in which the GJT was delivered). Robinson found that individual differences of working memory and aptitude did not have an effect on incidental learning of artificial grammar.

One limitation of this study is the test used to measure WMC and the fact that only one measure was used to assess WMC. Secondly, while sufficient item size was met (according to the standards set forth by Conway et al., 2005), learners in Robinson’s study had to recall words as opposed to digits, similar to Mackey et al. (2002). Robinson also used an absolutive scoring method, shown by Conway et al. to be problematic and rendering WM span tests incomparable.

Goo’s 2009 study is arguably the most robust investigation of the effects of WMC on L2 learning and feedback due to the methodology he employed with WM span tasks; an OPSAN and RSPAN test and followed the administration and scoring procedures as suggested by Conway et al. (2005). Goo explored working memory capacity and how it mediated the effect of
two types of correctional feedback: recasts and explicit grammatical feedback. 69 learners of English as a FL in Korea were assigned to three groups: the recast group, the metalinguistic group, and the control group. The linguistic item Goo explored was *that*-trace filter in question formations (i.e., Sarah thinks that John hates Mary \( \Rightarrow \) Who\(^{16} \) does Sarah think hates Mary?) All groups carried out a pretest and posttest (question formation, GJT task and written production), while the two experimental groups engaged in two sessions in which they received feedback, differing only in that they either got a full recast or a metalinguistic explanation. Goo found that both of the treatment groups performed statistically better than the Control group, and in fact, that there was no statistical difference between the two feedback groups. However, he found that WMC was a statistically significant predictor in the *efficacy of recasts* on learner performance, while it was *not so* for the metalinguistic group. In other words, he found that recasts may be quite sensitive to WMC variation, whereas metalinguistic feedback may not be.

The main limitation in Goo’s study was that WMC tests were presented to learners as a group, and not on an individual basis. Still, his study posed interesting questions on how WMC modulates the efficacy of implicit feedback.

Révész (in press, b) also explored the effects of recasts on the development of the English past progressive, examining how phonological short-term memory and WMC moderates this outcome. She measured WMC via a nonword repetition test, a digit-span test, and a reading-span test. 90 Hungarian learners of EFL participated in a pretest-posttest-delayed posttest design in which they partook in tasks of more or less cognitive complexity and in which recasts were delivered or not. The control group only participated in the assessments. The task was the same.

\(^{16}\) Note here that one can argue that the grammatical prompt should be “whom” and not “who.” Though not crucial for the present discussion on WMC, this could be considered a limitation in Goo’s design.
as in Révész (2009), where learners described a robbery scene and were either able to look at photos (+contextual support) or not (-contextual support), with the English past progressive as the linguistic item being explored. Learning gains were measured via a GJT, a written production task, and two oral production tasks. To measure phonological short-term memory (PSTM), Révész employed a digit span test and a nonword span test. For the digit span test, learners listened to a series of random numbers in Hungarian and had to repeat the numbers. Digit span was based upon the maximum list length that they could repeat. The nonword repetition task presented 36 nonwords that adhered to Hungarian phonotactics. Participants’ scores were based upon the maximum number of syllables that they could recall “at least twice out of the four nonwords for that length.” Lastly, WMC was measured via one reading-span test. Learners had to read sentences aloud, answer a comprehension question on the sentence, and then recall the final words of each sentence for each set. Their score was based on the average of the maximum final words remembered in each set length. Révész found that recasts resulted in learning, but differentially so for the assessments used to measure development: the greatest gains were made via the oral production tasks, then the written production task and finally on the GJT. Learners with higher PSTM scores developed more on the oral test, while those who had higher WMC developed more on the written production tests. As with her 2009 study, learners who did not receive recasts did not improve at all. This study was insightful in that it showed differential effects for PSTM and WMC on different assessments, which Révész associates as plausibly having to do with differences in declarative and procedural knowledge.\footnote{Interestingly, Mackey et al. (2002) made reference to this very same point, and suggested that learners’ phonological STM measure might being more related to task-based oral interaction, and not so much reading and writing tasks.} Like previous studies however, one limitation of this study is the fact that the verbal working memory test required
learners to recall words as opposed to digits. As put forth by Conway et al. (2005), this method favors learners with higher order cognition and reading comprehension abilities. Furthermore, Révész employed an absolutive scoring methodology when grading WM spans, shown to be problematic and disregarding of performance on other items in the span.

4.7.2 Working Memory Capacity in the CMC Mode

As far as the author knows, only three studies have investigated the construct of working memory capacity in relation to learner behavior or production in the computerized environment, that is, computer-mediated communication (CMC). Payne and Whitney (2002) carried out their investigation under the premise that practice in the CMC mode may lead to L2 speaking ability in the face-to-face (FTF) mode, and that furthermore, interaction in CMC is beneficial to learners with lower WMC. 58 learners of intermediate-level Spanish were divided into two groups: experimental and control, with the experimental group interacting in over 20 teacher-led chat sessions as well as 2 out of 4 regular classroom sessions, while the control group was exposed to the 4 out of 4 normal FTF classroom instruction sessions. The curriculum and activities were held constant across groups. Working memory capacity was measured via a recognition-based nonword repetition task and a reading span task. For the nonword repetition task, learners listened to audio files of nonwords on the computer, and then after each set, had to click on those words they thought they heard. The reading span task was also delivered via the computer. Learners were shown sentences in 7-second intervals. During this period, they clicked on a button that judged whether the sentence was made sense to them or not. After this, participants clicked on those words they thought they had seen as the final words for each sentence. The researchers examined learners’ oral production before and after treatment via scores on the oral proficiency interview (OPI) as based on the American Association of Teachers of Foreign
Languages (ACTFL) Oral Proficiency Guidelines, running an ANCOVA with the pretest and posttest gains as a covariate. They concluded that greater oral gains were achieved in the experimental group, hence, as a result of the interaction in the CMC mode. The researchers also reported indirect evidence that those with low working memory benefited most from the CMC mode. After running correlations between oral proficiency gain scores and nonword repetition scores, Payne and Whitney showed that the correlation was higher for the control group. Specifically, they found that learners with lower “phonological buffering capacity” were disadvantaged compared to some learners in the control group, but that this disadvantage was less so in the experimental group. Payne and Whitney concluded by saying that the CMC environment may be beneficial to learners “with lower ability to maintain verbal information in the Phonological Loop” (p. 23).

This study was the very first to suggest that interaction in CMC may help learners with low WMC. However, there are some major limitations in this study. First, while attempting to achieve high ecological validity in comparing the experimental group with a regular classroom group, the quasi-experimental nature of the research design of this study resulted in an inability to control for external factors as it took place over the course of a semester. Furthermore, all CMC interactions took place as a class. It has been suggested that interaction in CMC should be with one conversational partner, so as to reap the maximal benefits that CMC has to offer (Doughty & Long, 2003). Also, the two WMC spans employed here were a computerized, receptive nonword repetition task and a reading span task. The nonword task presented stimuli to learners orally, and then they had to click on radio buttons for those words they thought they had heard. The reading span task presented sentences to learners in intervals of 7 seconds, and also showed them final words via radio buttons. According to the literature on WMC in psychology,
stimuli should be *immediate* as a function of the learner’s own processing speed. Learners never had to repeat stimuli aloud, tests were not delivered orally, and words, as opposed to digits, had to be recalled (see Conway et al., 2005 for limitations). These tasks arguably do not tap the construct of WMC as well as OSPAN, RSPAN and CSPAN tests. Finally, using only the subjective OPI test as a measure of learners’ “oral proficiency” is a limited means to measure gains in L2 oral production. Perhaps a more robust means of examining improvement in oral proficiency would be via the use of various task-based interactions with production measures of accuracy, complexity and fluency. Such measures would also make the study generalizable to other findings.

In another study, Chun and Payne (2004) explored the relationship between learners’ WMC and free look-up behavior in *CyberBuch*, a CD-ROM program with multimedia annotations for vocabulary items. 13 students of second-year German read a short story on individual computers. A total of 102 annotated words were available in the story, with English and German translations available as well as image or video clips for some words. Using a built-in tracker, the researchers were able to “time-stamp” and record all look-up action performed by the learners, as well as performance on both the comprehension and vocabulary tests. Students also wrote a recall story after they finished, trying to remember everything in the story that they could in English. The two measures of WMC used in this study were tests of nonword repetition and reading span, the same spans employed by Payne and Whitney (2002).

Chun and Payne found a strong correlation between phonological working memory and look-up behavior. They also reported that learners with low WM measures looked up words three times more than learners with high WM, suggesting that students “use the multimedia look-up features of the CD-ROM to compensate for working memory capacity constraints” (p. 481).
This study is unique because it was the first to examine WMC and learners’ interaction with a CALL program. However, the same limitations present in Payne and Whitney (2002) apply to this study in regards to measures of WMC. Another limitation is the small N size, with only 13 participants total, which can potentially affect the power of the findings. Lastly, while learners carried out the comprehension and vocabulary tests after the treatment, they performed the story recall on the next day, after being allowed to read the story a second time.

In a follow-up study to Chun and Payne, Payne and Ross (2005) explored how individual differences in working memory capacity affect the frequency of repetition and other patterns of language use in chatroom discourse. The same measures of WMC as Payne and Whitney (2002) and Chun and Payne (2004) were employed for this study: a computer-delivered, receptive reading span and nonword repetition test. The researchers also investigated oral proficiency, doing so via one speaking task that solicited a 5-minute speech sample that was scored based on a holistic scale. Data collected from 20 chat sessions (from 24 learners of intermediate-level Spanish) was analyzed for occurrences of repetition and relexicalization, as well as average number or words, utterances, and turns generated per chat session. Their findings suggested a connection between working memory and language output in chat. For example, it was found that learners with lower phonological WMC learners produced a greater number of words on average per chat session than high-span learners. No significant differences were found between phonological WM groups for mean utterances or turns generated per chat session. The authors suggested that “students with low phonological working memory span exhibited a distinct chat style with longer utterances (sentences), but not more of them on average per turn” (Payne & Ross, 2005, p. 13).
While Payne and Ross’ study seems to provide indirect evidence that lower WMC learners might produce more in the CMC group, it would have been ideal to compare these same learners to their own interaction in the FTF group. It is therefore difficult to sum up the research exploring WMC as modulated by the CMC or CALL environment, due to the paucity of studies in this area. Limitations are also prevalent in the CMC literature, especially in regards to type of instruments used to measure WMC and methodologies employed to do so.

To summarize those studies in the field of SLA (FTF and CMC) that have explored WMC as a moderator variable, as well as tests and methodologies used, see Table 8 below.

**Table 8. Studies Examining WMC in FTF and CMC: Tests, Methods, and Findings**

<table>
<thead>
<tr>
<th>Author</th>
<th>WMC Test Used</th>
<th>Delivery of Test</th>
<th>Scoring Procedure</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mackey, Philp, Egi, Fujii &amp; Tatsumi (2002)</td>
<td>1 nonword repetition test, 2 listening span tests (one English, one Japanese): learners had to judge grammaticality and recall word in each sentence</td>
<td>Individually</td>
<td>Total number correctly recalled nonwords with up to one phoneme variation; total number correctly recalled words</td>
<td>Low WMC may inhibit learner from being able to notice feedback</td>
</tr>
<tr>
<td>Robinson (2005b)</td>
<td>1 reading span task in Japanese; learners had to recall underlined words</td>
<td>Individually</td>
<td>Total number of correctly recalled words</td>
<td>WMC not related to incidental learning of artificial grammar; was related to GJT in incidental learning of Samoan</td>
</tr>
<tr>
<td>Goo (2009)</td>
<td>OSPAN RSPAN</td>
<td>As a group</td>
<td>Partial-credit unit scoring (Conway et al., 2005)</td>
<td>WMC statistically significant predictor in the efficacy of recasts, not metalinguistic feedback</td>
</tr>
<tr>
<td>Révész, (in press, b)</td>
<td>Digit span and nonword repetition span for PSTM; reading-span test (comprehension + final word recall)</td>
<td>Individually</td>
<td>Max. list length repeatable; Max. number recalled syllables; Total final words recalled correctly</td>
<td>Differences in post assessments: High WMC led to greater gains on written posttests; High PSTM scores led to greater gains on oral posttests</td>
</tr>
</tbody>
</table>

| **CMC Environment** |  |
|---------------------|---------------------|---------------------|---------------------|---------------------|
| Payne & Whitney (2002) | Computer-delivered receptive nonword repetition test and receptive reading span test. | Individually (via the computer) | Max number correctly “clicked” nonwords, recalled words | CMC group achieved greater OPI scores than traditional classroom group. Indirect evidence that CMC better for low WMC learners |
| Chun & Payne (2004) | Computer-delivered receptive nonword repetition test and receptive reading span test. | Individually (via the computer) | Max number correctly “clicked” nonwords, recalled words | Strong correlation between PSTM and lookup behavior. Low WMC learners looked up words 3x more than H WMC learners |
| Payne & Ross (2005) | Computer-delivered receptive nonword repetition test and receptive reading span test. | Individually (via the computer) | Max number correctly “clicked” nonwords, recalled words | Correlation between WMC and language in chat: Low WMC learners produce longer and more utterances in CMC than High WMC learners. No difference for turn-taking |

### 4.8 Summary of the Chapter

The construct of WMC is highly attested in the field of psychology and has been reported to be related to greater cognitive abilities such as reasoning, paying attention, and even
comprehension/learning abilities. The field of SLA is just beginning to explore how this important construct affects L2 learning outcomes, and it seems that WMC is relatively sensitive to the noticing of implicit feedback in the form of recasts—an important finding given that recasts are most common type of feedback provided in the classroom setting (Lee, 2006; Lyster & Ranta, 1997; Mori, 2002; Panova & Lyster, 2002). Unfortunately, there is an overall disconnect between SLA methods and methodologies used in the field of psychology in order to measure WMC. So far, tests used to measure the construct of WMC in SLA are unmethodical, ungeneralizable, and dislocated in accordance with what psychological research indicates as to how they should be administered. This is especially the case for those studies investigating WMC in CMC and CALL contexts. While exploring the effect of WMC is critical to gain an understanding of SLA cognitive processes or how learners receive feedback during task interaction, SLA researchers must be more responsible in their methods and administrative techniques, and should refer to psychological studies and advice set forth by Conway et al. (2005) and Baddeley (2003). Given that recasts are the most common type of feedback and are one of the semantically contingent response types that facilitate the process of negotiation of meaning (e.g., Long, 1996) more studies are needed on this type of feedback. No study has ever looked at recasts in the CMC mode (with one-on-one interaction) and explored how WMC might modulate sensitivity to recasts in a computerized-context. In fact, no one has ever compared feedback in general in FTF versus CMC. A recent meta-analysis by Li (2010) highlighted this important gap in the literature:

Sagarra (2007) stated that feedback provided through the computer is more salient; one anonymous reviewer pointed out that computerized feedback might be more consistent. Therefore, the possibility exists that the mode of delivery may impact the effects of
feedback. To date, no empirical research has been done to compare the two modes of
delivery (Li, 2010, p. 8).

If it is the case that learners with lower WMC might benefit from interaction performed
in the CMC mode (Payne & Whitney, 2002), CMC may be very beneficial to certain types of
learners. One of the major reasons for which Robinson created his taxonomy of task design
features was so that teachers and researchers can best match tasks to learners; indeed,
environment could play a significant role in this learner-task type matching. As Mackey et al.
(2002) explain, “more empirical studies are needed in order for researchers to explore the role
WM may play in specific linguistic domains and learning conditions” (p. 186). This call is
particularly relevant for comparisons of the CMC and FTF environments. Further research in this
area is clearly warranted. The next chapter reviews literature on the CMC mode, describing this
environment for interaction and reviewing studies that have looked at conversation, uptake and
learning in CMC.
CHAPTER 5: THE ENVIRONMENT OF COMPUTER-MEDIATED COMMUNICATION (CMC)

In discussing the potentiality for modality to affect task-based interaction, Ellis (2003), sums up the overall paucity of research exploring this important design component:

There are strong theoretical and empirical reasons for believing that the discourse mode associated with a task will affect the extent to which participants modify their input and output in negotiation exchanges and the type of communication strategies they employ.

… It is no surprise, therefore, to find that discourse mode is an important dimension of tasks. It is somewhat surprising, however, that there has been relatively little attention to this dimension in task-based research to date (p. 93).

So far, task complexity had been tangential to the study of interaction in the computer-mediated communication (CMC) environment. This is surprising given that the incorporation of technological components into language learning curricula has exploded in the past decade, due to the demands of distance learning programs and new forms of technological media becoming popular in mainstream education. A considerable amount of literature has empirically demonstrated that interaction in the CMC mode provides for the same opportunities of interaction, negotiation for meaning, and opportunities for feedback as that which takes place in the oral mode (Darhower, 2002; Kötter, 2003; Iwasaki & Oliver, 2003; Lee, 2001; Pellettieri, 2000; Toyoda & Harrison, 2002; Tudini, 2003). The use of the CMC mode itself has engendered a considerable amount of research, many exploring CMC for different learning outcomes or foci of attention. Yet research exploring different types of tasks, or especially what kinds of tasks are best to promote interaction and learning in the CMC environment, is extremely limited.
5.1 Definition and Description of CMC

CMC is a real-time, synchronous conversation that takes place over the computer via the Internet. Typically, a program is set up that has a divided window for the online ‘conversation.’ On the top half, one can see what their partner has written, while one’s own typing is revealed on the bottom half. Some software incorporates all of the conversational components in one open chat box (for example, iChat). Even though the environment is a written one, it serves as a real-time chat and has been posited to be either synonymous to or as beneficial as face-to-face (FTF) interaction, that is, with the goal of gains achieved in the CMC mode being extendable to the FTF mode. Pellettieri (2000: 59) clearly states this postulation: “…because synchronous [CMC] chatting bears a striking resemblance to oral interaction, it seems logical to assume that language practice through [CMC] will reap some of the same benefits for second language development as practice through oral interaction.”

5.2 Theoretical Premise for Incorporating Tasks in CMC

The theoretical premise for incorporating tasks in CMC is that it is also a medium during which interaction, focus on form, and feedback can take place—all mechanisms of interaction that have been shown to facilitate SLA. Doughty and Long recognize the potential extension of interaction to the CMC mode: “… if learners participate in CMC discussion with one conversational partner, the interaction is very much like that observed in SLA research on negotiated interaction, particularly if task goals are clear” (Doughty & Long, 2003, p. 61). At the same time, CMC is different than FTF interaction in some ways. It does not pose the same immediate response demands as interaction in the FTF mode does. Because the interaction takes place on the computer, learners have the opportunity for extra processing time to respond and
formulate their answers in the CMC mode, processing time that they do not have in FTF. Some researchers see this as beneficial, especially for lower proficiency level learners. It has also been empirically demonstrated that partaking in interaction in the CMC mode equalizes participation amongst classmates (Kern, 1995). Learners then who typically do not participate in the classroom due to introverted personality traits or anxiety may feel more comfortable partaking in classroom discussions over the web. Secondly, the potential of the CMC mode to promote noticing is another key theoretical underpinning for its use. According to Schmidt (2001), noticing of input is crucial for second language acquisition to take place. It has been premised that CMC chat could have a greater potential to promote noticing due to the very fact that the “conversation” is written, it is slower, and allows for learners to review their output (Chapelle, 1998; Hegelheimer & Chapelle, 2000). This suggests that CMC could be an ideal medium in which to carry out tasks and explore different components of task design.

5.3 Research on CMC

Studies investigating the use of CMC chat can be divided into two main areas: those examining discourse in CMC and those that have explored the potential for CMC to promote learning. The majority of these studies comprises the first category and will be briefly discussed below.

5.3.1 CMC and Interaction

It has been observed that CMC does provide the opportunity to negotiate for meaning and to develop one’s interlanguage (Tudini, 2003). In her study, nine learners of Italian participated in real online chat sessions with native speakers. Tudini found that indeed, the CMC synchronous mode allowed for triggers on behalf of the native speakers to communicate their
miscomprehension to the learners, hence provoking a reformulation on behalf of the learner so that understanding could be achieved. She found that it was lexical and morphosyntactic items that most triggered negotiation moves between the NNSs and NSs. Tudini concluded that within CMC chat, negotiation for meaning does take place, and both implicit and explicit feedback are provided. Hence, she showed that the CMC chat environment can facilitate the learner’s second language acquisition as it has been shown to do in oral conversation.

Similarly, Toyoda and Harrison (2002) analyzed types of discourse moves in CMC between advanced level Japanese learners and native speakers (NNS with NS). The dyads were not given a specific task, but rather just chatted with each other using Japanese font. As in Tudini’s study, instances of nonunderstanding triggered negotiations of meaning in this environment. Toyoda and Harrison applied discourse analysis methods to divide the observed negotiation types into nine categories: recognition of a new word, misuse or misunderstanding of a word, ‘pronunciation’ or typing error, grammatical error, inappropriate segmentation (for example, separating a Japanese particle from a noun when it should be bound), abbreviated sentencing (on behalf of the native speakers), sudden topic change, slow response, and intercultural communication gap. They therefore identified many of the same interactional characteristics that take place in the face-to-face mode, and even highlighted some that may be unique to CMC. The authors conclude by suggesting that from a pedagogical point of view, students’ review of their chat logs might facilitate the improvement of their interlanguage, and that learners will definitely pay attention to the language in their logs because they “are their own products” (2002, p. 96). Toyoda and Harrison called for the creation of tasks specifically for the CMC environment so that the quality of chat communication can be improved.
Pellettieri (2000) tested different types of communicative tasks in the CMC mode, ranging from guided conversation to jigsaw activities. 20 intermediate-level learners of Spanish formed NNS-NNS dyads to complete the tasks. Referring to CMC as NBC (network-based communication) Pellettieri found that all of the tasks conducted in NBC fostered negotiation for meaning, and in fact, that “their patterns of interaction look[ed] very much like those seen in NNS oral conversation” (2000, p. 70). As in the oral mode, negotiation for meaning was triggered by clarification requests, confirmation checks, echo or repeated questions, explicit statements of miscomprehension, and inappropriate responses. Pellettieri observed that the negotiations that took place also pushed the participants to create more target-like forms. The majority of problems that triggered negotiation for meaning were lexical, followed by morphosyntactic items—the same findings reported in Tudini (2003). Participants in Pellettieri’s study also engaged in providing and receiving corrective feedback. She concluded by stating that

…because synchronous NBC fosters the negotiation of meaning and form-focused interaction, and because students communicating through this medium have more time to process and monitor their interlanguage, … NBC chatting can play a significant role in the development of grammatical competence among classroom language learners (p. 59).

Pellettieri also postulated that certain features of online chat, such as visually being able to see one’s own output, might be more beneficial in the beginning than oral or face-to-face environment for interaction, because “chatting allows students to practice and gain control over more cognitively demanding aspects of grammar that otherwise might not be so frequently practiced in classroom oral interaction” (p. 82). Like Toyoda and Harrison (2002), Pellettieri called for the creation of CMC tasks that are goal-oriented and designed so that learners are required to request and provide information, which is the only way that tasks may be completed.
successful in CMC. Her findings corroborate with those of Tudini, Toyoda, and Harrison, as well as many others (Blake, 2000; Fernández-García & Martínez-Arbeiaz, 2002; Smith, 2003) reporting that CMC chat gives learners the opportunity to negotiate for meaning, produce output, and develop conversational and grammatical competence, just as the FTF environment does. For a summary of these and other studies that have looked at discourse analysis and interaction in the CMC mode, please see Table 9 below:

Table 9. Studies Examining Discourse and Interaction in the CMC Mode

<table>
<thead>
<tr>
<th>Author</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beauvois (1992)</td>
<td>Online CACD provides unique opportunity for student-student and student-teacher interaction</td>
</tr>
<tr>
<td>Blake (2000)</td>
<td>Task type (jigsaw, information gap, decision-making) makes a difference in CMC; jigsaw promotes most negotiation for meaning. Lexical items negotiated more than syntactic ones</td>
</tr>
<tr>
<td>Chun (1994)</td>
<td>CACD promotes different interactional speech acts and provides learners with opportunity to ‘acquire and practice’</td>
</tr>
<tr>
<td>Darhower (2002)</td>
<td>CMC serves as a learner-centered discourse community</td>
</tr>
<tr>
<td>Fernández-García &amp; Martínez-Arbeiaz (2002)</td>
<td>CMC leads to negotiation of meaning just as in FTF mode</td>
</tr>
<tr>
<td>Iwaski &amp; Oliver (2003)</td>
<td>Japanese NS provide implicit negative feedback to the NNS in chat; NNS use this feedback in their production in this environment</td>
</tr>
<tr>
<td>Kotter (2003)</td>
<td>CMC leads to negotiation for meaning between learners of German and NSs. The CMC mode can contribute to SLA</td>
</tr>
<tr>
<td>Lee (2001)</td>
<td>CMC is an environment in which students negotiate for meaning and reformulate output</td>
</tr>
<tr>
<td>Negretti (1999)</td>
<td>Webchat forces learners to derive new ways for turn taking and chatting with instructor. CMC has unique paralinguistic features (emoticons, uppercase, onomatopoeia)</td>
</tr>
<tr>
<td>Pellettieri (2000)</td>
<td>Task-based synchronous NBC fosters negotiation of meaning</td>
</tr>
<tr>
<td>Shin (2006)</td>
<td>Joint social construction in CMC reveals that CM roles similar to learners’ social roles; learners work together to construct affordances as ‘active social agents’</td>
</tr>
<tr>
<td>Smith (2003)</td>
<td>Task type influences way learners negotiate for meaning in</td>
</tr>
</tbody>
</table>

132
<table>
<thead>
<tr>
<th>Toyoda &amp; Harrison (2002)</th>
<th>CMC leads to negotiation for meaning between NNS and NS of Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tudini (2003)</td>
<td>CMC allows learners to negotiation for meaning, modify interlanguage</td>
</tr>
<tr>
<td>Warner (2004)</td>
<td>CMC chat transcripts from intact German classes are playful in nature</td>
</tr>
<tr>
<td>Xie (2002)</td>
<td>CMC chat promotes communication and enhances reading/writing skills. However, writing in CMC poses unique challenges for learners of Chinese</td>
</tr>
</tbody>
</table>

### 5.3.1.2 Comparing Interaction in CMC versus FTF

Among those studies examining discourse in CMC, some have compared interaction in CMC to FTF. This research has shown that the environments have both similarities and differences. So far, the research shows that CMC and FTF are similar in that they (1) both promote negotiation for meaning (Fernández-García & Martínez-Arbelaitz, 2002); (2) are a platform for the provision of recasts (Lai & Zhao, 2006); (3) are environments in which learners can practice discourse and improve in syntactic complexity (Salaberry, 2000); (4) are both environments in which interactional features such as solidarity, humor, intersubjectivity, and off-task discussion take place (Darhower, 2002).

The research also shows that they are different in that (1) CMC may or may not reduce anxiety (Baralt & Gurzynski-Weiss, in press; Kern, 1995); (2) CMC equalizes participation more than FTF (Bohkle, 2003; Kern, 1995); (3) the first signs of syntactic development may be more visible in CMC than FTF (Salaberry, 2000); (4) there are more uses and types of strategies in pragmatic development in CMC than FTF (Sykes, 2005); and (5) the role of the teacher may be more minimized in CMC than in FTF (Kern, 1995; Meskill, 2005; Sullivan & Pratt, 1996).

Those studies that have compared the two environments are summarized below in Table 10:
### Table 10. Studies Comparing CMC versus FTF Interaction and Discourse

<table>
<thead>
<tr>
<th>Author</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bohkle (2003)</td>
<td>Interaction coded for c-units to compare distribution in CMC versus FTF discussion group. Participation is more evenly distributed in CMC</td>
</tr>
<tr>
<td>Kern, (1995)</td>
<td>Classroom discussions compared. Students had twice as many turns, produced 2-4 times more sentences, used more discourse functions in CMC than in FTF. Less attention to grammatical accuracy, less clear coherence is present in CMC; teacher authority also diminished more in CMC than in FTF</td>
</tr>
<tr>
<td>Lai &amp; Zhao (2006)</td>
<td>CMC chat elicited significantly more self-correction than in FTF mode. However, FTF elicited significantly more instances of negotiation of meaning. Recast frequency for both modalities about equal</td>
</tr>
<tr>
<td>Salaberry (2000)</td>
<td>Compared effects of CMC versus FTF on development of L2 morphosyntactic features. First signs of change in developmental stages are more clearly observable in CMC than in FTF</td>
</tr>
<tr>
<td>Sengupta (2001)</td>
<td>Examined socialization moves that occur in CMC compared to FTF. Students develop a sense of personal accountability more on the web</td>
</tr>
<tr>
<td>Sullivan &amp; Pratt (1996)</td>
<td>Compared writing processes in CMC versus oral classroom setting. Examined writing, student attitudes, teacher centeredness. Writing improved more in CMC mode, however oral mode led to more production. Less teacher-centeredness in CMC mode</td>
</tr>
<tr>
<td>Sykes (2005)</td>
<td>Measured effects of written chat, oral chat, and FTF on pragmatic development, comparing pre and post role plays. CMC synchronous discussion group type has an effect on pragmatic development. CMC chat group also outperformed other two groups in terms of complexity and variety of strategies</td>
</tr>
<tr>
<td>Warschauer (1996)</td>
<td>Compared amount of student participation in FTF versus CMC. Participation correlated to factors of nationality, language ability, time in the U.S., student attitude. CMC mode led to slightly more equal participation than FTF. Language production more lexically/ syntactically complex in CMC than FTF</td>
</tr>
</tbody>
</table>
5.3.2 CMC and Learning

Since 2003, five researchers have expanded the putative benefits of interaction in the CMC environment by exploring whether or not it can lead to learning, or specifically, the acquisition of lexical items (summarized in Table 11 below). Arguably the first to explore learning in the CMC mode, de la Fuente (2003) compared interaction in the CMC versus the FTF environments. She sought to investigate whether, in applying a controlled communicative task, instances of interaction in the CMC mode could be as effective as interaction in the FTF mode in the promotion of receptive and productive knowledge (both orally and written) of lexical items. Participants from three second-semester classes of Spanish were randomly assigned into two groups: oral interaction and virtual chat. The lexical items were 14 fruit, vegetable and seafood Spanish names. The study was a pretest-posttest-delayed posttest design. For the productive scores, participants both spoke (for oral measurements) and wrote (for written measurements) the name of the food item. For receptive scores, participants listened to names and tried to say their names in English (oral) and then were given a list of written words and were asked to write their translation into English (written). During the treatment, participants were given an information gap task in which they had to assign to each other certain food items that they needed their partner to retrieve from the market. The speaker assigning the shopping list was given a picture of seven of the targeted items and told his/her partner to go buy these items, describing them the best that he could in Spanish. The participant who received the shopping list had to negotiate for meaning of each word, while the list-giver had to modify his output so that understanding could be achieved. Participants then switched roles—the shopping list assigner now received a list of seven items to go and buy, while the participant who before received the shopping list had the opportunity to assign food items to purchase. The following day,
participants repeated the activity, but switched roles for assigner/receiver and items. In this way, each participant had the opportunity to negotiate the meaning of every lexical item. Importantly, during the treatment, participants in the oral interaction group (OIG) were given one minute only to negotiate the meaning of each shopping item, while the virtual chat group (VCG) had two minutes. De la Fuente found that both groups—OIG and VGC—had receptive and productive gains in the acquisition of the L2 vocabulary items. However, the OIG outperformed the VCG in written tests (both receptive and productive), a surprising find even though they had not engaged in CMC chat, a written mode. Those differences however were not significant. Both conditions appeared to facilitate the acquisition of L2 words, leading de la Fuente to conclude that “meaning negotiation during interaction (face-to-face or computer mediated) seems to promote acquisition of L2 vocabulary forms” (2003, p. 65). De la Fuente had hypothesized that the CMC group would outperform the FTF group because of previous claims that CMC may lead learners to pay more attention to the targeted forms because it is slower and learners can visually see their output and that of their partner’s. However, there were significant differences between groups for oral productive acquisition; the OIG group had higher productive acquisition one day and one week after the treatment, while the VCG group did not. De la Fuente concluded that type of medium does not affect learning, though face-to-face interaction might be more beneficial than CMC for short-term oral productive acquisition. She reported that CMC is as effective as FTF, but it is not necessarily better (especially in terms of oral production).

One limitation of de la Fuente’s study is the extremely controlled time limitation allotted to episodes of negotiation for meaning. In the CMC group, participants were allowed two minutes to negotiate the meanings of the vocabulary items, while the FTF group was given only one minute. De la Fuente does not provide a justification for this control on time. In fact, in
looking at her data, one can see that the participants appear rushed, and moved on to the next item before finishing the first item (and hence before achieving successfully an understanding of its meaning):

Example 5, Pair 1, Group 2, Day 1 (de la Fuente 2003, p. 72):

sth5> necesito frambuesas  [I need cranberries]
sth5> son rojas y pequenos  [they are red and small]
ag20> en ensaladas o no?  [in salads or not?]
sth5> no, es una fruto  [no, it is a fruit]
sth5> no es caliente  [it’s not hot]
ag20> es un tipo como cerezas?  [is it a type like cherries?]
sth5> no se  [I don’t know]
sth5> es un poco raro  [it is a bit strange]
ag20> no es importante  [it is not important]
ag20> pues, no se ...  [well, I do not know...]
ag20> vamos a tratar las dos y despues eso  [let’s try two and then this one]
sth5> bien  [ok]
sth5> necesito ciruelas  [I need plums]

As demonstrated in the above dialogue, the two participants in pair 1 do not achieve a successful negotiation for meaning before they decide to go on to the next item. Here, it appears that due to the time restraint, participant ag20 suggests that they try the next one and then come back to this one. De la Fuente does not report if participants returned to items that were not successfully negotiated. Therefore, her study was also one in which the imposition of tight timing control was a task design feature. Another limitation to her study is that treatment sessions were held on two consecutive days. De la Fuente did not employ a debriefing questionnaire to ensure that participants did not go back and look up any of the items, or that they had outside influence on top of the treatment itself that might have affected their posttests. Therefore, her results that the FTF environment was superior to the CMC environment (significantly so in oral production and somewhat in written production) are inconclusive;
without such a tight time constraint on negotiation for meaning, different results would have been found.

Smith (2004) did not compare CMC with FTF, but rather looked at the potential for learning of lexical items in just the CMC mode. He aimed to establish a relationship between negotiated interaction and items that were learned in CMC. 24 intermediate learners of ESL were paired together and had to complete jigsaw and decision-making tasks. In employing a pretest-posttest-delayed posttest design, Smith measured written receptive and productive gains. The treatment was conducted over a five-week period, with dyad partners not always the same for each participant. The first session was a training session with chat, while the remaining four sessions served as the treatment. During each session, tasks were limited to thirty minutes. The linguistic items employed were 32 lexical items, with eight for each task. Smith found that items that were negotiated in the interaction were retained significantly better than those items that were not negotiated, and where only preemptive input was given. He defined preemptive input as providing a description of a lexical item even though the partner’s miscomprehension of that item had not been established. His findings were congruous with de la Fuente’s, in that negotiation for meaning does take place in CMC and that it can promote successful L2 vocabulary acquisition. Like de la Fuente, his study does not go without limitations however; the strict thirty-minute time limit enforced proved to be problematic for the successful negotiation of items. Smith acknowledged that some of the participants were not able to finish the task because of the imposed time limitation.

Smith (2005) is an extension of his 2004 study. In this second study, he explored whether a relationship exists between the complexity of the negotiated interaction and learner uptake, and whether or not uptake is a significant predictor of L2 vocabulary acquisition. Simple negotiations
were operationalized as one-phrase negotiations (i.e., two turns) while complex negotiations referred to several phrases (turns) to successfully negotiate an item. Based on learners’ chat dialogues, Smith analyzed all episodes of interaction about a lexical item for instances of negotiation or preemptive input. Instances of uptake were operationalized as types of learner responses to a “negotiation-based focus on form episodes” (2005, p. 44). Three categories were used to code for uptake: No Uptake, Recognize, and Apply. No Uptake was when the learner did not produce any uptake where it was possible. Recognize occurred whenever the learner acknowledged the information that his partner gave him, such as okay, I see, all right, or mmm. Lastly, Apply was when the learner actually produced the target item. Recognize codings were considered as unsuccessful uptake, while Apply was referred to as successful uptake. Smith found that complexity of negotiation episodes (i.e., multiple phrases) did not influence learner uptake. He also found no relationship between the degree of uptake (none, unsuccessful, and successful) and acquisition of lexical items. However, as there were gains in lexical acquisition, Smith posited that simple recognition as a type of uptake may be all that is necessary for predicting acquisition for the CMC environment. Limitations of his study however warrant mentioning. Once again, a time limitation was enforced on learners’ interaction in this study. Given that CMC is slower in nature, imposing time restrictions may affect whether or not learners’ can achieve a successful negotiation of meaning for a word. Also, it is possibly that Smith’s strict definition of uptake led to his results. He only counted successful uptake as typed out production, and not acknowledgement (i.e., mm hmm, sí, etc.). Just because learners did not stop to produce a sentence in the negotiation may not mean that successful uptake has not been achieved. Based on his results, Smith suggested that uptake may have a minimized role in the CMC environment. More research is clearly needed that looks at uptake and how it plays
differently in CMC versus FTF. Smith posed some fascinating empirical questions in regards to uptake and learning in CMC, as his results gave empirical evidence that acquisition based on learner-learner negotiation of meaning in the CMC environment does occur.

A fourth study exploring learning in the CMC mode was carried out by Shekary and Tahririan (2006). The researchers investigated the noticing of linguistic items by analyzing collaborative dialogue via LREs (language related episodes). Sixteen intermediate to advanced EFL learners of English were paired over a course of one month to complete tasks, which were comprised of dictogloss tasks, jigsaw puzzles, and free discussion tasks. Linguistic items were not pre-selected, but rather were chosen based on those items that came up naturally during the chat interactions. Because of this, all immediate and delayed post-tests were individually tailored for each participant, based on what that participant had negotiated in a chat conversation with his partner. All LREs, consisting of a trigger, response and uptake, and were coded for linguistic focus, source, complexity, directness, timing, response, recast, elicitation, and uptake. Unlike Smith (2005), Shekary and Tahririan did consider responses such as ‘okay’ as successful uptake. Negotiation for meaning served as the basis for their operationalization of “noticing.” This is because, according to the authors, “online negotiation can be viewed as a way to promote noticing” (2006, p. 571). Shekary and Tahririan reported that participants focused on form and negotiated for meaning, and that learning took place. To examine the possible relationship between noticing and learning of items, all items that were negotiated in LREs were examined in post and delayed posttests. A relationship was found between noticing and learning; however, their data suggest that it was successful uptake that was the best predictor of learning. Therefore, just negotiation for meaning was not enough for L2 development. The researchers claimed that
successful uptake, where learners actually produced the forms, is what best predicted retention of that linguistic item.

This study was the first to claim that more than just negotiation of meaning is necessary for learning in the CMC environment. However, there are several limitations that make the generalizability of the findings small in scope. First, the concepts of noticing and negotiation for meaning, and of uptake and produced output, seem to be conflated in this study. Unlike Smith, the authors coded recognize as successful uptake; but it is not known that ‘okay’ is indeed indicative that they noticed the form. Second, the tasks they employ are not described nor are provided, eliminating the possibility of replicating their study. Third, while the tailor-made posttests for each participant is an ideal attempt to capture what was negotiated, the administration of the posttests ranged from one to five days after treatment, rendering the design limited. Plus, one half of the pretest items were put on individuals’ tailor-make posttest, while the other half was put on their tailor-made delayed posttests. In the end, the authors decided to combine the post and delayed posttests. This, as well as the time schedule for administering the tests, is a conflation of immediate and delayed retention—two very different and important measures that involve problems with retention of memory. In addition, both lexical and grammatical items were included on the posttests. The authors do not acknowledge that type of linguistic item could have had a differential effect on the findings, as has been shown in the literature (e.g., Gass, Sevetics, & Lemkin, 2003; Leow, 1993; 1995; Leow, Egi, Nuevo, & Tsai, 2003). The posttests employed also measured acquisition by including items of suppliance (where students had to provide a definition of the term), correction (where participants corrected their own nontargetlike output they produced during negotiations for meaning), and spelling (where participants judged the correct spelling of words that occurred in the LREs). These test
items were not motivated in the review of the literature as being valid assessments to measure L2 development. The quantity and depth of limitations of Shekary and Tahiririan’s study begs the question of whether their finding that “simply negotiation for meaning of linguistic items is not enough for learning to take place” can be generalized to the CMC context. However, they were the first authors to use tailor-made tests to try as robustly as possible to capture what learners themselves negotiated.

5.3.3 The Provision of Feedback in CMC to Promote Learning

The most recent study exploring acquisition in the CMC environment did not investigate lexical items, but rather a grammatical construction. It also paired learners with an expert interlocutor as opposed to other learners of the same proficiency level. Sachs and Suh (2007) investigated the effects of textually enhanced recasts on learner noticing and L2 acquisition. Thirty Korean intermediate and high-intermediate learners were randomly assigned into four groups, differing in regards to (1) whether or not the recasts were textually enhanced or (2) whether or not participants had to think aloud. The task they employed was a guided story retelling, in which participants read a story in their L1 and then retold the story to the researcher in English with prompts that attempted to solicit their use of the targeted linguistic item. The linguistic item they employed was the backshifting of verbs in the past tense to the present perfect. Textual enhancement was done via underlining of the matrix verb and bolding of the back-shifted verb. For example, “he said that she had lied about her job.” Feedback with textual enhancement was done via online recasts in CMC chat that contained the researcher’s targetlike reformulation of the non-targetlike output of the participant. Sachs and Suh also coded verbal think-aloud protocols for instances of levels of awareness.
Sachs and Suh reported significant gains in both groups from the pretest to posttest, demonstrating that learning did take place as based on recasts and interaction in CMC. Furthermore, they reported that higher levels of awareness led to better L2 acquisition. Sachs and Suh did not find an effect for written textual enhancement in their study. This study contributes to the small but promising body of literature showing that CMC can indeed serve as a medium for L2 development of grammar and also promotes noticing of forms made salient through the provision of recasts. For a summary of all studies that have explored learning in the CMC environment, please see Table 11 below:

**Table 11. Studies Examining Interaction and Learning Outcomes in CMC**

<table>
<thead>
<tr>
<th>Author</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>De la Fuente (2003)</td>
<td>NNS-NNS learners of Spanish. Compared learning of lexical items in CMC and FTF. Both modalities promote learning, however, CMC might not be as beneficial as FTF in post oral production</td>
</tr>
<tr>
<td>Sachs &amp; Suh (2007)</td>
<td>NNS-NNS learners of ESL. CMC promotes negotiation for meaning, opportunities for recasts and learning. Higher levels of awareness led to better L2 production. No effects for textual enhancement of recasts in CMC</td>
</tr>
<tr>
<td>Shekary &amp; Tahririan (2006)</td>
<td>NNS-NNS learners of EFL. CMC promotes negotiation for meaning and learning. Posttests tailor-made for linguistic items that came up during CMC chat. Successful uptake in CMC is the strongest predictor of learning</td>
</tr>
<tr>
<td>Smith (2004)</td>
<td>NNS-NNS learners of ESL. CMC leads to negotiation for meaning when carrying out task-based interaction. Negotiated items retained significantly better than items that were not negotiated</td>
</tr>
<tr>
<td>Smith (2005)</td>
<td>NNS-NNS learners of ESL. CMC promotes negotiation for meaning and learning. There is no relationship between uptake and learning</td>
</tr>
</tbody>
</table>
5.4 Extending the Study of Task Design to CMC

Though many researchers exploring the relationship between interaction and learning in CMC have made calls for the creation of tasks that are appropriate for the CMC environment, this area of research is so far untouched. Not much has been investigated regarding how different task design features are mediated by modality. Finally, no study has, as of yet, compared feedback provision and learning in the CMC versus FTF modes (Li, 2010). Given that there are theoretical implications for benefits of CMC over FTF, (i.e., CMC is slower and may allow more processing time; CMC may promote noticing more than FTF; CMC may post less restraints on WMC) more research that compares feedback provision (i.e., recasts) and learning in the two modes is clearly warranted.

5.5 The Cognition Hypothesis, Feedback, and CMC versus FTF

The Cognition Hypothesis posits that more complex tasks will result in greater noticing and retention of forms made salient in the input, such as via recasts. So far, only Révész (2009) has explored the combination of tasks of increased complexity and recasts, but in the FTF mode only. She found that recasts given to learners who interacted with the more complex task learned best. Could it be that the CMC environment, given that input is visually salient for learners, will lead to greater learning of forms with more complex tasks? How would interaction of task complexity and modality mediate the efficacy of recasts and learning in both the FTF and CMC modes? These are currently unattested empirical questions in the field, as no one has compared the task complexity differential alongside feedback in the two modes. It is also unknown whether individual differences affect task-based interaction contrastingly in FTF versus CMC. For example, many researchers have suggested that CMC assists in lowering anxiety as learners
carry out tasks (e.g., Abrams, 2003; Chun, 1994; Kelm, 1992; Kern, 1995; Meunier, 1997; Sullivan & Pratt, 1996; Warschauer, 1996), and that CMC may decrease the constraints of WMC in the CMC mode, thereby helping those learners with low WMC (Payne & Whitney, 2002). However, Baralt & Gurzynski-Weiss (in press) recently compared task-based interaction in both modes and found levels of learner state anxiety (measured with a questionnaire) to be almost identical both during and after the task. Qualitative exit-questionnaire comments indicated that their findings may be because learners view the two environments as promoting different outcomes, i.e., accuracy versus meaning. No study has looked at how differences in WMC mediate the learner’s noticing of and benefiting from feedback in CMC versus FTF, and how affective variables would also play a role. Empirical evidence as to how modality alongside individual differences modulates task-based interaction and learning is a necessary next step in the field.

5.6 Summary of the Chapter

The research presented here on the CMC environment for interaction has demonstrated that it is a mode in which negotiation of meaning, noticing, the provision and noticing of recasts, and learning take place, just as they do in the FTF environment. A multitude of studies have explored interaction and discourse moves in the CMC mode. Though this environment has some unique features (i.e., emoticons, equalization of participation, decentering of teacher authority), both the CMC and FTF mode share many characteristics, the most important being that they are environments in which negotiation for meaning, feedback provision, and learning occur, all tenants of Task-Based Language Teaching (TBLT). It seems that this is best accomplished when interaction in CMC is one-on-one and that the goal of the task is closed; not imposing time
limitations may also be necessary to reap the maximum benefits of interaction in CMC. Still, many features of CMC have yet to be explored; the putative benefits of CMC over FTF also need further empirical explanation. Finally, different design features of tasks, especially those of cognitive complexity, have yet to be empirically investigated in CMC and compared to FTF.

5.7 Purpose of the Study

In considering the research reviewed thus far, it is still unknown if task complexity differentially affects learning outcomes in the CMC and FTF environments, given that no study has explored this before. Furthermore, it is not known if a) recasts are differentially perceived in the CMC versus FTF modes, or b) how WMC may moderate feedback differently in different modes. Based on the findings that the FTF group did significantly better in oral posttests and somewhat better in written posttests than the CMC group, de la Fuente (2003) suggested that “face-to-face interactions seem to have a higher ‘noticing’ effect (conscious learning), and they seem to allow for a better mapping of input/output (oral) forms to semantic meaning” (p. 74). What de la Fuente might have been referring to is the postulation that input is richer experientially in the FTF mode than in the CMC, especially when recasts are considered. However, it would seem logical that, because recasts are written in the CMC mode, this visuality of feedback would assist in their being perceived better by learners. It could it be the case, as has been suggested in the literature, that interaction in the CMC mode places less demands on WMC, freeing up space for learners to be able to process feedback. Given that Robinson hypothesizes that increases in task complexity will push learners to think more syntactically and notice input more, especially input made salient (i.e., provision of recasts), and/or that the CMC mode assists
in making feedback more salient, it is possible that task complexity is experienced in distinct ways in the CMC and FTF environments.

Given that previous research indicates a positive effect for increased task complexity on measures of performance (either complexity or accuracy, or possibly even fluency) as well as learning, it is a worthwhile endeavor to examine the nature of increasing task complexity in the CMC mode. Because of the theoretical implications and benefits posed for CMC, the design characteristics contributing to complexity (along resource-directing and -dispersing dimensions) are well suited to apply to the CMC environment. An interesting and important issue is whether performing a CMC task, and tasks at levels of complexity, leads to language performance (i.e., complexity, grammaticization) as well as learning (i.e., uptake and memory for forms in the input) that could potentially transfer to FTF communication. Smith (2005) has suggested that uptake occurs less in CMC; might this be moderated by the complexity of the task, however? Thus, the empirical question of whether interaction in the CMC mode can be extended to the FTF mode (be it performance measures or gains in learning) needs to be explored. Given the increased demands for online instruction, to include computer-based courses for language learning, this is an important area of research in need of more empirical investigation. Indeed, the examination of task complexity and its extension to CALL is one direction in which the field is headed, especially with regard to synchronous CMC environments. A study design that examines task complexity in both the CMC and FTF mediums would address the above currently unattested empirical questions. This is the purpose of the present dissertation.
5.8 Research Questions

In light of the previous research discussed above, the purpose of this dissertation is to test the predictions of the Cognition Hypothesis and to extend task-based interaction and the task complexity differential to the CMC environment, comparing it with FTF. How WMC moderates learning outcomes will also be explored, as will uptake as a mediating variable for learning in tasks of more or less complexity. In considering again the predictions of the Cognition Hypothesis:

Predictions of the Cognition Hypothesis

Increasing the cognitive complexity of tasks along certain dimensions will:

In terms of L2 production:
1) push learners to greater accuracy and complexity of L2 production in order to meet the greater functional/communicative demands they place on the learner;

In terms of interaction, noticing, and uptake production:
2) promote interaction and negotiation work, and heightened attention to, noticing of, and incorporation of forms made salient in the input; and

in terms of individual differences:
3) individual differences in cognitive abilities (e.g. working memory) and affective factors (e.g. anxiety) will increasingly affect task-based performance and learning as tasks increase in complexity” (Robinson, 2001b, p. 56)

this dissertation will specifically explore the second and third prediction, that (1) increases in task complexity will promote interaction and negotiation work, and result in more attention to, noticing of, and incorporation of forms made salient in the input (via recasts) and (2) that
individual differences in WMC will increasingly affect task-based interaction and learning in more cognitive complex tasks. In accordance with all of these considerations, the following research questions are what guide this study:

5.8.1 **Research Questions 1-4: On learning**

1. What are the combined effects of task complexity and recasts on L2 development?
2. What are the combined effects of modality (FTF vs. CMC) and recasts on L2 development?
3. Is there an interaction effect between task complexity and modality when recasts are held constant?
4. For learners in the +complex groups (both in FTF and CMC): Are tailor-made items answered more accurately than non-tailor-made items on the assessments?

*Hypothesis 1 on learning:* Based on the predictions of the Cognition Hypothesis as well as empirical findings from the literature, it was predicted that more complex tasks plus recasts would lead to the most L2 development.

5.8.2 **Research Question 5: On uptake**

5. Does uptake *mediate* L2 development? In other words, if a treatment effect is observed, is it because of the production of uptake during interaction? If so, is the effect different under different experimental conditions?

*Hypothesis 2 on uptake:* Based on the predictions of the Cognition Hypothesis, it was predicted that uptake would mediate L2 development, especially in the more complex tasks.
5.8.3 Research Question 6: On working memory capacity

6. Does working memory capacity moderate L2 development? In other words, if a treatment effect is observed, is the effect different for learners of high or low WMC? If so, is the effect different under different experimental conditions?

*Hypothesis 3 on WMC:* Based on the predictions of the Cognition Hypothesis, it was predicted that WMC would moderate L2 development, especially in the more complex tasks.

5.8.4 Research Questions 7: On independent measures of cognitive complexity

7. How does the combination of task complexity and modality (i.e., FTF+C, FTF-C, CMC+C, CMC-C) affect learners’ reported independent measures of cognitive complexity (anxiety/perceived difficulty questionnaire, time judgments)?

*Hypothesis 4 on independent measures of cognitive complexity:* Based on the predictions of the Cognition Hypothesis, it was predicted that learners’ reported independent measures of cognitive complexity would be affected by increases in task complexity, with the more complex tasks resulting in higher anxiety/perceived difficulty and greater subjective time judgments of time on task.

In the next chapter, the design of the study and the methodological procedures are explained. A pilot study in which the design of the task was finalized is also described in detail. The pilot study was crucial in operationalizing cognitive complexity and Robinson’s Cognition Hypothesis as robustly as possible. It was also necessary to come up with a task designed so that learners could do the task with more or less complexity (1) in a dialogic way with the researcher; (2) in either the FTF and CMC mode; and (3) in a way that attempted to address the methodological limitations of studies reviewed in the past four chapters.
CHAPTER 6: STUDY DESIGN AND METHODOLOGY

This chapter discusses the methodology employed in the current study, first reviewing operationalization decisions in the examination of task complexity and modality concurrently with recasts. Specifically, operationalizations and/or explanations are provided for recasts, uptake, the linguistic item, forms of assessment, as well as decisions for independent measures of cognitive complexity. To follow is a description of the pilot study, in which the design of the task for both FTF and CMC was finalized. A brief history of the development of the task, as well as a justification for the current task used, is provided. Next is a description of the population, the design used in the present study, and the instruments to measure WMC and L2 development. A review of the procedure, coding, and scoring is provided, and finally, a summary of the statistical analyses used.

6.1 Explanation of Terms

6.1.1 Operationalization of Recasts

Generally speaking, a recast is a reformation of a learner’s erroneous utterance “by changing one or more sentence components …while still referring to its central meaning” (Long, 1996, p. 434). There are multiple ways in which recasts can be realized, to include a full or partial recast, a recast that has rising intonation (i.e., Doughty & Varela, 1998), and even recasts with other unique factors such as stress and the number of corrections present in the recast (see Loewen & Philp, 2006, as well as Lyster, 1998, for more detail). Some researchers have suggested that changing other features of the learner’s utterance besides the incorrect form may make the recast more explicit in nature (Ellis & Sheen, 2006). Therefore, in the present study, it was decided to keep recasts as the isolated, declarative type in which the learner’s entire
utterance is repeated and the only corrected component is the linguistic item in focus: the ill-formed verb which, in this study, was the past subjunctive in the dependent clause. Recasts of this nature were chosen due to two factors: (1) they are in accordance with those researchers who endorse recasts as ideal in their ability to draw learners’ attention to form in an implicit in tandem with a focus on meaning (Long, 2007; Long & Robinson, 1998) and (2) for comparability purposes with research done by Révész (2009) and Révész, Sachs, & Mackey (in press) in which isolated, declarative-type recasts were also used.

6.1.2 Operationalization of Uptake

Lyster and Ranta (1997) defined uptake as “a student’s utterance that immediately follows the teacher’s feedback and that constitutes a reaction in some way to the teacher’s intention to draw attention to some aspect of the student’s initial utterance” (p. 49). As with recasts, there are multiple versions of uptake that the learner may produce, to include simple acknowledgement of the recast (i.e., “uh huh”), a repetition of the error, a full repetition of the recast, and even no uptake at all. In a recent study by Egi (2010), uptake was coded for six different types: (1) no uptake, (2) repair: full uptake, (3) repair: partial uptake, (4) uptake with incorporation into a longer utterance, (5) needs repair: repetition of original error, (6) needs repair: simple acknowledgement. Robinson (2007a) explored both partial and exact uptake. Theoretically speaking, uptake is predicted to contribute to learning because it is posited to be equated with noticing the form (e.g., Schmidt’s noticing hypothesis premises that noticing is “the necessary and sufficient condition for converting input to intake,” 1990, p. 129). If uptake is indicative that learner noticed the form, then uptake may be predictive of SLA, which is why it is a variable that has been so heavily investigated. However, some researchers have warned that just because learners do not produce uptake does not mean that they haven’t noticed the form; at
the same time, the production of uptake could also be a mere “mimicking” of the correction and not reflect a true understanding (Gass, 2003, p. 236). Egi (2010) also highlights the fact that uptake could be simple parroting of what the learner’s interlocutor has said. Recall that Robinson’s Cognition Hypothesis predicts that tasks of higher levels of complexity will result in more interaction and more incorporation of forms that have been “made salient in the input” (p. 194). In this study, forms “made salient in the input” are recasts; uptake is operationalized as the “incorporation” of those forms. Here, all uptake was coded following Révész, Sachs, and Mackey’s (in press) scheme. In their codification scheme, uptake was coded as either 1 (full or partial correct repetition) or 0 (no uptake). Simple acknowledgement did not constitute uptake in this study. This means of coding of uptake allows correlations to be run so that uptake as a predictor of learning (which the Cognition Hypothesis postulates) can be explored. As with the operationalization of recasts, this decision was also made for comparability purposes to the existing literature (that by Révész, Sachs & Mackey, in press).

6.1.3 Operationalization of Task Complexity

The task complexity differential operationalized for this study was +/--intentional reasoning. Robinson (2001b, 2007, 2010) explains that by increasing the cognitive demand of tasks along resource-directing dimensions, the learner will be pushed to adjust and stretch their interlanguage resources in order to accomplish the increased demands of the task. This will then result in L2 development. By engaging in intentional reasoning, the learner is required to reflect on the intentional reasons and cognitive mental states that caused other people to do certain actions. This would constitute + intentional reasoning. -Intentional reasoning would simply mandate that the learner describe a person’s actions without having to reflect on the intentions behind those actions. For example, imagine the scenario of a party. John and Mary have gone to
a party together, but later on, John sees Mary dancing with Tom. John storms out of the party. Intentionally reasoning about John’s actions might give way to the following: John stormed out of the party → (i.e., because) John was jealous that Mary danced with Tom. Not having to think about the intentional reasons behind John’s actions mandates that the learner only report those actions: John stormed out of the party. Robinson (2010) explains the linguistic aspect that engaging in +/- intentional reasoning entails. He says that having to +/- intentionally reason:

- has the potential to direct learner attention to, and promote ‘noticing’ (Schmidt, 2001), and internalization of those aspects of linguistic code which can be used to meet these complex task demands (e.g., cognitive state terms such as ‘think’ and ‘wonder’, ‘doubt’ and the complex syntactic complementation that accompanies their use; ‘X wonders if Y’, ‘X doubts that Y believes Z’, etc.) (p. 243).

For this study, cognitive complexity will be operationalized as +/-intentional reasoning, in asking learners to either reflect on the intentions behind people’s actions or not. A detailed description of the task that was designed to naturally elicit intentional reasoning during task-based interaction is described below (section 6.8). As far as the author knows, the only other study to explore the +/-intentional reasoning variable was Robinson (2007a); that study looked at the effects of task complexity on measures of interaction, however, and not L2 development.

6.2 Justification for Linguistic Item

6.2.1 The Spanish Past Subjunctive

The linguistic item chosen for this study was the Spanish past subjunctive. The subjunctive involves a concept known as mood, which grammatically signals how the speaker views the world, a concept known as modality. Modality determines how some languages
grammatically mark verbs in accordance with their mood system. Collentine (2010) explains that modality is “any lexical or morphological expression of one's commitment to the truth-value of a statement,” while mood is “an inflectional representation of modality” (p. 40). Therefore, learners must first learn concepts of modality and then learn which concepts take indicative or subjunctive mood marking in grammar. Pérez-Leroux (2001) also elaborates on the difference between modality and mood, saying that modality, and the subsequent use of modal verbs and mood markers, is a way in which languages encode how the speaker perceives the actuality of an event; mood is “a grammatical category whose function is to describe the actuality of the event in terms such as possibility, necessity, or desirability” (p. 70). She cites Chung and Timberlake (1985) who describe three separate semantic systems for the use of the subjunctive in the Spanish language: epistemic, deontic, and epistemological. The latter semantic system, epistemological, is an attitude-based usage, and is the one used in the present study. For example, verbs that denote emotion, volition (i.e., wishes, desires, needs, wants) are epistemological and take the subjunctive in a dependent clause position. On the other hand, verbs that demonstrate the speaker’s knowledge, or facts, take the indicative mood in a relative clause. Pérez-Leroux contrasts this subjunctive versus indicative notion:

The principal generalization about mood selection with complements seems to be that attitude verbs select subjunctive, whereas neutral verbs select indicative. For instance, the emotive verb lamentar ‘to regret’ takes subjunctive complements, whereas the neutral factive verb saber ‘to know’ selects indicative complements (p. 71).

An example of the use of the subjunctive where emotion, doubt, or instances of recommendation and persuasion are expressed is provided below. Observe that verbs expressing volition are in the
main clause; these mandate that the verbs in the dependent clause be marked with the subjunctive mood grammatical system, as in examples 1 and 2.\(^{18}\)

1. Mamá quiere que vayas a la tienda.

‘Mom wants [that] you go-SUBJ to the store.’

2. Me da lástima que no te hayan llamado todavía.

‘It makes me sad that they still have-SUBJ not called you.’

The use of the subjunctive in the relative clauses takes place when (1) there are two clauses conjoined by the Spanish conjunction que (that); (2) there is a change of subject from the main clause to the subordinate clause (otherwise, the subordinate verb stays in the infinitive form), and (3), the verb in the main clause mandates the use of the subjunctive, hence expressing emotion, doubt, nonexistence, etc., (i.e., modality). To form the past subjunctive, one starts with the 3rd person plural form of the preterit past tense. The past tense number/person morphology marking is removed and is replaced with a morpheme marking mood morphology. Note that there are three class systems of verbs in the Spanish language: -ar verbs, -er verbs, and -ir verbs. For -ar verbs, the preterit ending -aron is replaced with -aran. For -er/-ir verbs, the preterit ending -ieron is replaced with -ieran. An example of this rule change is demonstrated in Table 12 below:

---

\(^{18}\) Many other uses of the subjunctive, be it in the present or past tense, abound, to include prepositional intersubjectivity, negative imperative forms, and time/place/temporal prepositions. These uses, and in particular, uses of the subjunctive in idiomatic phrases and/or main clauses, will not be addressed in this dissertation.
Table 12. Rule-Based System for Forming the Past Subjunctive in Spanish

<table>
<thead>
<tr>
<th>-AR VERBS IN SPANISH</th>
<th>Preterite conjugation for -ar verb</th>
<th>Past subjunctive conjugation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>hablar (to speak):</td>
<td>habló</td>
<td>hablara</td>
</tr>
<tr>
<td>habló</td>
<td>hablaste</td>
<td>hablaras</td>
</tr>
<tr>
<td>hablamos</td>
<td>hablasteis</td>
<td>hablara</td>
</tr>
<tr>
<td>hablaron → hablaron</td>
<td>hablan</td>
<td>habláramos</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hablarais</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hablaran</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-ER/-IR VERBS IN SPANISH</th>
<th>Preterite conjugation for -er verb</th>
<th>Past subjunctive conjugation:</th>
</tr>
</thead>
<tbody>
<tr>
<td>comer (to eat):</td>
<td>comió</td>
<td>comiera</td>
</tr>
<tr>
<td>comió</td>
<td>comiste</td>
<td>comieras</td>
</tr>
<tr>
<td>comimos</td>
<td>comisteis</td>
<td>comierais</td>
</tr>
<tr>
<td>comieron → comieron</td>
<td>comieran</td>
<td>comieran</td>
</tr>
</tbody>
</table>

6.2.2 Acquisition of the Spanish Subjunctive and Developmental Readiness

Collentine (2010) provides an excellent review of research up through the present on how learners acquire the Spanish subjunctive. Some researchers (Farley & McCollam, 2004; McCollam & Weibe, 2004, cited in Collentine, 2010) have examined the acquisition of the Spanish subjunctive via a processability framework (i.e., Johnston, 1995; Pienemann, 1998). So far, there seems to be a general consensus that learners must have reached a syntactic processing stage to be able to learn the Spanish subjunctive. This has also been informed by Collentine’s own work, and what he calls the Syntactic Deficiency Hypothesis. Collentine sets forth that learners must be able to process and produce syntactic complementation, because, “from a cognitive perspective, [the processing of the subjunctive] requires one to process information
across clauses … this is especially because the subjunctive is largely limited to subordinate (i.e., dependent) clauses” (2010, p. 40). Other studies have shown that (1) irregular forms of the subjunctive may be more noticeable to learners at earlier stages (e.g., Collentine, 1997; Gudmestad, 2006); (2) learners may not notice the subjunctive when they only have to interact with comprehensible input and/or are processing for meaning only (e.g., Lee & Rodríguez, 1997; Leow, 1993); (3) some learners may not notice the Spanish subjunctive because of their tendency to focus on lexical content of verbs as opposed to verbs’ morphological endings (Fernández, 2008; Farley, 2004); (4) the requirements of the task in which learners engage may have a large effect on their production of the subjunctive; for example, it has been found that monologic production in which learners only have to produce a phrase’s dependent clause results in significantly less production of the subjunctive (Gudmestad, 2008; Geeslin & Gudmestad, 2008, all cited in Collentine, 2010). Interestingly, it may be the case that lexical classes, and not developmental stages, mediate the acquisition of the subjunctive in Spanish. Gudmestad (2008) found that verbs of volition are most associated with early stages of subjunctive development; other categories, such as uncertainty, doubt, and idiomatic expressions that require the subjunctive are acquired in later stages. Lexical classes such as time reference (i.e., future, anticipated events) and hypotheticality were the poorest predictors of subjunctive use at all proficiency levels. Collentine provides the following synthesis on the research so far in regards to how learners acquire mood in Spanish:

[The research reviewed so far] may imply that as learners progress they initially incorporate the subjunctive only into the verbal system (i.e., it is largely seen as another "conjugation," without any particular communicative value). As they develop, the subjunctive accumulates lexical (or perhaps semantic) features. Another interpretation of
these data is that the syntactic-pragmatic interface can be breached over time (Collentine, 2010, p. 45).

In the present dissertation, it is argued that having to reflect on the intentional reasons for which people do certain actions (Robinson’s Cognition Hypothesis) is congruous with reflection on the epistemological mental states of others which necessitate the subjunctive in syntactic complementation. The Spanish subjunctive (i.e., mood) is therefore as an ideal linguistic item for the study of L2 development with the +/- intentional reasoning variable in regards to Robinson’s hypothesis. Recall Robinson’s 2010 explanation of how tasks that require intentional reasoning will push the learner to syntactic thinking: in reflecting on the cognitive mental state (i.e., she wanted, he was sad that, John was jealous, etc.) the learner must also communicate the “complex syntactic complementation that accompanies their use” (p. 243). This also takes into account what the literature says about the acquisition of the subjunctive: the first lexical class that appears to be acquired by learners of Spanish is volition, or epistemological use.¹⁹ (Note that the intermediate-level participants in this study had gone over concepts of modality and grammatical mood marking rules to form the present subjunctive in Spanish in class, but had never learned the past subjunctive.) In fact, in his review on the acquisition of the Spanish subjunctive, Collentine makes specific reference to the Cognition Hypothesis for contributing to our knowledge on how learners develop syntactically via the study of the subjunctive in embedded clauses:

What is interesting regarding the subjunctive and the issue of complex syntax is that much task-based research has attempted to identify the conditions where linguistic

¹⁹ In having to intentionally reason about people’s actions, the learner must reflect on the epistemological (to include verbs indicating volition) which require the past subjunctive in dependent complements.
complexity will occur (Robinson, 2001). It appears that, during the early stages of L2 development, learners will not be able to focus on the subjunctive's communicative value within a task because they will not have enough processing resources to attend to the subjunctive's formal properties (Foster & Skehan 1999). At later stages of development, however, the research suggests that learners will use the (structurally and semantically complex) subjunctive when they are forced to produce coherent messages, such as when they are to report on or provide some hypothesis about some situation/event, or when they have time to plan what it is they will have to communicate in a task (Robinson, 2001) … Interestingly, much task-based research tries to understand where linguistic complexity will occur based on the nature of the task and the learner's level of development. Thus, the Spanish subjunctive is a prime candidate as a target for understanding task-based instruction's potential and efficacy (Collentine, 2010, p. 47).

As Collentine articulated, more research is clearly needed in this area. The investigation presented here attempts to fill this important gap in the literature.

Finally, one other factor in selecting the past subjunctive for this study is the fact that it has served as the linguistic item of query for past studies (e.g., Bowles & Leow, 2005; Medina, 2009; Rosa & Leow, 2004). The replication of its examination can therefore contribute to this literature on the acquisition of this form.

6.3 Independent Measures of the Cognitive Complexity of the Task

6.3.1 Anxiety/Perceived Difficulty Questionnaire

Since Norris and Ortega’s (2009) call for independent measures of cognitive complexity (discussed in Chapter 3), there is a general agreement that it is critical to triangulate data sources
when investigating the thorny issue of task complexity. Up until now, the only means of gauging learners’ perception of increased cognitive complexity has been via a questionnaire scale measuring their anxiety and perceived difficulty for the task (e.g., Gilabert et al., 2009; Kim, 2009a, Révész, 2009; Robinson 2001a, 2007a). For comparability purposes, an anxiety/difficulty perception questionnaire was created and used in the present study. The other independent measures of cognitive complexity were done in two ways: (1) time judgments (i.e., asking learners after the task how long they believe it took them to complete the task) and (2) a tabulation of qualitative comments said during the task.

6.3.2 Time Judgments

Asking learners to estimate time on task as a measure of cognitive load is supported by literature in the field of psychology. These studies suggest that the greater processing demands are placed on the learner, the more time they will judge has passed when judging time once the task has been completed (indicating a linear relationship between cognitive demands and ‘guesstimations’ of time, e.g., Block et al., 2010; Dzaak et al., 2007; Fink & Neubauer, 2001; Hicks et al., 1976; Paas et al., 2003). In order to tap the construct of cognitive complexity, it was hypothesized in the present study that the more cognitive complex task (+intentional reasoning) would impose greater processing demands on the learner, thereby rendering longer time judgments.

6.3.3 Tabulation of Qualitative Comments in the During- and Post-task Phase

The third measure was a more organic approach to gain insight into learner processes, and included collecting learners’ on-task commentary reflecting their processes and feelings during the task. This included comments such as “this is hard!”, “that was fun!,” “that was easy,” thinking out loud, metatalk, and any other comments or thoughts they produced (which were
subsequently tabulated during data collection). It is important to mention that it was not assumed that not talking aloud indicated a lack of reflection on the requirements of the task; this decision was made once the study began because many learners, without any instruction to do so, said such comments both during and after the task. Tabulating learners’ qualitative comments therefore served to explicate the quantitative findings.

6.4 Measures of L2 Development: Productive and Receptive Assessments

Clearly, using multiple means of assessment as opposed to just one is a more robust way to capture whether or not the learner has developed in his or her linguistic repertoire. One means of doing so is via both receptive and productive tests, which aim to tap different processes (that of recognition and that of application). Both types of assessments were used in the present study, with the receptive test always given last so as not to prime learners with information before having to generate the form themselves in the productive tasks. Also, productive assessments in both the FTF and CMC modes were given. This allowed for a comparison of feedback provision in the two modes and the ability to observe if gains achieved in one mode were extendable to the other.

6.4.1 Tailor-made Component of Assessments in +Complex Groups

A tailor-made component was also added to the three assessments (FTF and CMC productive and receptive) for the complex groups. Recall that Nuevo (2006) questioned whether her assessments properly captured L2 development from complex tasks used in task-based interaction. She postulated that the use of tailor-made posttests, which reflect actual items that arise from learners’ interactions, might best capture what learners come up with in the complex tasks. Nuevo called for future studies to use customized assessments when researching L2
development and task complexity. As explained in Chapter 3, a benefit of tailor-made assessments is that they best capture what the learners themselves came up with during an interactive task. The problem with tailor-made posttests, however, is that they make comparability between participants problematic. To overcome this problem, a combination of ‘critical’ items (those that were tested on everyone) and tailor-made items (those that learners came up with themselves in the complex groups) was used in the assessments for this study. Specifically, two out of the ten items on the FTF and CMC productive tasks were tailor-made (maintaining 80% comparability), and five out of the 15 items on the receptive test were tailor-made (maintaining 66% comparability). Both critical and tailor-made items were epistemological verbs in the primary clause; learners were thus tested on whether or not they knew to mark the verb in the dependent clause with subjunctive mood morphology. It is important to point out that the majority of the ‘critical items’ verbs of volition that all participants, regardless of group, had communicated during the task. The difference is that the -complex groups were told the items during the task; the +complex group had to come up with it themselves. Extensive piloting is what contributed to the selection of the critical items that all participants were tested on, ensuring, for the most part, about an 80% overlap of items. That is to say, on average, the +complex group came up with the 80% of critical items chosen for the study, which the -complex groups were given (i.e., they did not have to come up with the intentional reasons themselves). This is explained in further detail in the pilot and procedure section below.

6.5 Operationalizing +/- Intentional Reasoning: the History and Development of the Task

One of the most difficult and yet important aspects of testing the Cognition Hypothesis is the way in which the researcher operationalizes the construct of cognitive complexity. This is
especially crucial for examining task-based interaction and learning; the researcher must ensure that the task naturally elicits those cognitive demands described by Robinson, i.e., reasoning, communicating an even in the there-and-then reference, taking a perspective on the topic, etc. Coming up with tasks of increased cognitive complexity to be carried out via two-way, dialogic interaction is a particularly exciting challenge for researchers. Looking at findings from research on interaction, Ellis (2003) put forth certain task design features so that interaction and cognitive complexity may best be combined to maximize L2 development. His synthesis of task dimensions thought to assist in SLA according to the interaction approach is particularly informative. Following Ellis’ advice, every attempts was made to come up with a task that (1) required information exchange during the interaction, (2) was of a two-way, information gap nature, (3) had a closed outcome, (4) was a familiar, human-ethical topic that is relative to learners’ lives, and (5) that required narration and collaboration. In total, nine different task versions were tested and tried in order to properly operationalize +/- intentional reasoning that could be done interactively and in both the FTF and CMC modes. In this next section, the pilot study describing each of the nine tasks is provided as well as the justification for why the current task used in the present study was the one chosen. For each task version that was tested, a description and example is provided as well the reason for which that task needed to be improved.

6.5.1 Pilot Study

During the summer of 2009, a pilot study was carried out in which nine different versions of a task were tested. 12 intermediate-level participants partook in the pilot. All participants were students taking intermediate-level Spanish at a private university in the Washington, D.C. metropolitan area. Due to the small population size, this study was not for the purpose of
achieving any statistical significance in findings, but rather to fine-tune all materials, test out the
assessments, technology and equipment, practice coding and analyses for the treatment tasks,
and most importantly, to finalize the interactional task so that +/-intentional reasoning
complexity differential was appropriately operationalized. While the task actually used for the
current study is described later in the materials section, the development of different task features
that were employed during the pilot is described here so that readers can see which features did
and did not work, and why the current task version best captured the notion of +/-intentional
reasoning where recasts could be provided in two different modes.

6.5.1.1 First Task

The task decided upon for this study was a story retell, for which participants had to retell
a story in the past tense and in Spanish. For the first version of this task, participants were given
one story in their L1 (English) that told of an event: a robbery carried out by a trusted
housekeeper. The participants retold the story in either the CMC or the FTF mode. The
+complex group was asked to retell the story and to reflect upon the intentions of the owner of
the house who accused the housekeeper of the robbery. The -complex group only had to tell the
events of the story (i.e., no intentional reasoning). No prompts (such as primary clauses
communicating volition, i.e., she was angry that) were given in this task. The main problem with
this version was that, in only having to tell the events of the story, the -complex group did not
produce, nor did they need to produce, any uses of the past subjunctive. Because recasts are held
constant in this study, all groups needed to produce the past subjunctive (with the idea that the
more complex task will lead to the highest levels of retention of the form and learning). The
researcher therefore needed to fine-tune the task so that the -complex participants did not engage
in intentional reasoning, but still were using the past subjunctive. Another problem was that the
task was too open-ended: responses in the +complex group varied greatly when learners were asked ¿Qué quería Sra. Martínez? (‘What did Mrs. Martínez want?’). Some learners gave one-word answers; others gave a lot of detail. It was decided that a task with better-designed prompts, that is, with the main clause provided, should be tried in order to elicit the past subjunctive in the dependent clause position.

6.5.1.2 Second Task

For the second task version, learners were once again given a story in their L1 (English). It was the same story as in task version 1: a robbery by a trusted housekeeper. In this version, prompts were given to learners that indicated how the owner of the house was feeling and what her intentions were when she fired the housekeeper and then felt intense guilt about her actions. Once again, the task was a story retell in either the CMC or FTF mode. In order to retell the story, learners followed written prompts given to them on paper. Prompts in this task looked like this:
She wanted [that] Srta. Gómez __________ (to change)

This task was an improvement from the first in that it better controlled the answers produced by participants. However, with this design, it became apparent that the +complex group simply had to choose a verb from one of the four options—they were not engaging in real intentional reasoning. Rather, the task design only required that they simply chose a verb. This component was thus improved for the third version.

6.5.1.3 Third Task

For the third version of the task, the +complex group was not given options for the dependent clause. For this version, prompts consisted of the following:
Though the new prompts seemed to improve the design, the +complex group still was not engaging in real intentional reasoning. The researcher realized that by giving the +complex learners the primary clause, she was erroneously giving them the intentional reason. A task needed to be developed so that they came up with the main clause (and subsequent epistemological notion) themselves. This led to the fourth design attempt, which took out the intentional reasoning from the first clause.

6.5.1.4 Fourth Task

For the fourth task, another story was given to learners, specifically, a situation between a boyfriend and a girlfriend, where the girlfriend sees some questionable text messages on his phone and then reacts. This time, for the +complex group, only the events that took place were described, and not the intentions, feelings, or desires of any of the characters. For the fourth task version, prompts looked like this:

-**complex:** Ella quería que Juan ____________ (cambiar).

  *She wanted [that] Juan ____________ (to change)*

-**complex:** Ella quería que Juan …

  *She wanted [that] Juan …*
Participants in the +complex group were told to reflect upon the intentional reasons of the main character in order to explain her actions. While getting closer to properly operationalizing the +/- intentional reasoning parameter, this kind of prompt resulted in massive variation in regards to the reasons the learners provided. For example, sometimes learners did not follow the prompts at all, regardless of the mode in which they were communicating (FTF or CMC). Some participants only provided the main clause expressing volition and did not produce any embedded clauses. For example, a few participants told the intentional reason/affective feeling of the main character, but then did not connect that reason to the action performed by the other character. It was obvious that the task goal was not clear enough for participants. The researcher needed to make the task clearer so that participants knew for whom to relay the intentional reasons and that they’d have to explain the cause of those reasons (i.e., Mary was angry that John called another girl). This then would result in syntactic complementation that required the past subjunctive.

Another problem with this task was related to modality. The use of written prompts, while successful in the CMC mode, created a hybrid in the FTF mode. Some participants in FTF group were confused as to whether the task should be a natural conversation or if they should just read the prompts. They struggled in having to look at the researcher and converse, but also having to refer to prompts in the form of sentences on a piece of paper. A solution for the elicitation prompts needed to be developed, so that a hybridized version was not present in the FTF groups. The researcher then decided to create homemade videos as the source of input for the original story. It was thought that if prompts were picture- or video-based (similar to Révész & Han, 2006), this would eliminate the hybrid problem of written prompts in the FTF mode.
6.5.1.5 Fifth Task

For the fifth task attempt, the researcher made a homemade video that depicted a couple having a disagreement. The idea was that the video could be used as input to eliminate the confusion of the written prompts in the FTF mode. Examples of the video are provided below in Figure 3.

Figure 3. Images from Homemade Videos for Fifth Pilot Task

While controlling the input and story, the homemade videos did not prove to be an effective source of input after it was tried with one participant. It was clear to the researcher that keen documentary-maker skills were needed to create videos, as the overall quality of the video was poor. For the sixth task, the researcher tried using videos from the Internet.

6.5.1.6 Sixth Task

The video for the sixth task was taken from the Monty Python’s Flying Circus, specifically, the ‘Parrot Sketch’ episode, and was tried with two participants. For this task version, the researcher and the participant watched the video together on a large screen in the laboratory. The participant was then shown a PowerPoint presentation that contained images taken directly from the video, similar to prompts used by Révész (2009). On each PowerPoint slide, a prompt was given, requiring participants to either come up with the intentions of the

---

20 This video was recommended to the researcher by one of the participants in the pilot study.  
21 This video can be accessed at [http://www.youtube.com/watch?v=npjOSLCR2hE](http://www.youtube.com/watch?v=npjOSLCR2hE).
characters or not. An example of one of the PowerPoint slides for this task version is provided in Figure 4 below:

**Figure 4. Example of Images from Video Used in Sixth Pilot Task, -Complex**

While the video alongside PowerPoint prompts eliminated the conflict of written prompts in the FTF mode, the monologic nature of the image prompts in a PowerPoint eliminated entirely the rich interaction yielded from the previous dialogic tasks. Plus, one of the participants said she had difficulty understanding the English accents in the video. It also proved very difficult for the researcher to find pre-made videos on the Internet that were of an appropriate genre type to serve as input for this project, and especially videos that, in retelling their events, would require participants to reason about the intentions of peoples actions. It was therefore decided to stick with a story retell so that the dialogic interaction could be maintained and so that the researcher could better control the content of the story. Prompts that were better suited for each mode (CMC and FTF) were subsequently tested out in the seventh task version.
6.5.1.7 Seventh Task

For this version, participants were once again given story in their L1, and this time, the story reported events only. (A husband and wife go to dinner together, to which the wife brings with her an old sweater. Upon arriving home that night, she realizes that she left the sweater there. When they go back to find it, it is gone. She gets very emotional, causing the husband to be quite confused). During the story retell, the researcher did not give participants written prompts, but rather online prompts. In other words, the researcher gave oral prompts in the FTF mode, and written prompts in the CMC mode during the interaction. This was done to eliminate the hybridization problem present in the FTF group in Task 4. Also, and unlike Task 4, the researcher once again gave prompts in this task to better control production. The difference was that for the +intentional reasoning, learners had to connect the intentions or emotional states of the main character (hence connecting epistemological to mood marking) themselves. For the -intentional reasoning group, learners did not have to do so—they were given both clauses. The -complex groups were thus told what the character’s intentions were by the researcher at the prompt. The following is an example of an elicitation prompt and recast in the +complex, FTF group that occurred during the pilot:

researcher: “Entonces, ella quería que …?”

‘So, she wanted [that] ...?’

participant: “Sí, um, ella quería que todos los hombres ayudaron.”

“Yes, um, she wanted that all of the men helped-IND.”

researcher: “Ella quería que todos los hombres ayudaran”

“helped-SUBJ”

participant: “ayudaran.”
This system of prompting assisted in controlled learner production and goals of the task. When the participant did not produce the past subjunctive form, the researcher recasted the entire utterance, replacing the indicative-marked verb with a subjunctive-marked one. In the example above, note that the learner repeated the corrected verb, producing what would be considered uptake of the form.\textsuperscript{22}

To see how this type of prompt was extended to the CMC mode, see the CMC dialogue provided below in Figure 5, also from the pilot study (note that the researcher is in blue speech bubbles, the participant is purple):

**Figure 5. Chat Excerpt from Seventh Pilot Task**

researcher: ‘For her, it was important [that] …?’

participant: ‘they not finish-IND looking’

researcher: ‘For her, it was important [that] they not finish-SUBJ looking’

\textsuperscript{22} Note too that Robinson (2007b) would consider this partial uptake.
Contrary to the +complex groups, the -complex groups did not have to think of the intentional reasons of the characters. The researcher assumed that by providing them with the main and dependent clause, they would not engage in any intention reasoning. So as to not give them the past subjunctive form, the researcher gave the main clause (epistemological) verb and then the verb in the dependent clause, but in the infinitive form. For example, a -complex, FTF elicitation prompt and recast (taken directly from the pilot data) was:

researcher: “Entonces, ella quería que todos los hombres AYUDAR” [prompt with hand and rising intonation of voice]

‘So, she wanted [that] all of the men HELP-INF?’

Participant: “Ella quería que todos los hombres ayudaron.”

‘She wanted that all of the men help-IND’.

researcher: “Ella quería que todos los hombres ayudaran.”

‘to help-SUBJ.’

learner: “ayudaran?”

In the -intentional reasoning group, this same elicitation prompt was done in the CMC mode: learners were given a prompt that told them what the intentional reasoning of the main character was as well as the related, subordinate clause. An example, taken from the pilot data, is provided below (note: researcher is in blue speech bubbles, while the participant is in pink):
Researcher: ‘So, obviously, it affected her greatly that he not to find-INF her sweatshirt.’

Participant: ‘It affected her greatly that I not find-IND [present tense] her sweatshirt.’

Researcher: ‘It affected her greatly that he did not find-SUBJ her sweater.’

Participant: ‘yes.’

The benefit of the seventh task version was that providing prompts online (i.e., during-task) prompts (either in the FTF or CMC mode) led to significantly richer conversational interaction, a key component that had been lost in the sixth task version. The participants also reported that they liked this version much better than the researchers’ previous PowerPoint prompts, and that the hybrid problem of reading prompts while conversing in FTF was no longer present. At the same time, it was still apparent to the researcher that a proper operationalization of the +/- intentional reasoning parameter was not successfully realized with task version seven. Admittedly, one could argue that this type of task was +/- verb provision, especially for the -complex groups.

After much consideration, it became clear that an operationalization of the +/- intentional reasoning parameter needed to be accomplished in a non-linguistic way. In other words, a task was needed where, for the -complex group, it was easy to come up with and reflect upon the intentional reasons of others; for the +complex group, it should not be so clear, and require great
reflection as to why a character did what he/she did. Up until now, the researcher had still been providing participants in the +complex groups with the intentional reasons, desires, and mental states, requiring only that the participants came up with the dependent clause (not doing so led to too much variation). While it is fine to require both groups to use the past subjunctive, a task was needed that better operationalized what Robinson refers to as +/- intentional reasoning, so that (1) it could be done so during conversational interaction and (2) it mandated the use of the past subjunctive. This led to the eighth task version, where two separate stories were created for both groups.

6.5.1.8 Eighth Task

For the eighth task version, the researcher created two different stories, one for the -complex group and one for the +complex group. They stories were written so that it was very obvious what the intentions were in the -complex story, but rather difficult to deduce the intentions behind actions in the +complex version. For the -complex group, a story was written about two lacrosse players who got into a fight on the field. The intentional reasons behind one of the characters actions were seemingly obvious (anger, jealously, frustration). For the +complex group, a different story about a family whose daughter did not want to help with the family chores was composed. The story was written with the assumption that it would be very difficult to come up with the daughter’s intentions. When the researcher tried these stories out with participants however, the two story versions resulted in such different outcomes that were incomparable. In addition, a participant in the -complex group had great difficulty in coming up with the intentional reasons of the lacrosse players, going off on unexpected tangents with the story. He would also often intentionally reason about actions that were not related to the character at all. It became obvious to the researcher that she was wrong in assuming the level of
difficulty and even complexity for someone else. Whether thinking of someone’s intentional reasons is ‘obvious’ or ‘difficult’ is a subjective experience, one that cannot be assumed by the researcher (this very much relates to Norris and Ortega’s 2009 critical point on independent measures to tap the construct of cognitive complexity).

Based on this result, it was decided that the best solution to overcome these methodological limitations would be to use the same stories for retell, but to include the intentional reasons of the characters in the -complex group’s original L1 story. The +intentional reasoning version of the task would get the same story, but their story would only tell the series of events; the intentional reasons behind peoples actions were not provided. The participants in the +complex group would be informed that they would have to think about the intentional reasons that caused specific actions in the story. The -complex group would be told that they only had to relay the sequence of events told in the story; of course, since the intentional reasons were provided in their L1 version, they would also have to produce the past subjunctive. Also, a task needed to be designed so that it was made very clear which actions needed to be related to, a problem inherent to task version 8. The researcher brainstormed, and figured that the L1 story for the ninth task might differ in the following way:

(L1 input)

- complex: Sra. Martinez ran up the stairs. She was angry that Srta. stole the pearls.

+ complex: Sra. Martinez ran up the stairs. [something here to prompt them to reflect on the intentions behind this specific action of running up the stairs. In other words, what caused her to run up the stairs?]
Recall that in the task version 4, learners in the +complex would often relate intentional reasons to actions not relevant to those intentions. The goals of the intentional reasoning component were obviously not clear enough. Therefore, some sort of prompt was needed that guided them to reflect upon the mental states of the character behind the very action that was just reported. The researcher decided that this should be done both in the L1 story and also in the prompt, such as with a directional prompt (action is #1, intentional reason behind that action is #2). Finally, an appropriate prompt that did not pose the same hybrid problems of task version four was needed. After much brainstorming, deliberation with colleagues, and further piloting, it was decided that comics would serve as the best means of input because they are picture-based and allow for minimal written input to be included. This led to the ninth and final task version.

6.5.1.9 Ninth Task

For the final task version, the challenges above were overcome by using colorful comic strips as prompts to elicit (1) a story retell, (2) actions and sequences of events, and (3) the intentional reasons behind those events for the complex group. In using pictures in a comic strip, the hybrid problem of written prompts in earlier versions of the task was eliminated. Also, empty thought bubbles in the comic strip worked extremely well as prompts; by providing the learners with empty thought bubbles, the intentional reasons being characters’ actions were successfully solicited. In order to make very clear the need to relate one’s intentional reasons to a specific action, actions in the comic were marked with a number 1, and then a yellow empty thought bubble, marked with a number 2, immediately followed. This let the learner know that he/she had to reflect on the intentions that caused the action (#1). Where there were empty thought bubbles in the +complex version of the comic strip, the -complex group had thought

---

23 I would like to thank Peter Robinson, Rebecca Sachs, and Laura Gurzynski-Weiss for their help in brainstorming this version of the task at my dissertation proposal defense.
bubbles where the intentional reasoning behind the character’s action were stated (they were
given only the main clause; note too that they were also given this information in the L1 story
before carrying out the task). Also, the L1 story was broken up into paragraphs to reduce the
demands made on participants’ memory. A short comic strip therefore followed after each story
section. Figure 6 below shows how this was done:

**Figure 6. Ninth Pilot Task: Final Task Version for the Dissertation**

-**Intentional Reasoning**

L1 blurb [seen first]:

The Martínez family had a house cleaner, Srta. Gómez. Srta. Gómez always arrived on time,
cleaned the house well, cooked good food, and cared for the children. The family really liked her
and had trust in her. Srta. Gómez recently mentioned that the economic situation of her own
family was bad, but overall, she was a positive and hard working person. Every morning, Sra.
Martínez would wake up and make coffee as she waited for Srta. Gómez. When Srta. Gómez
arrived, Sra. Martínez always smiled. She was happy that Srta. Gómez worked for them.

[Note: the intentional reason behind her action of always smiling is provided in the last line].

comic strip [seen second]:

![Comic Strip Image]
Intentional Reasoning

L1 blurb [seen first]:

The Martínez family had a house cleaner, Srta. Gómez. Srta. Gómez always arrived on time, cleaned the house well, cooked good food, and cared for the children. The family really liked her and had trust in her. Srta. Gómez recently mentioned that the economic situation of her own family was bad, but overall, she was a positive and hard working person. Every morning, Sra. Martínez would wake up and make coffee as she waited for Srta. Gómez. When Srta. Gómez arrived, Sra. Martínez always smiled. 🌼

[Note: the intentional reason behind her action of always smiling is not provided. Rather, an empty thought bubble is given in the story, prompting the participants so that they know they will have to reflect on the intentional reason that cause her to always smile].

comic strip [seen second]:

Observe that the +complex version has an empty thought bubble with a number 2 in it. Here, the participant must come up with the intentional reason behind the action of the character, marked with a number 1 and connect the two. Empty thought bubbles were also given in the L1 story, so that the participant knew where he/she would have to reflect on the intentional reasons of actions. Piloting with this version of the task proved that it successfully operationalized +/- intentional reasoning and worked very well in promoting dialogic interaction in both FTF and CMC. For the finalized tasks (L1 stories and all accompanying comic strips), please see Appendices A through D.
6.6 The Current Study: Research Design and Methodology

6.6.1 Participants

An original pool of 86 participants signed up for the current study. Participants were adult-aged students of Spanish as a Foreign Language and were enrolled in the 4th semester Intermediate II Spanish class at a private university in the Washington D.C. metropolitan area. They were all part of a program that was communicative in nature and used assessments measuring the four areas of speaking, listening, writing and comprehension. The students were placed in the Intermediate II level course based on either a) having successfully passed Intermediate I or b) testing into the level. Students of this course received three 50-minute-long sessions of instruction per week. At this point in the curriculum, students had extensively studied the Spanish past tense (both imperfect and preterit) and had covered the concept of the Spanish present subjunctive in the curriculum.

From this original pool, participants were eliminated from the study for: (1) demonstrating prior knowledge of the past subjunctive on the pretests; (2) missing sessions; (3) demonstrating lack of effort expenditure during the treatment (for example, one participant admitted he had not slept for 36 hours and could not function at his best); (4) engaging in rehearsal during the working memory tests or not meeting 85% accuracy on the processing component of one of the working memory tests, and (5) admitting to having looked up the past subjunctive form on the exit questionnaire. In the end, a total of 16 participants were eliminated for these reasons. The final group of participants for the study was 70 in total.

For all groups, participants were between the ages of 18 and 23 (the mean age was 19.26 years, SD = 1.31). Only 7 participants were not native speakers of English; 6 participants spoke Korean as an L1 and 1 learner spoke Mandarin Chinese. 2 participants reported that they had
learned creole as a first language, but had since reverted to English given that they were raised in the United States (their first languages were Haitian Creole and Ivory Coast Creole). The majority of participants reported that they had never traveled to a Spanish speaking country before, with the exception of 1 that had traveled to Cancun, Mexico, 3 who had traveled to Latin American for service trips, and 1 participant who had visited Spain with his family. Most students had studied Spanish since high school, with the mean overall length of Spanish study at 5.5 years (SD = 2.84). In terms of technology, all participants reported that they owned a computer, all participants had a Facebook account, and all used the Internet and word processors daily. Almost all engaged in some sort of CMC chat online to communicate with friends and family, with most reporting that they used Facebook chat or Gmail chat. Participants’ data in terms of age, gender, mean length of study of Spanish, and hours per week of reported chatting online are provided in Table 13 below:

Table 13. Biodata Information per Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Age M (SD)</th>
<th>Gender</th>
<th>Mean Length of Spanish study (SD)</th>
<th>Average hours per week chatting online (i.e., Gmail chat, Facebook) (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>19.2 (1.6)</td>
<td>11 females 7 males</td>
<td>5.7 (3.4)</td>
<td>2.0 (1.7)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>19.7 (1.5)</td>
<td>10 females 8 males</td>
<td>4.8 (2.2)</td>
<td>1.0 (1.2)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>18.8 (0.8)</td>
<td>8 females 9 males</td>
<td>6.8 (2.9)</td>
<td>3.5 (3.6)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>19.3 (1.1)</td>
<td>12 females 5 males</td>
<td>5.0 (2.6)</td>
<td>3.3 (2.8)</td>
</tr>
</tbody>
</table>
6.6.1.1 Group Comparability

To ensure group (FTF+C, FTF-C, CMC-C, CMC+C) comparability in terms of background, one-way ANOVAs were run on the variables of age, gender, mean length of Spanish study, and average hours per week engaging in CMC chat. There were no significant differences among the four groups in regards to any of these variables ($p = .63, p = .72, p = .46,$ and $p = .24,$ respectively).

As reported earlier, from the original group of 86 that signed up for the study, participants were eliminated from the final analysis if they showed prior knowledge on the three assessment tasks. Specifically, if learners produced any form in the past subjunctive on the FTF or CMC production pretest assessment, they were eliminated. If learners correctly answered more than 4 out of 15 items on the multiple-choice receptive test, they were eliminated. After those with prior knowledge were removed from the final pool, one-way ANOVAs were performed on the pre-receptive test scores and the two pre-productive test scores, FTF and CMC, and confirmed no significant differences among groups with regard to prior knowledge ($p = .09$ for the receptive test, $p = .06$ for the FTF productive test, and $p = .47$ for the CMC productive test).

6.6.2 Design

A pretest-posttest-delayed posttest design was employed for this study. The 70 participants were randomly assigned to one of the following four experimental groups:

1. FTF mode, +complex (FTF+C)
2. FTF mode, -complex (FTF-C)
3. CMC mode, -complex (CMC-C)
4. CMC mode, +complex (CMC+C)
Each group differed in terms of task complexity, i.e., whether or not they had to engage in intentional reasoning during a two-way, dialogic story retell task and modality, i.e., whether they carried out the interaction with the researcher in the FTF or CMC mode. Therefore, the independent variables, or between-group factors, were task complexity and modality. The dependent variables, or within-group factors, were gains made through time, i.e., from the pretest to posttest to delayed posttest on the three assessments. Other dependent variables were learners’ anxiety/perceived difficulty levels as measured by a questionnaire, learners’ time judgment of the duration of the interaction, and for the two complex groups, correct responses to tailor-made items versus non tailor-made items on the assessments (in other words, items that they themselves came up with during the interaction as a trigger for the past subjunctive in the dependent clause position). Finally, uptake and working memory capacity were the covariates, with uptake as a mediating variable and WMC as a moderating variable.

6.6.3 Treatment Task

The task type chosen for this study was an interactive dialogic story retell. This task required participants to read sections of a story in their L1 (English) and to then retell each section, all in Spanish and in the past tense, using a picture comic strip prompt to assist them. The task had six comic strip prompts in all, each preceded by a small section of the story in English. Therefore, the participant and the researcher both had a total of 12 cards (the same set of cards each). They first read the L1 blurb to themselves in English, and then moved on to the next card (the picture comic strip), where the participant had to retell that section of the story. They then went to the next card (and so on). Two versions of the task were created. One task was about a family living in Latin America and their housekeeper. One day, a set of pearls goes missing, and the family concludes that it must have been the housekeeper who stole them, even
though they have trusted her for years. They fire her, only to discover the lost pearls later on in a suitcase. They realize that they made a terrible mistake. The second story is about two young men in the city of Maracaibo, Venezuela, who love to play soccer. One day, a professional scout approaches them and invites them to practice with the city’s team. One of the boys is exceptionally good at soccer, and pushes his friend, who is not so good, to go to the tryouts with him. The story deals with their experience at the practice and how at the end, the better soccer player must choose between playing with a professional team or helping his friend. The comic strips for both stories were created with Comic Life™ version 1.4.1. All pictures for the comics were purchased from stock.xchng.

The two versions of the task were piloted with four native speakers of Spanish and with four intermediate-level learners of Spanish from the same population as the participants. The native speaker baseline data as well as the pilot data with learners indicated that the two tasks were similar in terms of task duration (time). Importantly, data from the native speakers and learners showed that the two tasks naturally elicited the past subjunctive. However, for the complex versions of the tasks, it was found that both native speakers and the learners had a tendency to use the conjunction ‘porque’ (because), i.e., Luis was sad because the other guys made fun of him at the practice. While also perfectly natural to do in the Spanish language, the porque conjunction does not elicit the past subjunctive test. In fact, most of the native speakers reported that their natural tendency was to use the Spanish copular verb estar (to be) with an adjective (i.e., she was sad) in an intransitive way, thus not requiring the subjunctive in the dependent clause. The use of copular estar +adjective with subjunctive in the dependent clause requires the use of the conjunction de que (that). But for some of the native speakers, the

---

24 Comic Life is created by Plasq.
25 http://www.sxc.hu/category/100
tendency was to use attitude-based verbs, and not intransitive estar + adjective, in conjunction with the past subjunctive. At the same time, all four of the intermediate-level learners overwhelmingly used copular estar + adjective when reflecting about the intentional reasons and mental states of the characters in the story. They reported that this was more natural for them. The researcher therefore decided on two things: (1) to include both transitive and intransitive items, specifically, 5 of each, as critical items to be learned and (2) to devise some strategies to employ during the conversational interaction so that learners would use the ‘de que’ conjunction and not ‘porque’ (because) following estar + adjective. This included saying ‘de que’ immediately after the learner provided the emotion or intentional reason, so that he/she would continue on with that prompt. With these strategies, Task A was consequently tried out with one more intermediate-level learner and the task was finalized.

6.6.3.1 Recasts During the Treatment

As stated above, recasts were kept constant throughout this study. Therefore, during the two treatment sessions, participants were provided with a recast whenever they made an error with the past subjunctive. Recall that Révész (2009) sought to compare the effects of recasts versus no recasts, and found that recasts in the more complex tasks led to the most learning. Her study therefore showed that recasts work. This study aimed to explore how task complexity and modality differentially modulate the efficacy of recasts, so recasts were kept constant. All recasts repeated the learner’s entire utterance, both the primary clause with the intentional reason/emotion and the dependent clause containing the new referent and verb marked with past subjunctive morphology. Similar to Révész (2009) and Révész, Sachs, and Mackey (in press), recasts were of a declarative, isolated-type nature and had no added emphasis to them. All effort
was made so as to not change anything else in the utterance besides the past subjunctive form. If learners did correctly produce the past subjunctive, a recast was not given.

6.6.4 Assessments

Three different assessments were employed to measure L2 development in this study: two productive tasks and one receptive task. As with the treatment, the two productive tasks were story retells done in both modes: CMC and FTF. Having learners carry out a brief story retell in both modes (during which the past subjunctive is required based on people’s emotions and intentional reasons) was deemed crucial so that conclusions could be made as to the potential of transfer of learning from one mode to the other (something which has never been empirically attested before in relation to task complexity and recasts). For example, whether or not performing a task in the CMC mode leads to language performance and learning (uptake and memory of forms) that transfers to the FTF mode, and vice versa, is important to investigate. The receptive task was a multiple-choice test delivered via the computer. The pilot study carried out during the summer of 2009 revealed that on the original paper-version of this test, learners would go back and change their answers. Some reported that they ‘figured it out’ and would go back to correct previously answered items. A computerized version of this assessment prevented learners from changing answers and/or from going back and looking at forms. All of the receptive multiple-choice tests were created using Blackboard©. There were multiple benefits in delivering multiple-choice assessments via the computer. First, all items could be statistically randomized. Second, the test features were designated so that participants could not go back and look at/change prior answers and so that the test had to be completed in one sitting. Lastly, a password could be created for each participant and each test in Blackboard©, and scores were uploaded to an online database in the Blackboard© grade center. Blackboard© also reported
188

information for the researcher such as date, time, and duration length for which the test was taken. For all receptive assessments, the following options in the Test maker in Blackboard® were selected: test availability, forced completion, password, and no feedback (such as test score) to be displayed to the participant upon test completion. An example of the test design interface in Blackboard® is provided below in Figure 7:

Figure 7. Blackboard© Test Design Features for Multiple-Choice Receptive Test

![Test Design Features](image)

6.6.4.1 Critical Items used for Treatment and Assessments

For each of the three assessments (FTF productive, CMC productive, and multiple-choice receptive), three versions were created. The ordering in which they were presented to each participant (pretest, posttest, and one-week delayed posttest) was counterbalanced so as to avoid
testing effects; however, the MC receptive task was always given last, after the participant had done the CMC and FTF productive tasks. This was so that it would not provide participants with the form or any other means of input. For all of the assessments, the following ‘critical’ items were selected (see Table 14 below). Half required the ‘que’ conjunction and half with *estar* plus adjective required the ‘de que’ conjunction in transitive constructions. All were verbs denoting emotion or subjunctive evaluation and/or judgment.

### Table 14. Critical Items used for Treatment and Assessments

<table>
<thead>
<tr>
<th><strong>estar</strong> + adjective + de que</th>
<th><strong>verb</strong> + que</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estaba contenta de que (was happy that)</td>
<td>le enfadó que (it angered her/him that)</td>
</tr>
<tr>
<td>Estaba triste de que (was sad that)</td>
<td>dudaba que (doubted that)</td>
</tr>
<tr>
<td>Se sentía mal de que (felt bad that)</td>
<td>quería que (wanted that)</td>
</tr>
<tr>
<td>Estaba sorprendida de que (was surprised that)</td>
<td>no pudo creer que (could not relieve that)</td>
</tr>
<tr>
<td>Estaba molesta de que (was annoyed that)</td>
<td>odiaba que (hated that)</td>
</tr>
</tbody>
</table>

All of the items were selected from the textbook used by the curriculum in which all participants were enrolled (*¡Avance!*). In one of the chapters on the present subjunctive, a list of almost 100 common items that take the subjunctive in the dependent clause position in Spanish is given. All of the above critical items came from that list.

### 6.6.4.2 Multiple-Choice Receptive Test

The multiple-choice receptive test was designed to assess participants’ recognition of the past subjunctive. It consisted of 32 sentences total, with four possible options for each sentence. 15 of the items represented standard uses of the past subjunctive. Of these 15, 10 were the critical items shown above, while 5 were other, similar attitude-based prompts. An equal amount of Spanish verbs from the -ar verb class and from the -er/-ir verb classes were provided in the dependent clause position for all of the multiple-choice items. 6 items tested learners’ knowledge of the present subjunctive, also with an equal divide of dependent clause verbs from either the -ar
or -er/-ir verb system in Spanish (3 with -ar verbs in the dependent clause, and 3 with -er or -ir verbs in the dependent clause). In addition, 6 items were used to test learners’ knowledge of the preterit past tense. These two categories of items on the receptive test were used to gain insight on learners’ developmental readiness for the past subjunctive, given that the preterit and present subjunctive must be learned before being able to deduce a rule for the past subjunctive (recall that the mood morphology marking is applied to the 3rd person plural preterit form). Lastly, 5 items on the test were distractor items. Distractor items covered basic grammar and lexical points that were covered in the Intermediate II course. Examples of each item type that learners saw on the multiple-choice receptive test are provided in Table 15 below:

**Table 15. Item Types Presented to Learners on the Multiple-Choice Receptive Test**

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Number of items per test</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Past subjunctive</td>
<td>15</td>
<td>Susana estaba triste de que Juan no la ________ (llamar).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marta deseaba que los chicos se ________ (portar) bien.</td>
</tr>
<tr>
<td>Present</td>
<td>6</td>
<td>Papá quiere que le ________ (ayudar) ahora en la cocina.</td>
</tr>
<tr>
<td>subjunctive</td>
<td></td>
<td>¡Qué bien que tú ________ (venir) a visitarnos!</td>
</tr>
<tr>
<td>Preterit past</td>
<td>6</td>
<td>La semana pasada, nosotros ________ (ir) a un restaurante buenísimo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>El año pasado, mis hermanos me ________ (hacer) una tarjeta para mi cumpleaños.</td>
</tr>
<tr>
<td>Distractor</td>
<td>5</td>
<td>José tiene una familia muy ________.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(cerca, unida, cercana, lejana).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yo he ________ la puerta.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(abro, abra, abierto, abierta)</td>
</tr>
</tbody>
</table>

Once item types were created for all four categories of past subjunctive, preterit, present subjunctive, and distractor, the three versions of the test were created by randomly selecting the designated number of each item type for each test. The three tests were piloted with two native
speakers of Spanish who correctly answered all items and gave feedback on some of the sentences, to include selections for the distractor items. The tests were subsequently modified to account for the native speakers’ recommendations. Next, all three tests were piloted with three Intermediate II level classes of 14, 13, and 13 students respectively, for a total of 40 students. Cronbach’s alpha was calculated to determine the internal-consistency reliability of each test, and all were in acceptable ranges of one another (.82, .86, and .80). For this piloting of the test versions, multiple-choice tests were administered to the students in a split-block design to avoid test effects.

6.6.4.3 FTF and CMC Production Tasks

Two productive tasks, one in the FTF mode and one in the CMC mode, were used to assess participant’s ability to generate the forms during both oral and chat interactive conversation. As stated earlier, it was important to do two productive tasks in both modes so that if gains were made in one mode during the treatment, it could be seen if those gains were extendable to the other mode at the post assessments. As with the treatment, these production tasks were story retells in the past tense. The researcher did not give recasts during these tasks, but only helped learners with lexical items if needed. Participants were first given a short story in their L1 and were told to read the story entirely before beginning the story retell. When they finished, their job was to retell the story in sections to the researcher to the best of their ability. The stories were all split into three sections. In between each section, the participant was told to read the L1 portion so as to not put too many demands on his or her memory of the story. Each productive assessment story retell had 10 critical items that required the use of the past subjunctive. The FTF assessments were all carried out in person with the researcher, while the CMC tasks were carried out via iChat version 4.0 on the computer; both were done one-on-one
with the researcher. Six versions total of the productive retell task were created, three designated for the FTF mode and three for the CMC mode. All six versions were piloted with the same group of native speakers and students as the multiple-choice receptive tests. Based on the native speakers’ feedback, minor revisions were made to each task, especially in terms of ensuring comparable word length and lexical density. The stories were also slightly modified so that vocabulary that students at the intermediate-level were familiar with was used. To test this, all of the six story versions were distributed to 29 more intermediate-level Spanish learners of the same population as the participants in this study. All of the students were asked to translate the passages as best as they could in Spanish and in the past tense. This was to ensure that participants at this level had no difficulty understanding the meaning of the stories and could translate them with ease. With the exception of a few lexical items, all participants translated the stories successfully, using indicative past tense verbs where the past subjunctive was needed.

To format each story for productive retell, all six stories were divided into three sections that the participant would have to retell at one time; each section ranged from 70-90 words. Lengths of the passages for each story version were: CMC version A, 271 words; CMC version B, 242 words, CMC version C, 269 words; FTF version A, 351, FTF version B, 275, and FTF version C, 295 words. Participants were told that they could look at the L1 story version if they needed to during the story retell. All three stories had the same number of obligatory contexts for 10 items requiring past subjunctive, with 5 -ar verbs and 5 -er/-ir verbs in the dependent clause position. A list of the verbs requiring subjunctive marking for all of the FTF and CMC production tasks is provided below in Table 16:
Table 16. Items for the Production FTF and CMC Assessment Tasks

<table>
<thead>
<tr>
<th></th>
<th>CMC Productive Task</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Version A</td>
<td>Version B</td>
<td>Version C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>-AR verbs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Planear</td>
<td>Internalizar</td>
<td>Estar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disfrutar</td>
<td>Estudiar</td>
<td>Ayudar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Regresar</td>
<td>Usar</td>
<td>Participar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Encontrar</td>
<td>Decepcionar</td>
<td>Desear</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Portarse</td>
<td>Mirar</td>
<td>Lastimar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>-ER/-IR verbs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Salir</td>
<td>Ser</td>
<td>Vivir</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sentir</td>
<td>Tener</td>
<td>Tener</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poder</td>
<td>Poder</td>
<td>Decir</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ir</td>
<td>Ser</td>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Entender</td>
<td>Mentir</td>
<td>Permitir</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FTF Productive Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the three versions of productive tasks for the CMC and FTF modes, please see Appendices E through I. The directions for the CMC and FTF were identical, except for the mode in which participants were told they’d be communicating. As with the multiple-choice
receptive tests, the three versions of the CMC productive and the FTF productive assessments tasks were distributed to participants in a split-block design to avoid test effects.

6.6.5 Background and Bio-Data Questionnaire

Before carrying out the experiment, all participants were given a background questionnaire. This instrument asked for their bio-data information such as age, gender, duration of time studying Spanish, and whether or not the participant spoke a different first language other than English and knew any other languages. Participants were asked about the type of activities they do in Spanish, reasons for studying the language, and how many hours a week they engage in speaking, reading, and listening activities in the Spanish language. In accordance with the purpose of the study, the second half of the questionnaire sought information about the participants’ background in regards to technology. Participants were asked to indicate how many hours a week they used a word processor, the Internet, if they were members of a social networking site, and whether or not they engaged in online CMC chat and/or other chat types online, such as video chat or voice chat. They were asked if they did any of these activities in Spanish or other foreign languages. The questionnaire instrument concluded with open-ended questions, soliciting insight on the participant’s perception of technology for language learning. They were asked to list technologies that they believe can or cannot assist in learning a foreign language, types of technologies that they have used or that their teachers have used in the past, as well as their opinion for online means of communicating in another language, to include CMC chat and even video chat. They were also asked about using Spanish in person versus via the computer, and preferences, pros and cons with each. Please see Appendix J for the Background and Bio-Data Questionnaire.
6.6.6 Anxiety and Perceived Difficulty Questionnaire

In accordance with previous literature, an anxiety and perceived difficulty questionnaire was created to gauge insight into how learners’ perceive the task (Appendix K). So far, studies have found an effect for task complexity on learners’ anxiety/perceived difficulty levels, with more cognitively complex task leading to higher reported levels of anxiety and perceived difficulty (Gilabert et al., 2009; Kim, 2009a, Révész, 2009; Robinson 2001a, 2007a). Such a questionnaire was therefore created for comparability purposes. For the present study, the goal was to measure learners’ state anxiety and perceived difficulty level as caused by task-based interaction with (1) a task of high or low complexity and (2) in the FTF versus CMC mode. A total of 15 items were present on both questionnaires. Learners were asked to indicate on a 6-point Likert-scale how much they agreed or did not agree with each statement (with options ranging from SD, strongly disagree, to SA, strongly agree. As suggested by Dörnyei (2003), both regular order and reverse order statements were included in the instrument.

To test the reliability of the instrument, Pearson’s Correlation was calculated for every item on both questionnaires using SPSS 18.0. The purpose of this analysis was to see if any items on the instrument were redundant and thereby needed to be discarded, and also to ensure that each item should be included on the questionnaire. No item had a negative correlation or had a correlation higher than .9 with any other item, rendering the inclusion of all items originally on the questionnaire. Correlations were then performed to measure the construct validity of the questionnaire. Analyses of reliability using Cronbach’s alpha were high, with .82 for the questionnaire after treatment 1 and .88 for the questionnaire after treatment 2, indicating that the instrument did set out to measure the construct of anxiety and difficulty perception after each treatment session.
6.6.7 Working Memory Spans

As is recommended by Conway et al. (2005), three working memory span tasks were used to determine learner’s working memory capacity (WMC): an operation span task (OSPN), a counting span task (CSPAN), and a reading span task (RSPAN). All three of the spans force “WM storage in the face of processing” (p. 773), and the reliability of these instruments has been highly attested in the psychology literature. All spans were created by Dr. Engle and associates at Georgia Tech University\(^{26}\) and were adapted to be used by PowerPoint. Their description follows below.

6.6.7.1 The Operation Span Task (OSPN)

This task was created by Engle et al. (2005).\(^{27}\) Though their task was meant for the E-prime software, the researcher adapted the span for PowerPoint so that it could be delivered in person and individually. As designed by Engle et al. (2005), the OSPAN task presents arithmetic problems to the participant on the computer. The participant is asked to read and solve the problem aloud. After solving a problem, the participant is shown a second screen where a possible answer was given. This screen also showed a prompt: “true” or “false.” Based on his/her calculation, the participant had to say aloud if that answer was true or not. Immediately

-----

\(^{26}\) Dr. Engle graciously offers his tests for researchers at his laboratory website: [http://www.psychology.gatech.edu/rengelelab/](http://www.psychology.gatech.edu/rengelelab/). Users must request a password and must cite the source of the span.

\(^{27}\) In their automated version (Engle et. al, 2005), no researcher is needed so that the participant can self-monitor their own test. In their version, E-prime automatically calculates how much time a participant will need to calculate the math problems, as determined during the participant’s practice session, and adds on a SD of 2.5 seconds. Unfortunately, E-prime could not be obtained for the present study; however, administering the test individually to each participant (where their own pace is set as the researcher shows the next slide immediately after each participant calculates an arithmetic string) with PowerPoint is arguably just as robust a procedure.
following this, the participant was shown a letter, which he/she also said aloud. Upon hearing the letter, the researcher immediately presented the participant with the next mathematical string. This prevented participants from being able to rehearse the digit sequences, but allowed adequate time to process the arithmetic operation. After each set, the participant saw a set of question marks centered on the computer monitor (“???”). This cue prompted him or her to write down all of digits that followed each mathematical string in the order that he/she had seen them. Sets ranged from three to seven, with three items each per set, for a total of 75 items in the task. Examples of the OSPAN stimuli are presented in Figure 8 below:

**Figure 8. Example of OSPAN Stimuli**
6.6.7.2 The Reading Span Task (RSPAN)

As with the OSPAN, the RSPAN task used for this experiment was created by Dr. Engle and associates at his lab at Georgia Technical University. Also like the OSPAN, Dr. Engle’s version was made for the software E-Prime; the researcher transferred the stimuli so that it could be delivered individually and in person via PowerPoint. For this task, participants were individually shown sentences on the computer, half of which were semantically anomalous and half of which were not. As soon as the participant finished reading the sentence, the researcher showed them the next screen, which read: “? YES or NO,” prompting the participant to judge the grammaticality of each sentence, for which they must say “yes” or “no” (depending on whether it makes sense or not) aloud. After making an oral judgment, participants were immediately shown the next slide with a letter on it. The participant had to say this letter aloud. The participant was immediately shown the next sentence, and the sequence continued. After each set of sentences (ranging from two to five, with three trials of each size, for a total of 42 sentences), the participant was shown a screen with three question marks (“???”) which was their cue to write down the final words of each sentence they had just seen. Examples of RSPAN stimuli that were presented to participants is provided in Figure 9 below:
6.6.7.3 The Counting Span Task (CSPAN)

The CSPAN task used in this experiment was created by Dr. Engle and associates at Georgia Tech University (Engle, Tuholski, et al., 1999). Like the OSPAN and the RSPAN, the CSPAN task requires that participants remember a final count total for a subsequent recall test. The difference is the type of stimuli and dual processing. For this task, each participant was asked to count dark blue circles against a backdrop where visually/colorfully similar items were placed, specifically, light blue circles and dark blue squares. The participant had to count the dark blue circles aloud without pointing at the screen, and then repeat the final count number. The final count total must be said aloud immediately. (For example: “one, two, three, four, FOUR.”) After repeating the final count total aloud, the researcher immediately showed the
participant the next slide, with another series of circles and squares, for which the participant had to start counting again. No pausing was allowed, so as to prevent rehearsal of any kind. When participants saw the “???” slide, that served as their cue to write down the letter sequences they saw in that set in the same order that they had seen them. Examples of stimuli from the CSPAN test are presented in Figure 10 below:

**Figure 10. Example of CSPAN Stimuli**

---

6.6.8 Exit Questionnaire

An exit questionnaire was created (see Appendix L) to gauge participants’ perception of the experiment and also to ask if they had looked up anything pertinent to the study while it was on-going. This questionnaire asked what participants thought the goals of the study might have been; whether or not they learned anything and whether they noticed any particular grammatical item. Finally, the exit questionnaire sought to gain insight into the participants’ perception of
conversational interaction in the FTF versus CMC mode, asking them about their preferences, views as to the pros and cons of each mode, and their overall perception of the study.

6.6.9 Equipment

In addition to the above-mentioned materials (the bio-data questionnaire, the two treatment tasks, the multiple choice receptive tests, the FTF and CMC productive assessments, the anxiety and perceived difficulty questionnaire, the WMC spans, and the exit questionnaire), various technological materials were needed to carry out the present experiment. Two digital voice recorders were used to record all FTF interactions, both the treatments and assessments. A mini cassette tape recorder was also used as a backup, in the case that the digital recorders stopped working. Three Mac OS X laptops were also used for the study, as well as one mobile printer. Two of these laptops were used for all of the CMC interaction. The iChat software, version 4.0, was downloaded on both laptops in order to interact. The third laptop, belonging to the researcher, was set up at another station and was connected to the mobile printer. This was so that the tailor-made posttests could be immediately accessed and printed out for the two +complex groups. The iShowU screen recorder software was used to record all of the CMC interaction during the treatment sessions, and recorded both the participants’ and the researcher’s laptops. Finally, three simultaneous and secure systems were used to backup data during the experiment: (1) the password-protected Mobile Me website, an online ‘cloud’ source for

---

28 These laptops were generously loaned to the researcher from CNDLS at Georgetown University specifically so that this research project could be carried out.

29 During the Summer 2009 pilot study, five different screen recorder software were tested: Jing, Screencast-o-matic, Copernicus, SMART™Board, and iShowU. iShowU was the only one that allowed for a limitless amount of time to be recorded, and one that created automatic QuickTime files of computer activity. It far surpassed the other software options in terms of clarity, facility of use, file-saving options, and recording quality.
guarding and saving data; (2) an external hard drive; and (3) a password-protected driver set up on both experimental laptops so that data could also be saved there.

To sum, the technological equipment used in this experiment was:

1. Three MAC laptops: 2 set up with the iChat software and the iShowU screen recorder software for treatment sessions
2. one mobile printer
3. two digital audio recorders
4. one mini cassette recorder for backup
5. an online secure site for uploading data, as well as an internal and external drive for data backup

All of the above equipment were tested and tried during the summer 2009 pilot study. Once the experiment began, all data (to include consent forms, questionnaires, print-outs of the CMC interaction, and the mini cassettes from the cassette recorder) were placed in binders according to group. A sleeve pocket was created for each participant and placed in the binder; all binders were kept in a locked filing cabinet belonging to the researcher.

6.7 Procedure

This experiment was carried out over a period of two semesters. All of the 70 participants attended four sessions each over the course of four to five weeks. Participants were recruited via class visitations by the researcher. Before the study began, the researcher visited all sections of Intermediate II level Spanish, explaining the study and the time involved, and inviting students to participate. Those who participated were offered six extra credit points for completion of all four sessions. Students were told to email the researcher if they wanted to participate in the study so
that the first session could be arranged. Since all sessions were one-on-one, appointments had to be made for each participant and for each of the four sessions. This involved extensive preplanning with participants. Data collection took place during the weekdays, in the evenings and also on Saturdays and Sundays in order to accommodate participants’ schedules and the one-on-one nature of the treatment.

On the first day of the experiment, participants were given an informed consent form (Appendix M), and were then administered the three WMC tests and the three pretests (the FTF and CMC productive retells tasks, followed by the multiple-choice receptive test on the computer). At the end of the session, the participant and the researcher made an appointment for the second, third, and fourth sessions. The average time for the first session ranged from one hour and fifteen minutes to an hour and a half.

The treatment consisted of two sessions, Days 2 and 3 of the experimental schedule. The first day of the treatment took place a week or two weeks after the pre-test session of the experiment. The second treatment day took place one or two days (maximum) after the first treatment day. This was a critical component when scheduling appointments with each participant because, if due to unexpected circumstances, more than two days passed between the treatment days and/or more than a week between the immediate post and delayed posttest, they had to be eliminated. For the two treatment sessions, participants were first given the instructions and explained that they would be carrying out a story retell task with the researcher. Those participants in the -complex groups were told that they would be reading a story in parts in English, and that they had to retell the story as best as they could to the researcher in Spanish and in the past tense. The participants in the +complex groups were told the same thing, but were also informed that for some of the actions in the story, they would have to reflect upon the
intentional reasons behind those actions. (For the different instructions given to the simple and complex groups, please see Appendices N and O). All participants were told that if they needed assistance or wanted to know how to say a word in Spanish, to simply ask the researcher during their conversation with her.

All interactions in the FTF mode were carried out in a classroom or office with the participant and researcher facing each other at the same table. Both the participant and the researcher had the same set of cards: the L1 story sections and the accompanying comic strips. They started the treatment by reading the first L1 card. When the participant was ready, they moved on to the first comic strip, which the participant used to help him/her retell the story in Spanish. They then moved on to the next card, which was the second story blurb in the L1. The task was interactive and dialogic in nature, with the participant and the researcher working together to retell the story. The researcher would often start each comic strip by asking the participant questions (i.e., So then what happened?). Whenever the participant made an error with the past subjunctive, the researcher recasted that error, repeating their entire utterance (primary and secondary clause). Each FTF session was recorded with the digital and cassette recorders.

For the CMC interactions, participants were given the same instructions as the FTF group and were told that they would be carrying out the conversation with the researcher via iChat on a Mac laptop. The participant was given the laptop and was then instructed to log into iChat. After ensuring that the two laptops were connected on iChat, the researcher pressed the start record button for the iShowU screen recorder on both laptops. Unlike the FTF interactions, the CMC interactions—both treatments and assessments—were carried out in separate rooms. The researcher explained to the participant that that was so the true nature of computerized
interaction could be explored (as opposed to them sitting and typing in front of each other). When the participant indicated that he/she was ready, the researcher stepped out of the room and went to a nearby office/cubicle/hallway to begin the session. All participants in the CMC groups were told not to worry about typing accents in Spanish and to just focus on communicating the content of the story as best as they could. Like those in the FTF groups, the CMC participants were told to ask the researcher if they had any questions (i.e., vocabulary items) in chat just as they would in FTF conversation.

During all treatment sessions, the researcher took notes on the intentional reasons the participant had come up with for the +complex groups, and any during-task or post-task comments that were indicative of concurrent processing, meeting the task demands, and formation of the rule. This was done easily while communicating in the CMC mode; note-taking in the FTF groups was done while the participant was reading the L1 portion of the story so as to not detract from the conversation. The mean time for the first treatment session ranged from 21 to 58 minutes (depending on group assignment: FTF+C, FTF-C, CMC+C, CMC-C); the mean time for the second treatment session ranged from 16 to 48 minutes, also depending on group.

Immediately after each treatment session concluded, all participants were asked to write down how many minutes they believed it took them to carry out the task they had just completed with the researcher. After writing down their time judgment, participants were then told to fill out the anxiety and perceived difficulty questionnaire based on their perception of the task.

On the second treatment day (Day 3), upon completion of treatment 2, the time judgment, and the anxiety/perceived difficulty questionnaire, participants carried out the 3 immediate posttests. They did the two productive FTF and CMC retell tasks first, and then signed into Blackboard© on their laptop to take the multiple-choice receptive test via the computer. The
delayed posttests were done one week after Day 3 of the treatment. For this final session, participants once again did the two productive tasks in the FTF and CMC modes, and then took the multiple-choice receptive test on Blackboard©. After the delayed posttests, all participants filled out the exit questionnaire. As with the treatment sessions, the pretest and immediate and delayed posttests (three each per session) were all completed one-on-one with the researcher.⁴⁰

Recall that for the immediate and delayed posttests, both of the +complex groups had partially tailor-made assessments (2 out of 10 items on the productive tasks, and 5 out of 15 items on the receptive test). As the researcher took notes during the treatment on the intentional reasons the participants had come up with, the tailor-made posttests could be created immediately after the second treatment for the immediate posttests. To do this, the researcher had a separate computer station with a mobile printer set up that was in the next room or in a cubicle outside the researcher’s office. While the participant was filling out the Anxiety & Perceived Difficulty Questionnaire, the researcher took the notes she had taken during the two treatments and went to the laptop station, where the two word documents of the productive retell stories that participant was to take were open. (This required extensive prior planning so that the correct story for each participant was open and ready on the laptop). On the FTF and CMC story retell documents, all 10 critical items were highlighted in yellow. The researcher quickly inserted two of the participant’s own intentional reasons into the story, replacing two of the critical items with the participant’s items. The researcher then de-highlighted the entire document and printed each

---

⁴⁰Note that during the second semester of data collection, a second researcher assisted in collecting delayed posttest data for a total of five of the participants; this was due to some scheduling conflicts. Specifically, this researcher carried out the delayed FTF and CMC production assessment tasks with the five participants and then set them up on Blackboard© for the multiple-choice receptive tests. Before doing so, the researcher went through two training and observation sessions with the principle researcher. She was a female, non-native speaker of Spanish with five years of teaching experience. Her level of education was two Master’s Degrees in Spanish and in second language acquisition; she was a fourth-year Ph.D. candidate at the time.
out at the researcher station. The researcher then returned to where the participant was finishing up his/her questionnaire. The total time to create the tailor-made posttests took about 2 to 3 minutes; not one participant in the +complex groups had to wait to begin their retell assessment task.

The tailor-made receptive tests were created at this same computer station while the researcher was carrying out the CMC productive assessment task with the researcher. On the third laptop, the researcher opened up Blackboard© and created a new test for that particular participant, copying the test contents from one of the original A, B, or C versions into the test template. While tending to the CMC conversation on the experimental laptop, the researcher quickly scrolled down and changed five of the past subjunctive items, replacing the critical item primary clauses with intentional reasons that the participant had come up with him or herself. The researcher then assigned a new title to the test, gave it a password, ensured that the test features of randomization and no feedback were selected, and finalized the test. When the participant was ready to take the multiple-choice receptive test, they were given the new password so that he/she could access his/her own tailor-made tests. Therefore, a unique tailor-made posttest and delayed posttest was created for every participant in the +complex groups.

Because sessions with participants were often back-to-back, this tailor-made component of the design required extensive pre-planning on the researcher’s part. Each morning, a schedule of who was carrying out a session was made indicating the day’s participants, whether or not they required a tailor-made test (+/- complex status), and which CMC and FTF assessments they were assigned to do (i.e., version A, B, or C, according to the split-block schedule). The documents for each +complex participant were always opened, ready to go and with subjunctive prompts highlighted so that the researcher could tailor-make two of the items immediately,
remove the highlighting and then print out the document. Tailor-made items for the delayed posttest could obviously be made in advance. In all, approximately 700 hours were dedicated to the data collection sessions with participants for this experiment.

To control for outside exposure of the form, the Director of the Intermediate II program approved removing a reference to the past subjunctive from the curriculum throughout both semesters. Also, all instructors of Intermediate II were asked to not cover the past subjunctive inside or out of class during the period of the experiment.

The average times for the pre-, immediate, and delayed posttests, as well as for the treatments, are provided below in Tables 17 and 18 respectively:

**Table 17. Mean Times for the Productive Assessment Tasks**

<table>
<thead>
<tr>
<th></th>
<th>Pre-</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FTF</strong></td>
<td>14.9 minutes (4.2)</td>
<td>12.2 minutes (2.8)</td>
<td>12.1 minutes (2.8)</td>
</tr>
<tr>
<td><strong>CMC</strong></td>
<td>18.7 minutes (5.2)</td>
<td>17.6 minutes (5.7)</td>
<td>22.2 minutes (5.1)</td>
</tr>
</tbody>
</table>

Mean (SD)

**Table 18. Mean Times for the Treatment Sessions by Group**

<table>
<thead>
<tr>
<th></th>
<th>Treatment Session 1</th>
<th>Treatment Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FTF+C</strong></td>
<td>25.2 minutes (4.9)</td>
<td>21.8 minutes (5.7)</td>
</tr>
<tr>
<td><strong>FTF-C</strong></td>
<td>21.2 minutes (7.0)</td>
<td>16.5 minutes (5.5)</td>
</tr>
<tr>
<td><strong>CMC-C</strong></td>
<td>43.5 minutes (8.8)</td>
<td>33.0 minutes (7.7)</td>
</tr>
<tr>
<td><strong>CMC+C</strong></td>
<td>58.0 minutes (13.7)</td>
<td>48.7 minutes (14.0)</td>
</tr>
</tbody>
</table>

Mean (SD)
The experimental procedure is provided below in Figure 11:

**Figure 11. Experimental Procedure**

| Day 1 | Consent form, background and bio-data questionnaire  
|       | Working memory spans (OSPAN, RSPAN, CSPAN)  
|       | Pretests: FTF productive, CMC productive, multiple-choice receptive |
| (one to two weeks interval) |
| Day 2 | Treatment 1: FTF+C, FTF-C, CMC-C, CMC+C  
|       | Time judgment of treatment time  
|       | Anxiety / Perceived Difficulty Questionnaire |
| Day 3 | Treatment 2: FTF+C, FTF-C, CMC-C, CMC+C  
|       | Time judgment of treatment time  
|       | Anxiety / Perceived Difficulty Questionnaire |
|       | Immediate posttests: FTF productive, CMC productive, multiple-choice receptive  
|       | (20 % tailor-made productive tests, 35% tailor-made receptive test for +complex groups) |
|       | (one week) |
| Day 4 | Delayed posttests: FTF productive, CMC productive, multiple-choice receptive  
|       | (20 % tailor-made productive tests, 35% tailor-made receptive test for +complex groups)  
|       | Exit Questionnaire |
6.7.1 Transcription and Preparation of the Data

All of the CMC chat dialogues, both treatment and assessments, were printed out by the researcher for coding purposes; all of the FTF oral tasks were transcribed by the researcher for the production of the past subjunctive and also for uptake during the treatments. Once the FTF assessment tasks had been transcribed, coding and scoring of the tests could consequently be carried out.

6.8 Coding and Scoring Procedures

6.8.1 Bio-Data

Following the answers from the learners’ background questionnaires, participants’ bio-data was quantitatively coded for so that group comparability analyses could be run. Variables coded for were: sex (1 for female, 2 for male), and complexity (+/- intentional reasoning), modality (1 for FTF, 2 for CMC), group (1 for FTF+C, 2 for FTF-C, 3 for CMC-C, 4 for CMC+C). Also, age, total hours spent chatting online per week, and other L1 was tallied and entered into SPSS 18.0.

6.8.2 FTF and CMC Production Tasks

For both the FTF and CMC production assessment tasks, participants’ production of the past subjunctive was scored in two ways. An ‘absolute’ scoring was done first, assigning 1 point to past subjunctive forms and 0 to indicative past tense forms or the present subjunctive. Deviations to the past subjunctive, be they in writing, pronunciation, or in production (i.e., starting with the first person present indicative form instead of the third person plural preterit form to create the past subjunctive, which was the rule base to start with for creating the present...
subjunctive) were accepted. Examples of deviate forms accepted were: *tengaba(n)* (produced by 4 different learners), *fueron sean, estudieron, fuyera, hacera, haciera, quedaran, seran, intentiera, herira, permitra, supara, quersera,* and *diciera.* Interestingly, it was often observed that the same participant could deviate in their production of the past subjunctive, mixing up the endings of *-ar* verbs (*-ara*) and *-er/-ir* verbs (*-iera*). For example, one learner produced both *tengara* and *teniera* in the same story retell. Another learner produced *diciera* in the FTF retell, but later on the CMC retell productive assessment, wrote the correct form *dijera.* Once absolute scores had been compiled, a second, more sensitive way of coding for the past subjunctive was done. For the ‘sensitive’ scoring, 1 point was assigned to the past subjunctive form (deviations included), .5 point was assigned to the *present* subjunctive form (produced where the past subjunctive should have been) and 0 points was given to the indicative past forms. There were two reasons for doing this. First, the goal of this study was to see if, through conversational interaction, learners could acquire mood morphology marking in the past tense as necessitated by intentional reasons, epistemological, and attitude-based verbs in the primary clause position. Therefore, marking a verb with the subjunctive morphology while telling a story in the present could be indicative that the learner recognized that the verb needed subjunctive mood marking, but just did not know how to do this in past tense. Theoretically, the goal was that the learner recognized that intentional reasons, verbs denoting volition, etc. in the primary clause mandated the use of the subjunctive in the dependent clause, so if the learner used the present subjunctive, it might be the case that they had recognized this rule. The decision to allocate points to present subjunctive forms was further supported by the following comment by a learner in the FTF+C group once he finished the posttest:
It should be past subjunctive, I'm just really rusty so that's why like ... fuera? I'll be like 'sea'. Because I know the present subjunctive.

His unexpected comment seemed to support a sensitive procedure of coding for forms. Also, another learner in the CMC-C group, who had started producing the past subjunctive form during the treatment, once reverted back to the present subjunctive:

\[
\text{Estaba molesto de que Luis no tomara el invitacion seriomiento} \\
(\text{He was angry that Luis did not take-SUBJ, PAST the invitation seriously})
\]

12 lines later: \(\text{El dudaba que ellos pueda hacer esta} \)  
(He doubted that they can-SUBJ, PRESENT do this)

researcher recast \(\rightarrow \text{el dudaba que ellos pudieran hacer esto}\)  
(He doubted that they can-SUBJ, PAST do this)

\[Uptake \rightarrow \text{Si}\]  
(Yes)

This regression to marking the verb with the subjunctive morphology in the present tense seemed to be indicative of backsliding or U-shaped behavior that is documented in the literature as the nonlinear process of L2 development.\(^{31}\) Later, the participant went back to using the past subjunctive form. This U-shaped behavior was observed on more than one occasion.

The second reason for allowing points to be assigned to the present subjunctive was the fact that some modern-day dialects in Spanish (i.e., Ecuadorian Spanish) use the present subjunctive where the past subjunctive should be in the dependent clause position.\(^{32}\)

In coding for production according to both absolute and sensitive coding schemes as described above, the goal was the ability to perform statistical analyses for both score systems, allowing the researcher to get a better overall idea of how to best capture L2 development.

\(^{31}\) See Kellerman (1985), Sato (1990), Selinker (1972) and Selinker & Lakshamanan (1992) for more discussions on the non-linear process that comprises SLA.

\(^{32}\) The researcher thanks Dr. Héctor Campos for pointing this out.
6.8.2.1 Interrator Agreement

After the researcher scored all versions of the FTF and CMC productive tasks (both absolute and sensitive coding), a second researcher also coded 20% of the data. To test the reliability of the scoring, the two rators’ coding for all of the data was compared. Interrator reliability was high for all of the assessments: FTF production, absolute coding, version A: 97%; version B: 98%; version C: 99%, and for the sensitive coding: version A: 98%; version B: 96%; and version C: 97%. For CMC production, the interrator reliability was also high: CMC production, absolute coding, version A: 98%; version B: 98%; version C: 97%, and for the sensitive coding: version A: 96%; version B: 97%; and lastly, version C: 97%.

6.8.3 Multiple-Choice Receptive Tests

Each multiple-choice receptive test (A, B, and C) for each learner was printed out and coded for correctly answered past subjunctive forms. 1 point was given to correct forms, while 0 points was given to incorrectly answered items.

6.8.3.1 Interrator Agreement

As with the production assessments, a second researcher coded 20% of the multiple-choice receptive data. Interrator reliability was at 100% for all versions of the tests.

6.8.4 Working Memory Spans

In accordance with the recommendations of Conway et al. (2005), the performance on all WMC span tasks (OSPAN, CSPAN, RSPAN) was first checked to ensure that participants achieved at least 85% accuracy in processing component of the task (correct math calculations of the OSPAN, correct grammatical judgments of the RSPAN, and correct counting during the CSPAN). This was critical to ensure that participants were actively engaged with the processing
component of each task. If a learner achieved less than 85% accuracy on one of the tasks, he or she was eliminated from the study.

The coding procedures for the spans also followed the recommendations of Conway et al. (2005), where a partial-credit unit scoring (PCU) procedure was used to code and grade the spans as opposed to an absolute scoring procedure. To explain this coding procedure, recall that for each span, there were two sources of data present given the dual-task nature: the processing component and the storage component. In the past, many researchers have used absolute scoring, starting a participant with an item of 2 and continuing the task until he/she reached a maximum threshold—the total amount of perfectly recalled items (say, 4) was thus considered their WM span value, following Daneman and Carpenter (1980). However, as Conway et al. (2005) pointed out, one issue with this is that information from other trials is discarded for that participant. Furthermore, different factors might determine the assumed ‘difficulty’ of a WM span task; this could be “threatening [to] span reliability across tasks” (Conway et al., 2005, p. 774). Instead of using absolute methods, the researchers recommend assigning 1 point to all accurately recalled items and 0 points to all other responses, irrespective of their position in the recall. Then, the data can be aggregated for the four different coding procedures, which they explain:
1) **Partial-credit unit scoring** (PCU): the “mean proportion of elements within an item that were recalled correctly”

2) **Partial-credit load scoring** (PCL): “the sum of correctly recalled elements from all items, regardless of whether the items are perfectly recalled or not” (i.e. serial position)

3) **All-or-nothing unit scoring** (ANU): “the proportion of items for which all the elements were recalled correctly”

4) **All-or-nothing load scoring** (ANL): “the sum of the correctly recalled elements from only the items in which all the elements are recalled in correct serial order.”

“Partial-credit” refers to correctly recalled elements within an item regardless of their serial position. “All-or-nothing” refers to correctly recalled elements only in the correct serial position of their original presentation. The “unit” component of scoring refers to the counting of all elements equally (an averaged proportion, i.e. one correctly recalled element out of a two item span counts as 1/2 or .50). Finally, the “load” method means more weight is assigned to lengthier items, so the researcher must calculate the mean of all correctly recalled elements. To provide an example of how these four score methods differ, compare the following span scores for one participant on the CSPAN, taken from the pilot study (see Table 19 below):
Table 19. Scoring Procedure Options for WM Span Tasks (Conway et al., 2005)

Data from Participant 5, Pilot Study, CSPAN task

<table>
<thead>
<tr>
<th>Element</th>
<th>Score</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6,4,8</td>
<td>4,6,8</td>
</tr>
<tr>
<td>2</td>
<td>5,7,6,3,9,</td>
<td>5,7,6,9,</td>
</tr>
<tr>
<td>3</td>
<td>7,5,6</td>
<td>5,7,6</td>
</tr>
<tr>
<td>4</td>
<td>9,4,3,6,8,5</td>
<td>6,4,9,8,5,</td>
</tr>
<tr>
<td>5</td>
<td>4,9</td>
<td>4,9</td>
</tr>
<tr>
<td>6</td>
<td>7,5,3,8,4</td>
<td>5,7,8,4,</td>
</tr>
<tr>
<td>7</td>
<td>7,4,6,9</td>
<td>7,6,4,9</td>
</tr>
<tr>
<td>8</td>
<td>8,5</td>
<td>8,5</td>
</tr>
<tr>
<td>9</td>
<td>7,3,9,6,8,5</td>
<td>7,6,8,5,</td>
</tr>
<tr>
<td>10</td>
<td>4,9,3,7</td>
<td>,8,7,3</td>
</tr>
<tr>
<td>11</td>
<td>5,6,9</td>
<td>5,6,9</td>
</tr>
<tr>
<td>12</td>
<td>3,6,4,8,9</td>
<td>6,8,9,</td>
</tr>
<tr>
<td>13</td>
<td>5,3,6,7,8,4</td>
<td>,6,8,4,</td>
</tr>
<tr>
<td>14</td>
<td>9,6</td>
<td>9,6</td>
</tr>
<tr>
<td>15</td>
<td>7,8,3,5</td>
<td>7,8,3,5</td>
</tr>
</tbody>
</table>

RSPAN SCORE: .83 .53 .8 .37

For the present dissertation, partial-credit unit scores for all participants were calculated as Conway et al. (2005) suggested that it may be the most advantageous. Therefore, for all WMC span answers, one point was assigned to each correctly recalled element regardless of its serial position.

6.8.4.1 Interrator Agreement

To ensure that the spans were scored correctly, another researcher scored all of the working memory spans for every single participant. The two researchers compared their scores and found a difference for one of the spans for one participant. One of the researchers went back
and recalculated that span’s score and saw she had made an error; the scores were subsequently compared again and 100% agreement on participants’ WMC scores was achieved.

6.9 Statistical Analyses

First, descriptive statistics were calculated for participants’ bio-data according to group, as well as for each group’s performance on the pretest, to ensure comparability. Next, descriptive statistics were calculated for each group’s immediate post, and delayed posttests performance. The means and standard deviations were computed in order to examine the central tendency, variability, and distribution of group-based scores for each assessment. Next, to ensure comparability between participants’ pretest scores in the four Experimental Groups (+Complex/CMC, -Complex/CMC, +Complex/FTF, -Complex/FTF) a one-way ANOVA (where Group was the independent variable and pre-test scores were the dependent variable) was run for each assessment test (MC receptive, FTF and CMC productive), with post-hoc Scheffé to determine statistically significant differences. Following this analysis, two outliers were eliminated so that the groups were comparable with no statistically significant difference.

6.9.1 RQs 1 - 4

To address the first four research questions, the combined effects of recasts and task complexity, modality, and their interaction on L2 development (i.e., the extent of gains made from pre- to posttest performance with the past subjunctive form) a repeat-measures ANOVA was conducted, with Time (pretest, posttest, and delayed posttest scores) as the within-subjects factor, Task Complexity and Modality served as the between-subjects factors, and test scores were the dependent variable. This analysis was performed in order to show the effects of both task complexity and modality and their intercept on learners’ retention of the past subjunctive
form, and was conducted for each assessment test. In addition, to get a clearer picture of learning gains made in each of the different groups, a one-way ANOVA was performed for each group (i.e., (+Complex/CMC, -Complex/CMC, +Complex/FTF, -Complex/FTF and Control), with Group as the independent variable and posttest performances on each assessment test as the dependent variables. The mean gains, standard deviations, and difference comparisons for each was consequently computed and reported. Plots showing the learning gains for complexity, modality, and then for each of the four groups, were drawn.

The fourth research question examined whether tailor-made items on the assessments would be answered better than non-tailor-made items for the +complex groups. To do this, a tailor-made score and a non-tailor-made score was computed for each immediate and delayed post assessment (FTF productive, CMC productive, and multiple-choice receptive) for the FTF+C and CMC+C groups. Paired t-tests were then calculated, comparing the mean differences of tailor-made items with non-tailor made items for each assessment.

6.9.2 RQ 5: On Uptake

To address the fifth research question, uptake as a mediating variable on learning, a simple regression analysis was calculated to explore the relationship of uptake and learning on the posttests. Output was specified by group to examine the predicting effects of uptake on L2 development in the four experimental conditions (FTF+C, FTF-C, CMC-C, CMC+C).

6.9.3 RQ 6: On WMC

In order to explore the moderating variable of working memory capacity (measured with the OSPAN, RSPAN, and CSPAN), internal reliability (coefficient alphas) was first assessed. Then, participants were divided into high and low WMC groups. Correlations were computed for WMC as a moderator of gains made on the immediate and delayed posttests. This was so that it
could be determined whether higher WMC was better for learning to take place, and also to test the hypothesis that low WMC capacity learners do better in CMC (Payne & Whitney, 2002).

6.9.4 RQ7: On Independent Measures of Cognitive Complexity

Finally, for the sensitive measures of cognitive complexity, various statistical performances were calculated. To address the effect of task complexity on time judgments, a time difference score was calculated for each participant, where the real treatment time was subtracted from participants’ guessed time. This was performed for both treatment sessions. A one-way ANOVA was calculated with Task Complexity as the independent variable and time difference as the dependent variable. A follow up analysis was done with Group as the independent variable as well to get a clearer picture of the results.

For the Anxiety and Difficulty Perception Questionnaires, an overall anxiety/difficulty score was computed for each participant and for each treatment session (the total sum of Likert-scale items answered on both questionnaires). Then, one-way ANOVAs were run with Task Complexity as the independent variable and anxiety score as the dependent variable for both treatment sessions. As with the time judgments, a one-way ANOVA was calculated for the effect of Group (FTF+C, FTF-C, CMC-C, CMC+C) on anxiety score to get a more in depth picture of how task complexity and modality affect learners’ task perceptions.

Where necessary, descriptive statistics were collected and reported. All analyses were performed with the Statistical Package for the Social Sciences (SPSS) 18. An alpha level of $p < .05$ was set for all tests.
CHAPTER 7: RESULTS

This chapter presents the results of all quantitative analyses performed on the mean scores obtained on the three assessments (FTF production, CMC production, and multiple-choice reception) in order to address the effects of the independent variables of Task Complexity and Modality on L2 development. Descriptive statistics are provided and box plots are drawn to show the variance occurring in the different levels of complexity. How learners in the complex groups performed on the tailor-made items is also provided. To follow is a presentation of the data obtained for uptake as a mediating variable and WMC as a moderating variable of learning. Finally, data obtained from the analyses of the individual measures of task complexity (anxiety and perceived difficulty questionnaires and time judgments) are presented. Descriptive statistics and reliability analyses are provided for each computed analysis where necessary. The chapter concludes with a presentation of post hoc analyses that were performed: (1) number of recasts received by each group per treatment, (2) total amount of uptake produced by each group per treatment, (3) total production of the past subjunctive form per group during the treatment sessions and (4) measures of noticing and awareness of a rule. These decisions were coded for and analyzed as potential variables that further assisted in explicating the findings, as is explained in detail below.

7.1 RQs 1-4: The Effects of Task Complexity and Modality on L2 Development

This first three research questions were:

1. What are the combined effects of task complexity and recasts on L2 development?
2. What are the combined effects of modality (FTF vs. CMC) and recasts on L2 development?
3. Is there an interaction effect between task complexity and modality, when recasts are held constant?

To address these questions, findings for FTF production are presented first, followed by findings for CMC production, and then findings for multiple-choice receptive development. For each of these assessments, the research questions were addressed by performing a repeated-measures ANOVA with Task Complexity and Modality entered as the between-subject factor, and Time (pre, immediate, and delayed-posttests) as the within-subject factor. To show the effect size of the independent variables and their interaction, partial eta-squared ($\eta^2$) values were also obtained. In order to explicate the findings in terms of differences in learning for each group, the same repeated-measures ANOVA was calculated with Group (FTF+C, FTF-C, CMC-C, and CMC+C) entered as the between-subject factor and Time (pre, immediate, and delayed-posttests) as the within-subject factor.

### 7.1.1 FTF Production Task

#### 7.1.1.1 Descriptive Statistics

The mean scores and standard deviations for the two independent variables (Task Complexity and Modality) by Time are reported below in Tables 20 and 21 below and are graphically displayed respectively in Figures 12 and 13:

**Table 20. FTF Production: Mean Scores and Standard Deviations for Task Complexity**

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Complex</td>
<td>.00 (.00)</td>
<td>3.19 (3.41)</td>
<td>2.71 (3.37)</td>
</tr>
<tr>
<td>+Complex</td>
<td>.11 (6.8)</td>
<td>2.20 (3.21)</td>
<td>1.91 (3.07)</td>
</tr>
<tr>
<td>$p =$</td>
<td>.32</td>
<td>.22</td>
<td>.30</td>
</tr>
</tbody>
</table>
Table 21. FTF Production: Mean Scores and Standard Deviations for Modality

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF</td>
<td>.11 (.67)</td>
<td>2.43 (3.20)</td>
<td>2.31 (3.35)</td>
</tr>
<tr>
<td>CMC</td>
<td>.00 (.00)</td>
<td>3.00 (3.48)</td>
<td>2.32 (3.14)</td>
</tr>
<tr>
<td>p</td>
<td>.34</td>
<td>.50</td>
<td>.98</td>
</tr>
</tbody>
</table>

Figure 12. FTF Production: Mean Scores for Task Complexity x Time
7.1.1.2 Inferential Statistics

A repeated-measures ANOVA was calculated and revealed no significant main effect for Task Complexity \( (F(1, 66) = 1.81, p = .183; \text{partial } \eta^2 = 0.03) \), no significant main effect for Modality \( (F(1, 66) = .11, p = .74 \text{ and partial } \eta^2 = 0.01) \), and a significant interaction between Task Complexity and Modality \( (F(1, 66) = 13.2, p = .001; \text{partial } \eta^2 = 0.17) \). A significant main effect was also found for Time \( (F(2, 132) = 33.47, p > .001; \text{partial } \eta^2 = 0.34) \), and finally, a significant triple interaction between Task Complexity, Modality, and Time \( (F(2, 132) = 7.28, p = .001 \text{ and } \eta^2 = 0.10) \).

The ANOVAs for the FTF production task are reported in Table 22 and 23 below:
Table 22. FTF Production: Between-Subject Effects

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>199.82</td>
<td>1</td>
<td>199.82</td>
<td>56.65</td>
<td>.000</td>
<td>1.0</td>
</tr>
<tr>
<td>Complexity</td>
<td>6.37</td>
<td>1</td>
<td>6.37</td>
<td>1.81</td>
<td>.183</td>
<td>.263</td>
</tr>
<tr>
<td>Modality</td>
<td>.388</td>
<td>1</td>
<td>.388</td>
<td>.110</td>
<td>.741</td>
<td>.062</td>
</tr>
<tr>
<td>Complex. * Modal.</td>
<td>46.55</td>
<td>1</td>
<td>46.55</td>
<td>13.20</td>
<td>.167</td>
<td>.947</td>
</tr>
<tr>
<td>Error</td>
<td>232.80</td>
<td>66</td>
<td>3.53</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 23. FTF Production: Within-Subject Effects

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>285.56</td>
<td>2</td>
<td>142.78</td>
<td>33.47</td>
<td>.000</td>
<td>1.0</td>
</tr>
<tr>
<td>Time * Complexity</td>
<td>13.69</td>
<td>2</td>
<td>6.83</td>
<td>1.60</td>
<td>.205</td>
<td>.334</td>
</tr>
<tr>
<td>Time * Modality</td>
<td>4.16</td>
<td>2</td>
<td>2.08</td>
<td>.487</td>
<td>.615</td>
<td>.128</td>
</tr>
<tr>
<td>Time * Complexity * Modality</td>
<td>62.10</td>
<td>2</td>
<td>31.05</td>
<td>7.28</td>
<td>.001</td>
<td>.932</td>
</tr>
<tr>
<td>Error</td>
<td>563.13</td>
<td>132</td>
<td>4.27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Given the significant triple interaction between Task Complexity, Modality, and Time, the data of this study are addressed in light of this triple interaction (i.e., per group).

First, a one-way ANOVA was calculated to ensure group comparability on the FTF oral production pretest. The results revealed no significant difference among groups (FTF+C, FTF-C, CMC-C, and CMC+C; \( F(3, 66) = .96, p = .42 \)), ensuring that any difference found on the posttest could be attributed to the treatment.
7.1.1.3 Descriptive Statistics Per Group

Next, descriptive statistics were calculated. The descriptive statistics for the four experimental groups over Time are reported in Table 24 below and are visually displayed in Figure 14:

Table 24. FTF Production: Mean Scores and Standard Deviations per Group

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>18</td>
<td>.22 (.94)</td>
<td>3.00 (3.70)</td>
<td>3.17 (3.72)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>18</td>
<td>.00 (.00)</td>
<td>1.86 (2.58)</td>
<td>1.44 (2.78)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>17</td>
<td>.00 (.00)</td>
<td>4.59 (3.69)</td>
<td>4.06 (3.50)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>17</td>
<td>.00 (.00)</td>
<td>1.35 (2.42)</td>
<td>0.58 (1.31)</td>
</tr>
</tbody>
</table>

* p = .25, .02*, .00*

Figure 14. FTF Production: Mean scores per Group
A close examination of Figure 14 reveals that the CMC-C group and, to a lesser extent, the FTF+C group performed better when compared to the CMC+C and FTF-C on both the immediate and delayed posttests.

7.1.1.4 Inferential Statistics Per Group

Next, a one-way ANOVA was performed on the mean scores of the immediate posttest as well as the delayed posttest with Group (i.e., FTF+C, FTF-C, CMC-C, and CMC+C) as the between-subjects factor. The analysis revealed a significant main effect for Group at both immediate posttest \( (p = .02) \) and at the delayed post \( (p = .00) \), demonstrated below in the ANOVA tables 25 and 26:

**Table 25. FTF Production: Group Differences at Immediate Posttest**

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>105.74</td>
<td>3</td>
<td>35.25</td>
<td>3.54</td>
<td>.019</td>
<td>.761</td>
</tr>
<tr>
<td>Intercept</td>
<td>510.10</td>
<td>1</td>
<td>510.10</td>
<td>51.21</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>105.74</td>
<td>3</td>
<td>35.25</td>
<td>3.54</td>
<td>.019</td>
<td>.761</td>
</tr>
<tr>
<td>Error</td>
<td>657.40</td>
<td>66</td>
<td>9.96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 26. FTF Production: Group Differences at Delayed Posttest**

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>129.08</td>
<td>3</td>
<td>43.03</td>
<td>4.82</td>
<td>.004</td>
<td>.887</td>
</tr>
<tr>
<td>Intercept</td>
<td>374.69</td>
<td>1</td>
<td>374.70</td>
<td>41.99</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>129.08</td>
<td>3</td>
<td>43.03</td>
<td>4.82</td>
<td>.004</td>
<td>.887</td>
</tr>
<tr>
<td>Error</td>
<td>589.00</td>
<td>66</td>
<td>8.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A post hoc Scheffé was subsequently performed to identify exactly where the significant
differences existed among the groups. At the immediate posttest time, the Scheffé revealed that
the CMC-C group did significantly better than the CMC+C group \( (p = .04) \). At the delayed
posttest, this significant difference between the CMC-C and CMC+C held \( (p = .01) \). While the
two FTF groups also had a mean difference of 1.12 points at the immediate post and 1.72 at the
delayed posttest, their differences were not significant.

In sum, for FTF oral productive development, the combination of -complex and CMC
mode (CMC-) led to the highest gains. The FTF+C group did the second best, followed by the
FTF-C. The CMC+C group did the worst, averaging only a gain score of approximately 1 point
on the FTF productive post assessments. Note that while the CMC-C, CMC+C, and FTF-C
demonstrated a decrease in performance on the delayed posttest, the FTF+C group was the only
group to continue increasing in performance at Time 3.

7.1.1.5 +/- Complex: Variance in Performance

One observation when coding data for participants’ FTF oral production was that in both
of the complex groups (i.e., FTF+C and CMC+C) there was great variance in performances. In
other words, participants in these groups seemed to either get all answers correct or not, and the
researcher questioned whether ANOVAs, where means are compared, failed to showed this. To
explore this potential, box plots were drawn in order to examine the variance of scores between
the two complex and simple groups. Below, in Figure 15, are two box plots showing the variety
achieved in gains on the immediate oral posttest:
The data from this box plot demonstrate that there were major outliers for both of the complex groups. On the contrary, for both of the simple groups, all participants stayed within the range. For the complex groups, the standard deviation was greater than three, and some participants went outside of this range. In other words, Task Complexity resulted in greater variance for gains made in the FTF immediate posttest.

---

33 As can be seen, both box plots have a black line in the middle of them, representing the median. The space below the median line represents the lower quartile, while the space above represents the upper quartile. Both box plots have one ‘whisker’ line drawn from the top. These lines are drawn up to the adjacent top value from the upper quartile, a value that mathematically is still not an outlier. The circles above the whisker lines represent the outliers, values that exist outside the threshold of the upper quartile. These box plots therefore visually demonstrate that there were major outliers for the +complex groups.
7.1.2 CMC Production Task

7.1.2.1 Descriptive Statistics

The mean scores and standard deviations for the two independent variables (Task Complexity and Modality) by Time are provided below in Tables 27 and 28, and are then graphically displayed in Figures 16 and 17 respectively:

Table 27. CMC Production: Mean Scores and Standard Deviations for Task Complexity

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Complex</td>
<td>.06 (.24)</td>
<td>3.20 (3.63)</td>
<td>3.04 (3.61)</td>
</tr>
<tr>
<td>+Complex</td>
<td>.14 (.69)</td>
<td>2.31 (3.30)</td>
<td>2.40 (3.76)</td>
</tr>
<tr>
<td><em>p</em></td>
<td>.49</td>
<td>.29</td>
<td>.47</td>
</tr>
</tbody>
</table>

Table 28. FTF Production: Mean Scores and Standard Deviations for Modality

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF</td>
<td>.17 (.70)</td>
<td>2.24 (3.12)</td>
<td>2.50 (3.79)</td>
</tr>
<tr>
<td>CMC</td>
<td>.03 (.17)</td>
<td>3.31 (3.78)</td>
<td>2.96 (3.59)</td>
</tr>
<tr>
<td><em>p</em></td>
<td>.27</td>
<td>.20</td>
<td>.61</td>
</tr>
</tbody>
</table>
Figure 16. CMC Production: Mean Scores for Task Complexity x Time
7.1.2.2 Inferential Statistics

A repeated-measures ANOVA was performed with Task Complexity and Modality entered as the between-subject factors, and Time (CMC pretest, immediate post, and delayed posttest) as the within-subject factor. As with the FTF production assessments, there was no significant main effect for Task Complexity ($F(1, 66) = 1.10, p = .30$; partial $\eta^2 = 0.02$), no significant main effect for Modality ($F(1, 66) = .84, p = .36$; partial $\eta^2 = 0.1$); however the interaction between Task Complexity and Modality was significant ($F(1, 66) = 12.13, p < .001$ and partial $\eta^2 = 0.16$). A significant main effect was found for Time ($F(2, 132) = 35.35, p < .001$; partial $\eta^2 = 0.35$), as well as a significant interaction between Task Complexity, Modality and Time ($F(2, 132) = 7.97, p < .001$; partial $\eta^2 = 0.11$).

The ANOVAs for the CMC production test are presented below in Table 29 and 30:
Table 29. CMC Production: Between-Subject Effects

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>243.58</td>
<td>1</td>
<td>243.58</td>
<td>54.35</td>
<td>.000</td>
<td>1.0</td>
</tr>
<tr>
<td>Complexity</td>
<td>4.94</td>
<td>1</td>
<td>4.94</td>
<td>1.10</td>
<td>.298</td>
<td>.179</td>
</tr>
<tr>
<td>Modality</td>
<td>3.76</td>
<td>1</td>
<td>3.76</td>
<td>.84</td>
<td>.363</td>
<td>.147</td>
</tr>
<tr>
<td>Complx. * Modal.</td>
<td>54.34</td>
<td>1</td>
<td>54.34</td>
<td>12.13</td>
<td>.001</td>
<td>.929</td>
</tr>
<tr>
<td>Error</td>
<td>295.78</td>
<td>66</td>
<td>4.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 30. CMC Production: Within-Subject Effects

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>328.06</td>
<td>2</td>
<td>164.03</td>
<td>35.35</td>
<td>.000</td>
<td>1.0</td>
</tr>
<tr>
<td>Time * Complexity</td>
<td>10.29</td>
<td>2</td>
<td>5.14</td>
<td>1.11</td>
<td>.333</td>
<td>.241</td>
</tr>
<tr>
<td>Time * Modality</td>
<td>12.80</td>
<td>2</td>
<td>6.40</td>
<td>1.38</td>
<td>.255</td>
<td>.292</td>
</tr>
<tr>
<td>Time * Complexity * Modality</td>
<td>73.96</td>
<td>2</td>
<td>36.98</td>
<td>7.97</td>
<td>.001</td>
<td>.952</td>
</tr>
<tr>
<td>Error</td>
<td>612.51</td>
<td>132</td>
<td>4.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In light of the significant triple interaction between Task Complexity, Modality, and Time, the data are once again addressed per group to demonstrate this triple interaction. First though, a one-way ANOVA was calculated to ensure group comparability for the CMC production pretests. There was no significant difference among groups, $F(3, 66) = 1.01, p = .39$,.
ensuring that any gains on the posttests were due to the treatment. Once group comparability had been established, descriptive statistics could be calculated.

### 7.1.2.3 Descriptive Statistics Per Group

The descriptive statistics for the four experimental groups over Time are reported in Table 31 below and are visually displayed in Figure 18:

**Table 31. CMC Production: Mean Scores and Standard Deviations per Group**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>18</td>
<td>.28 (.96)</td>
<td>2.86 (3.24)</td>
<td>3.61 (4.31)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>18</td>
<td>.06 (.24)</td>
<td>1.61 (2.95)</td>
<td>1.39 (2.90)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>17</td>
<td>.06 (.24)</td>
<td>4.88 (3.59)</td>
<td>4.79 (3.53)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>17</td>
<td>.00 (.00)</td>
<td>1.74 (3.35)</td>
<td>1.12 (2.63)</td>
</tr>
<tr>
<td>p</td>
<td></td>
<td>.39</td>
<td>.02*</td>
<td>.01*</td>
</tr>
</tbody>
</table>

**Figure 18. CMC Production: Mean scores per Group**
A close examination of Figure C reveals that the CMC-C group performed much higher than the other three groups at the immediate and delayed posttests. The FTF+C group performed second best, and then the CMC+C and FTF-C, respectively, came in last. Also, while the CMC+C and FTF-C groups demonstrate a decrease in performance on the delayed posttest, only the FTF+C group demonstrated a linear increase in performance by Time 3. The CMC-C group’s scores appear to stay constant by the delayed posttest, decreasing only minimally.

7.1.2.4 Inferential Statistics Per Group

A one-way ANOVA was then performed on the mean scores of the immediate posttest and the delayed posttest with Group as the between-subjects factor. The analysis revealed a significant main effect for group at both immediate posttest ($p = .003$) and at the delayed post ($p = .005$), which is demonstrated in the ANOVA tables 32 and 33 below:

**Table 32. CMC Production: Group Differences at Immediate Posttest**

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>118.37</td>
<td>3</td>
<td>39.46</td>
<td>3.65</td>
<td>.017</td>
<td>.775</td>
</tr>
<tr>
<td>Intercept</td>
<td>537.62</td>
<td>1</td>
<td>537.62</td>
<td>49.73</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>118.37</td>
<td>3</td>
<td>39.46</td>
<td>3.65</td>
<td>.017</td>
<td>.775</td>
</tr>
<tr>
<td>Error</td>
<td>713.50</td>
<td>66</td>
<td>10.81</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 33. CMC Production: Group Differences at Delayed Posttest

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>162.97</td>
<td>3</td>
<td>54.32</td>
<td>4.66</td>
<td>.005</td>
<td>.876</td>
</tr>
<tr>
<td>Intercept</td>
<td>520.50</td>
<td>1</td>
<td>520.50</td>
<td>44.68</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>162.97</td>
<td>3</td>
<td>54.32</td>
<td>4.66</td>
<td>.005</td>
<td>.876</td>
</tr>
<tr>
<td>Error</td>
<td>768.85</td>
<td>66</td>
<td>11.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A post hoc Scheffé revealed that at the immediate posttest time, the CMC-C group did significantly better than the FTF-C group, averaging +3.27 points more ($p = .04$). The difference between the CMC-C and CMC+C group approached significance ($p = .06$), with a mean difference of +3.15 more for the CMC-C group. The two FTF groups also had a mean difference of 1.25 points (with the FTF+C scoring higher); however, these differences were not significant.

Next, the Scheffé showed that at the delayed posttest, the significant difference between the CMC-C and FTF-C groups held ($p = .04$) and that this time, the CMC-C group also did significantly better than CMC+C ($p = .03$). The FTF+C averaged +2.22 points higher than the FTF-C at the delayed posttest, but as with the immediate post, this difference was not significant ($p = .29$).

In sum, for CMC productive development, the combination of -complex and CMC mode (CMC-) led to the highest gains, just as it did for oral productive development. The FTF+C group did the second best, followed by the CMC+C group and FTF-C group. The FTF+C group averaged a gain of .75 points at the delayed posttest; all groups experienced a natural decrease.
7.1.2.5 +/- Complex: Variance in Performance

As with the FTF production assessment task, it was observed that much variance existed in the CMC production scores of the two complex groups (FTF+C and CMC+C). To explore this further, box plots were once again drawn in order to examine the variance of scores between the +complex and -complex conditions. The box plots are provided below in Figure 19 to illustrate this variance at the immediate CMC posttest:

**Figure 19. CMC Production: Box Plot Showing Performance for +/- Complex Groups**

The box plots demonstrate that all participants in the simple groups stayed within the range of the median at the posttest level; there were no outliers. Contrastingly, the two complex conditions did result in outliers (shown via the circles above the top value adjacent to the upper quartile, which is represented by the whisker line). As with the gains made on the FTF production data, the data from this box plot demonstrate that there were major outliers for the complex groups on the CMC production assessments. Once again, Task Complexity resulted in
greater variance for gains made in the CMC immediate posttest, regardless of the mode in which they had carried out the treatment.

7.1.3 Multiple-Choice Receptive Test

7.1.3.1 Descriptive Statistics

The mean scores and standard deviations for the two independent variables (Task Complexity and then Modality) by Time are reported below in Tables 34 and 35 below, and are then graphically displayed in Figures 20 and 21:

**Table 34. M-C Reception: Mean Scores and Standard Deviations for Task Complexity**

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Complex</td>
<td>1.71 (1.84)</td>
<td>7.20 (6.18)</td>
<td>7.49 (6.24)</td>
</tr>
<tr>
<td>+Complex</td>
<td>2.11 (2.51)</td>
<td>7.40 (5.93)</td>
<td>7.14 (6.32)</td>
</tr>
<tr>
<td>p =</td>
<td>.45</td>
<td>.89</td>
<td>.82</td>
</tr>
</tbody>
</table>

**Table 35. M-C Reception: Mean Scores and Standard Deviations for Modality**

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF</td>
<td>1.78 (2.00)</td>
<td>7.50 (6.21)</td>
<td>6.94 (6.48)</td>
</tr>
<tr>
<td>CMC</td>
<td>2.06 (2.40)</td>
<td>7.09 (5.89)</td>
<td>7.71 (6.04)</td>
</tr>
<tr>
<td>p =</td>
<td>.60</td>
<td>.78</td>
<td>.61</td>
</tr>
</tbody>
</table>
Figure 20. M-C Reception: Mean Scores for Task Complexity x Time

Estimated Marginal Means of MEASURE_1

- COMPLEXITY
  - SIMPLE
  - COMPLEX

TIME
7.1.3.2 Inferential Statistics

Next, a repeated-measures ANOVA was calculated with Task Complexity and Modality set as the between-subject factor and Time (i.e., pre-, immediate post, and delayed post multiple-choice receptive test scores) set as the within-subject factor. No significant main effect was found for Task Complexity \( (F(1, 66) = .013, p = .99; \text{partial } \eta^2 = 0.00) \), nor was there a significant main effect for Modality \( (F(1, 66) = .008, p = .93; \text{partial } \eta^2 = 0.00) \). A significant interaction effect was found however for Task Complexity and Modality \( (F(1, 66) = 15.27, p < .001; \text{partial } \eta^2 = 0.19) \). There was also a significant main effect for Time \( (F(2, 132) = 45.16, p < .001; \text{partial } \eta^2 = 0.41) \), and a significant triple interaction between Task Complexity, Modality and Time \( (F(2, 132) = 5.33, p = .006; \text{partial } \eta^2 = 0.75) \). The ANOVAs for the receptive multiple-choice test are reported in Table 36 and 37 below:
As with the two productive tests, the significant triple interaction between Task Complexity, Modality, and Time required a demonstration of the data for the receptive tests necessary.

Before doing so, a one-way ANOVA was performed that examined the multiple-choice receptive pretest scores for all groups. There was no significant difference between any of the groups ($F(3, 66) = .463, p = .71$), ensuring that any gains made were attributable to the treatment.
7.1.3.3 Descriptive Statistics Per Group

The descriptive statistics for the four experimental groups over Time are reported in Table 38 below and are visually displayed in Figure 22:

Table 38. M-C Reception: Mean Scores and Standard Deviations per Group

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Immediate Post</th>
<th>Delayed Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>18</td>
<td>2.17 (1.69)</td>
<td>9.78 (5.46)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>18</td>
<td>1.39 (2.25)</td>
<td>5.22 (6.21)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>17</td>
<td>2.06 (1.25)</td>
<td>9.29 (5.59)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>17</td>
<td>2.06 (3.21)</td>
<td>4.88 (5.48)</td>
</tr>
<tr>
<td>p</td>
<td>.71</td>
<td>.02*</td>
<td>.003*</td>
</tr>
</tbody>
</table>

Figure 22. M-C Reception: Mean scores per Group

An close look at Figure 22 shows that both FTF+C and CMC-C performed better than the FTF-C and CMC+C groups, with the FTF+C group doing the best, followed closely by the
CMC-C group. It also shows that these two sets of groups (FTF+C and CMC-C, and then FTF-C and CMC+C) performed quite similarly. After the CMC-C group, there is a nearly 5-point decrease, at which the FTF-C group appears and is followed closely by the CMC+C group. The two FTF groups experienced a natural decline by the delayed posttests. The CMC+C group’s scores stayed constant by Time 3. The CMC-C group is the only one to increase in performance by the delayed post.

7.1.3.4 Inferential Statistics Per Group

A one-way ANOVA was then performed on the mean scores of the immediate posttest as well as the delayed posttest with Group entered as the between-subjects factor. The analysis revealed a significant main effect for group at both immediate posttest ($p = .02$) and at the delayed post ($p = .003$), which are demonstrated below in the ANOVA tables 39 and 40, respectively:

**Table 39. M-C Reception: Group Differences at Immediate Posttest**

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>355.18</td>
<td>3</td>
<td>118.40</td>
<td>3.65</td>
<td>.017</td>
<td>.775</td>
</tr>
<tr>
<td>Intercept</td>
<td>3721.25</td>
<td>1</td>
<td>3721.25</td>
<td>114.69</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>355.18</td>
<td>3</td>
<td>118.40</td>
<td>3.65</td>
<td>.017</td>
<td>.775</td>
</tr>
<tr>
<td>Error</td>
<td>2141.52</td>
<td>66</td>
<td>32.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 40. M-C Reception: Group Differences at Delayed Posttest

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Sig.</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>498.05</td>
<td>3</td>
<td>166.01</td>
<td>5.02</td>
<td>.003</td>
<td>.900</td>
</tr>
<tr>
<td>Intercept</td>
<td>3753.00</td>
<td>1</td>
<td>3753.00</td>
<td>113.36</td>
<td>.000</td>
<td>1.00</td>
</tr>
<tr>
<td>Group</td>
<td>498.05</td>
<td>3</td>
<td>166.02</td>
<td>5.02</td>
<td>.003</td>
<td>.900</td>
</tr>
<tr>
<td>Error</td>
<td>2185.00</td>
<td>66</td>
<td>33.11</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In order to see where the significant difference(s) existed among the groups, a post hoc Scheffé was performed. At the immediate posttest time, the Scheffé revealed that the FTF+C and CMC-C both did significantly better than the FTF-C and CMC+C. The FTF+C and CMC-C averaged 9.78 point and 9.29 points more (respectively), compared to the FTF-C and CMC+C average means of 5.22 and 4.88, respectively. There was no significant difference between the FTF+C or CMC-C group however, nor was there a significant difference between the FTF-C and CMC+C groups. The post hoc Scheffé showed significant differences between groups at the delayed posttest, as well. This time, only the CMC-C performed significantly better than the FTF-C and the CMC+C groups. At the delayed posttest, FTF+C averaged +4.78 points more than the FTF-C and +4.51 points more than the CMC+C groups, though these differences were not significant.

In sum, for receptive L2 development, the combination of +complex and FTF mode (FTF+C) led to the highest gains, as did the combination of -complex and CMC mode (CMC-C). There was no significant difference between these two groups at the immediate post or delayed posttest. By the delayed posttest, significant differences were found for the CMC-C over the FTF-C and CMC+C groups.
7.1.3.5 +/- Complex: Variance in Performance

The fact that the two complex groups (FTF+C versus CMC+C) performed so discordantly is probably what led to the lack of variance showed in scores. As can be seen in the box plots below, the +Complex task did not result in outliers for gains made in L2 receptive development of the past subjunctive. Variance was comparable in the -Complex and +Complex groups, with median values at approximately 6 and whisker lines indicative of both upper and lower adjacent values to the top and bottom quartiles:

Figure 23. M-C Reception: Box Plot Showing Performance for +/- Complex Groups

7.1.4 +Complex Performance on Tailor-Made Items

The fourth research question was:

4. For learners in the +complex groups (both in FTF and CMC): Are tailor-made items answered better than non-tailor-made items on the assessments?

As explained in the methodology chapter, 2 of 10 items were customized on the productive assessments, and 5 of 15 items on the receptive assessment for the two complex groups (FTF+C
and CMC+C). In order investigate if these groups answered tailor-made items significantly better than the ‘critical’ items (i.e., those intentional reasons they had come up with themselves during the treatment vs. predetermined items), two variables were computed: TailorMade items and NonTailorMade items. All items on each of the learning assessments were coded according to (1) tailor-made status and (2) whether or not it was produced (FTF and CMC production task) correctly or answered (multiple-choice reception task) correctly. This then allowed for paired t-tests to be run in order to investigate any significant differences in mean number of items answered correctly. Each paired t-test analysis is reported below according to assessment.

7.1.4.1 Tailor-Made Items in FTF Production

A paired t-test was performed to compare answers to tailor-made items versus non-tailor-made items on the FTF productive immediate post. The results revealed a significant difference in the scores of tailor-made (Mean = .6, SD = .85) and non-tailor-made (Mean = .18, SD = .32) items; $t(34) = 3.79, p = .001$. A similar significant difference in scores on the FTF delayed posttest was also revealed between the delayed tailor-made items (Mean = .43, SD = .78) and non-tailor-made items (Mean = .18, SD = .32); $t(34) = 2.62, p = .01$. In other words, items that were customized led to production of the past subjunctive on the FTF production task significantly better than items that were not customized.

7.1.4.2 Tailor-Made Items in CMC Production

Next, a paired t-test was performed to compare answers on tailor-made items versus non-tailor-made items on the CMC immediate post and delayed posttests. Like the FTF assessments, a significant difference was found in the scores of tailor-made (Mean = .49, SD = .74) versus non-tailor-made (Mean = .21, SD = .33) items; $t(34) = 3.38, p = .002$. This significant difference held through the delayed posttest in the scores of delayed tailor-made items (Mean = .43, SD =
and delayed non-tailor-made items (Mean = .23, SD = .37); t(34) = 2.39, p = .02. As with FTF production, items that were customized led to production of the past subjunctive in the CMC mode significantly better than items that were not customized.

### 7.1.4.3 Tailor-Made Items in Receptive Development

Lastly, a paired *t*-test was calculated in order to compare answers on tailor-made versus non-tailor-made items on the multiple-choice receptive posttest and delayed post. At the immediate post receptive test, no significant difference was found between the scores of tailor-made (Mean = .51, SD = .43) versus non-tailor-made (Mean = .48, SD = .41); *t*(34) = .875; *p* = .39. The opposite was found at the delayed receptive posttest however. On the delayed test, there was a significant difference between tailor-made items (Mean = .53, SD = .43) versus non-tailor-made items (Mean = .42, SD = .44); *t*(34) = 2.70, *p* = .01.

In sum, these results suggest that customizing items on post assessments for the +complex groups may have had an effect on their answering that item correctly. Specifically, when the prompts in test items were tailor-made, (i.e., were those prompts that the participants had come up with themselves during the treatment), participants produced the past subjunctive significantly more in the FTF and CMC modes, both at the immediate and delayed posttests. A significant effect for tailor-made items was found on the receptive test only at the delayed posttest.

To conclude this section on the effects of Task Complexity and Modality on L2 development of the past subjunctive (RQs 1-4), it was found that the combination of -Complex and CMC mode led to the best FTF and CMC productive development. This group was followed

---

34 By prompts it is meant the first clause of intentional reasoning that the participant came up with him or herself, which mandates the use of the past subjunctive in the dependent clause. Examples are: quería que (he/she wanted [that]), estaba triste de que (he/she was sad [that]), estaban muy contentos de que (they were happy [that]), etc.
closely by the FTF+C group on both productive measures. In addition, box plots showed that a great deal of variation in performance took place in the +complex groups irrespective of mode. For L2 receptive development, the best performances were achieved by both the FTF+C and CMC-C groups. Together, the CMC+C and FTF-C groups did significantly worse than FTF+C and CMC-C, which had no significant difference between them. Unlike the production assessments, no variance in performance was observed in the two complex groups (FTF+C and CMC+C) for receptive development as shown by the box plots. Finally, customizing items on assessment tasks may have an effect on their being answered correctly. The +complex groups answered customized items (i.e., those intentional reasons they came up with themselves) significantly better than the critical items. This stayed true for both the FTF and CMC production tasks at the immediate and delayed posttest; this finding only held at the delayed posttest for receptive development however.

7.2 RQ 5: Uptake as a Mediator of L2 Development

The fifth research question was:

5. Does uptake mediate L2 development? In other words, if a treatment effect is observed, is it because of the production of uptake during interaction? If so, is the effect different under different experimental conditions?

To answer this question, simple regression analyses were performed to assess whether the production of uptake mediated learning (in other words, whether learning was due to uptake). First, the amount of uptake produced per participant was tallied. Given that there were two treatment sessions with ten obligatory contexts for the past subjunctive each, there was a possibility of receiving a total of 20 recasts and subsequently producing uptake a total of 20
times. Following Révész et al. (in press), both repetitions of the entire recasted utterance or just part of the utterance (i.e., just the form) were considered uptake; simple acknowledgment did not count as uptake. Once a total uptake amount had been tallied for each participant, their rate of uptake was regressed onto scores at the immediate post and delayed posttest for all three assessments. Below, each of the regression analyses is reported with SPSS analysis output split by groups (i.e., each of the four different experimental conditions).

7.2.1 Uptake as a Mediator for FTF Production Development

Two linear regression analyses were performed with participants’ rate of uptake regressed on their scores for the FTF production assessment tasks. Output was designated by groups. The first regression analysis was computed with scores on the immediate posttest, and the second with scores from the delayed posttest. Regressions for uptake per group and per assessment are provided in Table 41 below:

Table 41. Uptake as a Mediator for L2 Development in FTF Production

<table>
<thead>
<tr>
<th>Group</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>-.57**</td>
<td>-.64**</td>
</tr>
<tr>
<td>FTF-C</td>
<td>.05</td>
<td>.03</td>
</tr>
<tr>
<td>CMC-C</td>
<td>-1.34</td>
<td>-1.54</td>
</tr>
<tr>
<td>CMC+C</td>
<td>.65</td>
<td>-.10</td>
</tr>
</tbody>
</table>

** p < .001

As can be seen above, uptake was found to be a highly significant mediator for the FTF+C group only, and negatively so. In other words, the more uptake produced by the FTF+C group led to significantly less development. Uptake was not found to be a significant mediator for development of FTF production for any of the other groups.
7.2.2 Uptake as a Mediator for CMC Production Development

To examine the mediating effect of uptake for development on the CMC production task, two linear regression analyses were performed with participants’ rate of uptake regressed on their scores for the CMC production assessment tasks. The first was computed with scores on the immediate posttest, and the second with scores from the delayed posttest, with output set according to group. Table 42 below shows the regressions results for uptake per group and per assessment:

**Table 42. Uptake as a Mediator for L2 Development in CMC Production**

<table>
<thead>
<tr>
<th>Group</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>-.64**</td>
<td>-.89**</td>
</tr>
<tr>
<td>FTF-C</td>
<td>-.02</td>
<td>-.03</td>
</tr>
<tr>
<td>CMC-C</td>
<td>-.98</td>
<td>-1.43</td>
</tr>
<tr>
<td>CMC+C</td>
<td>1.20</td>
<td>.66</td>
</tr>
</tbody>
</table>

** p < .001

Once again, uptake was found to be a highly negative significant mediator for the FTF+C group only. This indicated that the more uptake produced by the FTF+C group resulted in significantly less development. Just as with development on the FTF production tasks, uptake was not a significant mediator for development of CMC production for any of the other groups.

7.2.3 Uptake as a Mediator for Receptive Development

Lastly, to examine the mediating effect of uptake for development on the multiple-choice reception tests, two linear regression analyses were performed with participants’ rate of uptake regressed on their scores (with the first analysis performed with scores on the immediate posttest,
and the second with scores from the delayed posttest). Table 43 below shows the regressions results for uptake per group and per assessment:

**Table 43. Uptake as a Mediator for L2 Receptive Development**

<table>
<thead>
<tr>
<th>Group</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>-.82**</td>
<td>-1.03**</td>
</tr>
<tr>
<td>FTF-C</td>
<td>-.08</td>
<td>-.04</td>
</tr>
<tr>
<td>CMC-C</td>
<td>-1.52</td>
<td>-1.31</td>
</tr>
<tr>
<td>CMC+C</td>
<td>4.99**</td>
<td>2.50</td>
</tr>
</tbody>
</table>

** p < .001

Just as with the production assessments, uptake was found to be a highly negative significant mediator for the FTF+C group, indicating that the more uptake produced by the FTF+C group resulted in significantly less development. However, this time, uptake was found to be highly significant predictor of development in the CMC+C group at the immediate posttest. In other words, the more uptake produced in the CMC+C group, the more L2 development occurred on the multiple-choice receptive immediate posttest. Uptake was not a significant mediator of learning at the delayed posttest for CMC+C however. Finally, uptake was not a mediator for L2 development for the FTF-C or CMC-C groups.

**7.3 RQ 6: WMC as a Moderator of L2 Development**

The sixth research question asked:

6. Does working memory capacity moderate L2 development? In other words, if a treatment effect is observed, is the effect different for learners of high or low WMC? If so, is the effect different under different experimental conditions?
In order to address this research question, the correlation coefficient was calculated for all three WM spans, the OSPAN, RSPAN and CSPAN. All of the spans were significantly correlated with each other at the \( p < .01 \) level: CPSPAN with OSPAN, \( r(70) = .59, p < .00 \), CPSAN with RSPAN, \( r(70) = .60, p < .00 \); OSPAN with RSPAN, \( r(70) = .77, p < .00 \). This confirmed that the three spans were measuring the same construct of WMC and that further analyses could subsequently be performed. Next, a composite WMC score for each participant was calculated via the sum of each span. Descriptive statistics were calculated to establish the mean WM score, which was 2.08. A new variable was subsequently calculated, WMC_L_H, which established the low WMC learners as those in the range of 0 to the mean (2.08) and high WMC as those in the range of the mean to the highest score. After the low and high WMC learners had been identified, correlations between WMC scores and immediate and delayed posttests, splitting the output by low vs. high WMC learners. Results are provided below for each assessment: FTF production, CMC production, and the multiple-choice receptive test.

### 7.3.1 WMC as a Moderator for FTF Production Development

To address whether WMC was a moderating variable for development on the FTF production tasks, a bivariate correlation was performed between WMC and scores on the FTF production assessments with data output split according to low and high level WMC learners and also by Group (i.e., FTF+C, FTF-C, CMC-C, CMC+C). Pearson’s correlation coefficients are reported below in Table 44:
Table 44. WMC as a Moderator of L2 Development in FTF Production

<table>
<thead>
<tr>
<th>Group</th>
<th>WM Group</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>Low WMC</td>
<td>.275</td>
<td>.524</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>.571</td>
<td>-.222</td>
</tr>
<tr>
<td>FTF-C</td>
<td>Low WMC</td>
<td>.557</td>
<td>.440</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>.295</td>
<td>.287</td>
</tr>
<tr>
<td>CMC-C</td>
<td>Low WMC</td>
<td>-.129</td>
<td>-.237</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>-.596</td>
<td>-.731</td>
</tr>
<tr>
<td>CMC+C</td>
<td>Low WMC</td>
<td>.571</td>
<td>-.222</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>-.501</td>
<td>-.675</td>
</tr>
</tbody>
</table>

As Table 44 shows, WMC was not significant correlated with performance on either the immediate post or delayed posttest, in any of the four experimental groups. This held true for both low and high WMC learners. In the CMC groups, WMC was for the most part negatively correlated with development, which seemed to indicate that higher WMC did not lead to more development in this mode. WMC was not a moderator for development on the FTF production assessments.

### 7.3.2 WMC as a Moderator for CMC Production Development

Next, to examine whether WMC differentially moderated L2 development on the CMC production tasks in the four experimental groups, a bivariate correlation was once again performed with data output split according to WMC level and according to Group. Table 45 below provides Pearson’s correlation coefficients:
Table 45. WMC as a Moderator of L2 Development in CMC Production

<table>
<thead>
<tr>
<th>Group</th>
<th>WM Group</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>Low WMC</td>
<td>.224</td>
<td>.261</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>-.082</td>
<td>-.398</td>
</tr>
<tr>
<td>FTF-C</td>
<td>Low WMC</td>
<td>.659</td>
<td>.564</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>.385</td>
<td>.340</td>
</tr>
<tr>
<td>CMC-C</td>
<td>Low WMC</td>
<td>-.183</td>
<td>-.087</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>-.419</td>
<td>-.526</td>
</tr>
<tr>
<td>CMC+C</td>
<td>Low WMC</td>
<td>.001</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>-.428</td>
<td>-.514</td>
</tr>
</tbody>
</table>

As with FTF production development, WMC was not significantly correlated with performance on either the immediate post or delayed CMC posttests, in any of the four experimental groups, for low or for high WMC learners. For the high WMC learners, WMC was negatively correlated with development in the FTF+C, CMC-C, and CMC+C groups, indicating that high WMC did not result in more development. WMC was negatively correlated with development for both low and high WMC learners in the CMC-C group, which seemed to indicate that WMC played no role in development for this group. Overall, WMC was not a moderator of CMC production development for low or for high WMC learners.

7.3.4 WMC as a Moderator for Receptive Development

As a final step, a correlation was performed to examine the relationship between WMC and development demonstrated on the multiple-choice receptive tests. Once again, the analysis
output was divided by low versus high-level WMC learners and per group. Pearson’s Correlation Coefficients are provided in Table 46:

### Table 46. WMC as a Moderator of L2 Receptive Development

<table>
<thead>
<tr>
<th>Group</th>
<th>WM Group</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>Low WMC</td>
<td>.162</td>
<td>.196</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>.095</td>
<td>-.171</td>
</tr>
<tr>
<td>FTF-C</td>
<td>Low WMC</td>
<td>.090</td>
<td>-.171</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>.695*</td>
<td>.464</td>
</tr>
<tr>
<td>CMC-C</td>
<td>Low WMC</td>
<td>.307</td>
<td>-.015</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>-.933**</td>
<td>-.720</td>
</tr>
<tr>
<td>CMC+C</td>
<td>Low WMC</td>
<td>-.296</td>
<td>-.497</td>
</tr>
<tr>
<td></td>
<td>High WMC</td>
<td>.116</td>
<td>-.276</td>
</tr>
</tbody>
</table>

*p = .027  
**p = .002

Unlike the production tasks, WMC was found to significantly moderate receptive development for the high WMC learners in the FTF-C; this was at the immediate posttest. This meant that higher WMC assisted in receptive development in the FTF-C group. Contrarily, WMC was found to be highly significantly correlated, negatively so, with development at the immediate posttest for the CMC-C group. In other words, high WMC did not lead to more receptive development in this group. Technically speaking, the significant negative inverse relationship indicated that high WMC would lead to worse developmental outcomes in CMC-C. Finally, WMC was not significantly correlated with performance on the immediate post or delayed posttest for low or for high WMC learners at the immediate nor delayed posttest—it therefore did not moderate learning for high or low WMC learners in the other groups.
In sum, WMC did not moderate learning on the FTF or CMC production tasks. On the multiple-choice receptive test, WMC did moderate learning for FTF-C high WMC learners at the immediate posttest. The opposite was true for high WMC learners in the CMC-C group: WMC was negatively and significantly correlated with development at the immediate post.

7.4 Independent Measures of Cognitive Complexity

The seventh research question was:

7. How does the combination of task complexity and modality (i.e., FTF+C, FTF-C, CMC+C, CMC-C) affect learners’ reported independent measures of cognitive complexity?

In order to probe deeper into the nature of increases in cognitive complexity during task-based interaction, two quantitative measures were collected: (1) time judgments of the duration of the treatment compared to the real treatment times and (2) measures of anxiety and perceived difficulty as determined by a questionnaire. This section reports on the statistics and results for each of these variables.

7.4.1 Time Judgments

Recall that immediately after both treatment sessions, all participants were asked to guess the time duration they believed had passed for them to be able to complete the task. Therefore, for both treatment sessions 1 and 2, two times were recorded for each participant: Guess Time and Real Time. Based on these numbers, a Time Difference score was calculated by subtracting the real time from the guess time. The descriptive statistics for the time judgments (i.e., the difference from the real times) are provided below in Table 47:
Table 47. Descriptive Statistics for Time Judgments

<table>
<thead>
<tr>
<th>Group</th>
<th>T1 Real Time</th>
<th>T1 Guess Time</th>
<th>T2 Real Time</th>
<th>T2 Guess Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>25.17 (4.9)</td>
<td>28.06 (9.3)</td>
<td>21.83 (5.73)</td>
<td>24.5 (10.45)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>21.17 (6.96)</td>
<td>17.17 (5.77)</td>
<td>16.50 (5.47)</td>
<td>16.28 (6.87)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>43.47 (8.84)</td>
<td>36.32 (11.04)</td>
<td>33.00 (7.66)</td>
<td>27.06 (7.91)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>58.00 (13.62)</td>
<td>63.82 (15.76)</td>
<td>48.65 (14.04)</td>
<td>49.00 (12.18)</td>
</tr>
</tbody>
</table>

Mean Times (SD)

Next, a one-way ANOVA was calculated with Group (FTF+C, FTF-C, CMC-C, CMC+C) entered as the independent variable and Time Difference entered as the dependent variable. A significant main effect was found for Group ($F(3, 66) = 12.41, p < .001$). A Post hoc Scheffé revealed that significant differences in time judgments for Treatment 1 existed between the FTF+C and the FTF-C group, with the FTF+C judging an average of 6.89 minutes more for their treatment than the real time ($p = .04$). A significant difference also existed between the FTF+C group and the CMC-C group; the FTF+C group on averaged guessed 10.04 minutes more than the CMC-C group ($p = .001$). This significant difference in time judgment was also seen for the CMC+C group in relation to CMC-C and the FTF-C groups. The CMC+C group on average judged their Time 1 as 9.82 minutes more than the FTF-C group ($p = .002$) and 12.98 minutes more than the CMC-C group ($p < .001$).

To explore the differences of time judgments at Treatment 2, a one-way ANOVA was then computed with Group (FTF+C, FTF-C, CMC-C, and CMC+C) entered as the independent variable and Time Difference (for Treatment 2) as the dependent variable. A significant main effect was found for Group at Treatment 2 ($F(3, 66) = 4.94, p = .004$). A post Hoc Scheffé
revealed that, this time, significant differences in time judgments only existed between the FTF+C and the CMC-C group. On average, the FTF+C judged Treatment 2 as taking 8.61 minutes more than the judgment CMC-C group ($p = .005$). Differences in time judgments between the other groups (i.e., CMC+C always guessing more time, and FTF-C always guessing less), but they were not significantly different at Treatment 2. Table 48 below provides the mean time differences judged by all four groups at Treatment 1 and 2:

**Table 48. Mean Judgment Time Differences per Group**

<table>
<thead>
<tr>
<th>Group</th>
<th>T1 Time Difference</th>
<th>T2 Time Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>+2.89* (8.53)</td>
<td>+2.67* (9.25)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>-4.0* (5.37)</td>
<td>-0.22 (4.05)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>-7.15* (7.4)</td>
<td>-5.94* (5.92)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>+5.8* (6.55)</td>
<td>+0.35 (7.40)</td>
</tr>
</tbody>
</table>

Mean Difference in Times (SD)

In sum, experimental condition significantly affected how much time learners guessed took them to carry out the task. The time judgment differences were most notable for Task Complexity, not Modality. The FTF+C and CMC+C always guessed more time than the real time for both treatments. Contrarily, the FTF-C and CMC-C always guessed less time. By Treatment 2, the differences between Guess Time and Real Time lowered. Thus, Task Complexity appeared to have affected how much time learners’ judged the treatments to last, with the more complex task resulting in higher time judgments than the less complex task.
7.4.2 Anxiety and Perceived Difficulty Questionnaire

In order to explore the effects of experimental condition (i.e., FTF+C, FTF-C, CMC-C, CMC+C) on reported levels of anxiety and difficulty perception, a total anxiety and perceived difficulty variable was computed for each participant for each treatment session (in other words, for questionnaire 1 and questionnaire 2). This was done via the sum of each of the 15 items on the questionnaire, calculated for Treatment 1 and for 2. The descriptive statistics of reported anxiety and perceived difficulty level per group and at each treatment are provided in Table 49 below. Note that anxiety and perceived difficulty levels are reported in descending order, with the group with the highest levels at the top, and the group with the lowest levels at the bottom.

Table 49. Reported Anxiety and Perceived Difficulty Level by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Anxiety T1</th>
<th>Anxiety T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>49.39 (9.38)</td>
<td>48.25 (9.02)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>45.28 (7.36)</td>
<td>43.28 (11.08)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>41.24 (8.60)</td>
<td>40.71 (10.27)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>39.53 (9.64)</td>
<td>37.09 (8.60)</td>
</tr>
</tbody>
</table>

A quick glance at Table 49 shows that all four groups decreased somewhat by Treatment 2 in reported anxiety and perceived difficulty level. Notably, the two FTF groups reported higher anxiety and perceived difficulty than the CMC groups. FTF+C always had the highest scores, followed by FTF-C, then CMC+C, and lastly, CMC-C.
To explore this statistically, a one-way ANOVA was performed with Group as the independent variable and Anxiety T1 as the dependent variable. A significant main effect was found for Group ($F(3, 66) = 4.42, p = .007$). A post hoc Scheffé revealed that the significant difference was between FTF+C and CMC-C. The FTF+C group, on average, had 9.86 points higher of anxiety and perceived difficulty than the CMC-C group ($p = .016$). In fact, the FTF+C group had the highest level of anxiety and perceived difficulty after Treatment 1 but the difference in means was only significant when compared to CMC-C.

Next, a one-way ANOVA was calculated with Group as the independent variable and Anxiety T2 as the dependent variable. As with Treatment 1, a significant main effect was found for Group on Anxiety and perceived difficulty ($F(3, 66) = 4.02, p = .011$). A post hoc Scheffé revealed that once again, a significant difference in anxiety/perceived difficulty levels occurred between the FTF+C and CMC-C groups, with the FTF+C averaging 11.16 points more in reported anxiety/difficulty than the CMC-C group ($p = .015$).

In sum, experimental condition significantly affected learners’ reported anxiety and perceived difficulty of a task. The two FTF groups had the highest levels of anxiety and perceived difficulty, followed by the two CMC groups. Unlike the time judgments, modality is what primarily affected anxiety and perceived difficulty levels, and this was combined with complexity. That is, within modes, the +complex groups always had higher anxiety/perceived difficulty than the -complex groups, but significance was only found between the FTF+C and CMC-C groups. All participants, irrespective of group, reported decreases in anxiety and perceived difficulty by the second treatment day.
7.5 Summary of the Results

To briefly summarize the findings so far, each research question and corresponding result is provided below.

Research Questions 1-4: On learning

1. What are the combined effects of task complexity and recasts on L2 development?
   It was found that the combination of the -complex condition and recasts led to the most development on the FTF productive assessment as well as the CMC productive assessment. The +complex condition led to the most development for the receptive test. However, no significant differences between +complex and -complex conditions were found.

2. What are the combined effects of modality (FTF vs. CMC) and recasts on L2 development?
   The combination of the CMC mode and recasts led to the most development on the FTF productive assessment as well as the CMC productive assessment. Contrarily, the FTF mode led to the most development for the receptive test. However, no significant differences between FTF mode and CMC mode were found.

3. Is there an interaction effect between task complexity and modality when recasts are held constant?
   Yes, and the triple interaction effect between Task Complexity, Modality, and Time proved to be the most revealing, rendering a presentation of the results by Group (FTF+C, FTF-C, CMC-C, and CMC+C) necessary. The CMC-C group outperformed the other three groups on both the FTF and CMC productive assessments (with significant differences found between CMC-C and CMC+C for FTF production; significant
differences between CMC-C and FTF-C at the posttest for CMC production, and significant differences between CMC-C and FTF-C, as well as CMC+C, at the delayed post CMC production test). The FTF+C group was a close second, and was the only group that continued to increase in development at the two productive delayed posttests. The FTF+C group performed the highest on the receptive test, with significantly higher scores than both FTF-C and CMC+C. The CMC-C group did second best, also significant higher than FTF-C and CMC+C. There were no significant differences between FTF+C and CMC-C on any of the assessments.

4. For learners in the +complex groups (both in FTF and CMC): Are tailor-made items answered more accurately than non-tailor-made items on the assessments?

Tailor-made items were answered significantly better than non-tailor-made items on both of the production tasks at both times 2 and 3. Tailor-made items were answered significantly better on the multiple-choice receptive test at the delayed test only. This indicates that customizing posttests for the complex groups may have an effect on their being answered correctly.

Hypothesis 1 on Learning: Based on the predictions of the Cognition Hypothesis as well as empirical findings from the literature, it was predicted that more complex tasks plus recasts would lead to the most L2 development.

The first hypothesis was confirmed, but for the FTF mode only. That is, the combination of more cognitively complex tasks and recasts leads to greater L2 development in the FTF mode, but does not in the CMC mode.
Research Question 5: On uptake

5. Does uptake mediate L2 development? In other words, if a treatment effect is observed, is it because of the production of uptake during interaction? If so, is the effect different under different experimental conditions?

Uptake did significantly mediate development for only one group and on only one assessment: the CMC+C group on the immediate receptive test. Therefore, the CMC+C group’s production of uptake significantly predicted receptive development. Uptake was significantly and negatively correlated with learning for the FTF+C group, however, on all three assessments. Uptake did not mediate learning for the FTF-C and CMC-C groups. The treatment effect was therefore not because of the production of uptake during the interaction, except for the CMC+C group on the receptive test.

Hypothesis 2 on uptake: Based on the predictions of the Cognition Hypothesis, it was predicted that uptake would mediate L2 development, especially in the more complex tasks.

The second hypothesis was partially confirmed, for the CMC+C group only. Uptake did not mediate learning for the other groups.

Research Question 6: On working memory capacity

6. Does working memory capacity modulate L2 development? In other words, if a treatment effect is observed, is the effect different for learners of high or low WMC? If so, is the effect different under different experimental conditions?

WMC did moderate learning for one group and on one assessment only: the high WMC learners in the FTF-C group and on the receptive immediate posttest. WMC was significantly and negatively correlated with development on the receptive test for the CMC-C group. WMC did not moderate learning differentially for high or low WMC
learners in any of the four groups on either of the productive assessments. The experimental effect was therefore not dependent on WMC except for high WMC learners, in the FTF-C group, on the receptive test.

**Hypothesis 3 on WMC**: Based on the predictions of the Cognition Hypothesis, it was predicted that WMC would moderate L2 development, especially in the more complex tasks.

Hypothesis 3 was rejected. WMC did not moderate L2 development on the more cognitively complex tasks, and notably, not at all in the CMC mode.

**Research Questions 7**: Independent measures of cognitive complexity

7. How does the combination of task complexity and modality (i.e., FTF+C, FTF-C, CMC+C, CMC-C) affect learners’ reported independent measures of cognitive complexity (time judgments, anxiety/perceived difficulty questionnaire)?

For time judgments:

For Treatment 1, both of the +complex groups, regardless of the mode in which they had carried out the task, reported the time for task as significantly more than the real time it had taken them. Both of the -complex groups, regardless of the mode in which they had carried out the task, reported the time for task as significantly less than the real time. Task complexity therefore led to a significant discrepancy in time judgment, with the +complex groups always guessing more time and the -complex groups always guessing less. By Treatment 2, the two complex groups still estimated more time and the two simple groups still estimated less time than their real time, but the difference between guess time and real time decreased. For the second treatment, a significant difference was only found between the FTF+C and CMC-C groups.
For anxiety and perceived difficulty:

Modality in which the task was carried out appeared to affect reported anxiety and perceived difficulty the most, with the two FTF groups reporting the highest levels, followed by the two CMC groups. Within the FTF and CMC groups, task complexity affected anxiety and perceived difficulty: the complex groups reported higher levels than the simple groups. During both treatments, the order, from highest to lowest, of reported anxiety and perceived difficulty was:

FTF+C > FTF-C > CMC+C > CMC-C

A significant difference was only found between the FTF+C and CMC-C groups for both treatment sessions however, with the FTF+C always reporting significantly more anxiety and perceived difficulty than the CMC-C group.

Hypothesis 4 on independent measures of cognitive complexity: Based on the predictions of the Cognition Hypothesis, it was predicted that learners’ reported independent measures of cognitive complexity would be affected by increases in task complexity, with the more complex tasks resulting in higher anxiety/perceived difficulty and great subjective time judgments of time on task.

Hypothesis 4 was confirmed for only one of the measures: time judgments. This measure seemed to best address the construct of cognitive complexity, as it held across modality. The measure of anxiety and perceived difficulty seemed to be affected by a combination of complexity and mode, with significant differences only between FTF+C and CMC-C.
7.6 Justification for Further Analyses

These results rendered the pursuit of additional analyses necessary. For example, if uptake and WMC did not affect L2 development, what could have explained the group differences? What was unique about the way Task Complexity differentially interacted with Modality, i.e., the performances of participants in CMC-C versus FTF+C? Also, much variance was observed in the +complex condition, as demonstrated by the box plots. Could a deeper probe into the performance of the outliers that answered correctly versus those that did not be insightful? Indeed, it did. First, the researcher tallied and compared the total number of recasts, the amount of uptake, and the number of produced forms during the treatment. Second, a close examination of the concurrent processes as demonstrated by qualitative comments observed by participants during the treatments was carried out. It was found that on many occasions, learners noticed the form during the treatment and/or tried to develop a rule for the form. These follow-up analyses proved to be the clinching factors in explicating the results summarized thus far, are explained in greater detail below.

7.6.1 Recasts, Uptake, and Production-of-Form During the Treatment

The descriptive statistics for mean number of recasts, mean amount of uptake, and mean times that the past subjunctive was produced during the treatment are provided for each group in Table 50 below:
Table 50. Recasts, Uptake, and Production-of-Form During the Treatment

<table>
<thead>
<tr>
<th>Group</th>
<th>Total # Recasts during treatments</th>
<th>Total # Uptake during treatments</th>
<th>Total # Production of form during treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>16.44 (4.31)</td>
<td>11.50 (4.00)</td>
<td>3.56 (4.31)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>17.72 (2.47)</td>
<td>12.78 (3.89)</td>
<td>2.28 (2.47)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>11.12 (5.70)**</td>
<td>0.53 (1.01)**</td>
<td>8.88 (5.70)**</td>
</tr>
<tr>
<td>CMC+C</td>
<td>17.44 (3.77)</td>
<td>0.53 (0.72)**</td>
<td>2.56 (3.77)</td>
</tr>
</tbody>
</table>

Mean (SD) **p < .01

Table 50 shows the following results. First, the CMC-C group clearly received fewer recasts than the other three groups; where participants did not receive a recast, they were producing the form correctly. A one-way ANOVA with Group as the independent variable and Total Number of Recasts as the dependent variable revealed that this difference was significant ($F(3, 69) = 9.27, p = .00$), with the CMC-C receiving significantly less recasts than the three other groups ($p = .00$). This was indicative that participants in the CMC-C group arrived at the structure faster than the other three groups.

Table 50 also shows that, after receiving a recast during the treatment, the two CMC groups produced noticeably less uptake than the FTF groups. A one-way ANOVA with Group selected as the independent variable and Total Number of Uptake as the dependent variable revealed that the different in uptake production was significant ($p = .00$). A post hoc Scheffé showed that the significant difference was between the two CMC groups and the two FTF groups. In other words, the production of uptake was dependent on mode. The CMC-C group produced significantly less uptake (mean: 0.53) than the FTF+C ($p = .00$) and the FTF-C group ($p = .00$). The CMC+ group also produced significantly less uptake (mean: 0.53) than the FTF+C ($p = .00$) and the FTF-C group ($p = .00$). Note that the total amount of uptake produced
by the CMC-C and CMC+C groups was identical. This showed that modality, not task complexity, significantly affected whether or not participants produced uptake, with uptake having drastically less of a presence in the CMC mode. A visual representation of this difference is provided in Figure 24 below:

**Figure 24. Mean Production of Uptake According to Group**

![Bar chart showing mean production of uptake by group](image)

Lastly, Table 50 showed that the CMC-C group seemed to produce the form during the treatment much more than the other three groups. To test if this was statistically significant, a one-way ANOVA with group as the independent variable and total production as the dependent variable revealed a significant difference among groups; indeed, the CMC-C group produced the past subjunctive form significantly more than the other three groups ($p = .00$). Since uptake was found to be significantly affected by mode and not complexity, the researcher then explored whether or not production of the form during the treatment was predictive of development. Linear regressions were performed between total amount of production and all of the post and
delayed posttests: immediate FTF productive post, FTF productive delayed posttest, the immediate CMC productive post, the CMC productive delayed posttest, and the immediate receptive post and delayed receptive posttest. The regression between production of the form and performance on each assessment was highly significant at the \( p = .00 \) level. Regression reports are given below in Table 51:

**Table 51. During-Task-Production as a Predictor of L2 Development**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>During-Task-Production as a Predictor of L2 Development</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Immediate post</td>
</tr>
<tr>
<td>FTF production development</td>
<td>(.520^{**})</td>
</tr>
<tr>
<td>CMC production development</td>
<td>(.571^{**})</td>
</tr>
<tr>
<td>Receptive learning</td>
<td>(.761^{**})</td>
</tr>
</tbody>
</table>

\*\*\( p = .00 \)

Therefore, production of the form during the treatment was significantly predictive of development for all of the post and delayed posttests, not uptake or WMC. This showed that the more a participant tried to produce the form during the treatment, the better he/she performed on the assessments.

These follow-up analyses showed that amount of recasts and production of uptake did not predict learning outcomes, and that WMC was also not related to L2 development. But what else could explain the superior performance of the CMC-C group, and the comparability of CMC-C with FTF+C? The CMC-C group produced the form (during the treatment) significantly more than the other groups, and they received significantly less recasts and produced significantly less uptake. The FTF+C group received significantly more recasts and produced significantly more uptake, and their performance was not too far behind the CMC-C group;\(^{35}\) FTF+C even outperformed the CMC-C group on the receptive immediate posttest. How was carrying out a —

\(^{35}\) FTF+C also reported significantly higher anxiety/perceived difficulty than CMC-C
complex task in the CMC mode similar qualitatively to a +complex task in the FTF mode? Recall that a great deal of variance was observed in the +complex groups, irrespective of modality; this was shown by the box plots. If WMC nor uptake was related to development, and if production of form was highly significantly in the CMC-C group only, how could the results be explained? More follow-up analyses were carried out in an attempt to explain this by re-visiting the treatment sessions. Attention was drawn to online processes that occurred during the treatment sessions as being potentially explanatory of results.

7.6.2 Analysis of Concurrent Processes During the Treatments

The researcher noted that, during the two treatment sessions, participants often made comments about the nature of the task and the linguistic item for which their erroneous production was being recasted. Originally for the purpose of gaining qualitative data to explore the construct of task complexity, the researcher tabulated every comment and/or action by participants that might have been relevant. Given that all FTF production was recorded and that all CMC production was screen recorded, the interactions could be listened to and watched again. Because CMC interaction took place with the researcher and participant in separate rooms, the participant was often in a room or office alone. Some participants in the CMC groups made comments to themselves or spoke aloud, and the screen recorder captured this audio. Also captured were their moves on the computer, such as mouse movements, output that was reformulated or erased, and even cases of learners scrolling back up to see earlier sections of the CMC chat conversation. Alongside FTF comments and other behaviors, actions in the CMC mode were also tabulated. In both modes, it was observed that some participants indicated they noticed the past subjunctive form, while some went beyond noticing, and tried to formulate a rule concurrently during the task. These participants engaged in online formulation of a rule, or
in other words, demonstrated of awareness at the level of Understanding (i.e., Schmidt’s 1990 noticing hypothesis). Evidence of noticing and awareness at the level of Understanding occurred in both the FTF and CMC modes. Therefore, for the purpose of the present analysis, instances of noticing a form and of formulating a rule will be discussed, as they proved to be relevant for the purpose of the current study.

7.6.2.1 Evidence of Noticing

In both the CMC and FTF modes, there were numerous occasions of participants’ noticing (that is, awareness at the level of Noticing) the form. Examples include:

7.6.2.1.1 FTF Mode

In the FTF Mode:

1. ¿Qué tenso es gustara? (What tense is gustara?)
2. invitara? oh is that like the um ... perfect? Na never mind
3. Qué tiempo es pensara, despidiera? (What tense is pensara, despidiera?)
4. Qué es reaccionara? Is that the future perfect? (What is reaccionara?)
5. What are the -era forms?
6. Mirara o miraba? Qué es mirara? (Mirara o miraba? What is mirara?)
7. I keep missing the .... ... -iera? What is that?
8. Would do it? [This learner thought the form was the conditional]

7.6.2.1.2 CMC Mode

In the CMC Mode:

8. diera, ok
9. ara*

(Next line) ok
10. Que tenso usaba? (What tense were you using?)

11. creyera? Que tenso usabas? (creyera? What tense were you using?)

12. porque usaste la tensa futuro? (why did you use the future tense?)

13. Por que el subj. ... ok (Why the subj. ... ok [the learner erased this and never sent it])

14. After the treatment, the participant called the researcher over to his laptop, scrolled up in the CMC conversation to a past subjunctive form, and said to the researcher:

“What is this form? I don’t think I’ve learned it yet.”

15. futuro … ok (future … ok [This learner erroneously thought it was the future tense]

16. line 1: si, porque usa la forma (yes, why use the form)
   line 2: de fuera (of fuera)
   line 3: en este situation? (in this situation?)

17. Porque es fuera y no fue (Why is it fuera and not fue)

   All of these comments/actions serve as examples of the participant having noticed the form. Some indicated that they were not familiar with the form by saying “what is that form?” Others indicated that they had not learned it yet or asked the researcher what it was.

7.6.2.2 Evidence of Awareness at the Level of Understanding

   At the same time, some participants went beyond noticing the form, and demonstrated a higher level of awareness by minimally identifying the targeted structure as the subjunctive or the past subjunctive. In some instances there were exemplars of hypothesis testing and rule formations, features associated with awareness at the level of Understanding (i.e., Leow, 1997; Rosa & Leow, 2004; Rosa & O’Neill, 1999). Exemplars include:

---

36 Following Smith and Sauro (2009), text with a strikethrough indicates text that had been typed and was then deleted before the message was sent.
7.6.2.2.1 FTF Mode

In the FTF Mode:

1. Fuera? Es el subjunctive pero yo no sé el subjuntivo … (Fuera? It’s the subjunctive but I don’t know the subjunctive…)

   (later during the treatment): quisiera ... quisier ... a... [whispered to self] has to be subjuntivo … (subjunctive)

2. (Immediately after Treatment 1, FTF+C Group): That was HARD. That has to be the past subjunctive.

3. Hmmm ... it’s the subjunctive I think ... (Later during treatment): estaba avergonzada de que despedía, despidi... despidie ...

despidieron, despidieron a ella? (she was embarrassed that they fire, fire … fire … fired-
[IND] … fired-[SUBJ] her?)

4. pienso que es el subjuntivo. (I think that it’s the subjunctive).

5. I haven’t done the past subjunctive. I think I missed it in high school.

6. Quería que ... I don’t know how to say it in the subjunctive, and I know it has to be the subjunctive.

7. Oh ... it’s the subjunctive! It took me forever to figure it out!

   I’m always forgetting the emocional ... hiciera?

   All right. Well I know how to use it. I figured it out.

8. Mirara? Mirara .... mirara es past subjunctive? (mirara is past subjunctive?)

9. I just realized it’s ... what you were saying ... ara? is the past subjunctive.

10. rompiera .. is that the past perfect subjunctive?

---

37 IND refers to indicative morphology, SUBJ refers to subjunctive morphology
In the CMC Mode

11. (Immediately after treatment) “It should be past subjunctive, I’m just really rusty so that’s why like … fuera? I’ll be like ‘sea’. Because I know the present subjunctive.”

12. (Same Participant, Treatment 2): “That is definitely the past subjunctive. Can I ask you about qué vs. de qué? I definitely noticed a pattern.”

13. Pasado del subjuntivo? No lo se … (Past subjunctive? I don’t know)

14. Este es el subjuntivo pasado? (Is this the past subjunctive?)

15. I’ve never heard of so much past subjunctive.

16. cambiara … o, cambiera? (change … or, change?)

researcher → dime tú. (you tell me).

participant: es emocion, no? So, subjuntivo? (it’s emotion, no? So, subjunctive?)

17. Oh … subjunctive? No aprendo past subjunctive. (Oh, subjunctive? I don’t learn past subjunctive)

18. oh ok subjunctive past tense ..

(later in treatment) right, again, the subjunctive

19. Porque es fuera y no fue?

(Later during treatment): Ah, si, usando el sujentivo, no? Porque fueron es (Ah, yes, using the subjunctive, no? Because fueron is [participant erased this last part, sending the message without it])

20. dijeron

next line → no no dijeran porque es subjuntivo (no, no, dijeran because it’s the subjunctive)

21. este es el subjuntivo pasado_ (this is the past subjunctive_)
22. por que el no creia que el scout queria hablar con ellos (because he didn’t believe that the
scout wanted-IND to talk with them)
next line → o quisiera? (or wanted-SUBJ?)
next line → si past subjunctive ok (yes past subjunctive ok [erased this part before sending])

23. invite invitaron invitar [5 second pause] invitara(n) (invited-IND invited-PL invited-
[5 second pause] invited-PL-SUBJ]
[All strikethroughs were erased by participant before sending]

7.6.3 Coding for Noticing and Awareness

In order to investigate whether noticing the form or indication of awareness at the level of
Understanding (e.g., Leow, 1997; 2001) could have explained the results, all of the participants’
interactions were coded for (1) as having demonstrated noticing as one variable, and then (2) for
having indicated awareness at the level of understanding as another. (Note, it is important to state
here that just because a learner did not state anything explicitly, speak or whisper aloud, or type a
comment about noticing or developing a rule for the form, does not mean that they did not notice
the form or were not aware of it.) The percentages (with standard deviations) of number of
participants per group who reported noticing or reported awareness at the level of Understanding
are reported below in Table 52:
Table 52. Means and Standard Deviations of Noticing and Awareness by Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Reported Noticing</th>
<th>Reported Awareness at the Level of Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTF+C</td>
<td>78% (.43)</td>
<td>44% (.51)</td>
</tr>
<tr>
<td>FTF-C</td>
<td>44% (.51)</td>
<td>22% (.43)</td>
</tr>
<tr>
<td>CMC-C</td>
<td>76% (.44)</td>
<td>47% (.51)</td>
</tr>
<tr>
<td>CMC+C</td>
<td>53% (.51)</td>
<td>6% (.24)</td>
</tr>
</tbody>
</table>

Means (SD)

Next, two linear regressions were performed to address the extent to which reported noticing and reported awareness at the level of Understanding predicted L2 development. Regression reports provided below in Table 53:

Table 53. Reported Noticing and Reported Awareness at the Level of Understanding as Predictive of L2 Development

<table>
<thead>
<tr>
<th></th>
<th>FTFpost</th>
<th>FTFdpost</th>
<th>CMCpost</th>
<th>CMCdpost</th>
<th>Receppost</th>
<th>Recepdpost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noticing</td>
<td>.809</td>
<td>.095</td>
<td>.820</td>
<td>.446</td>
<td>2.24</td>
<td>1.43</td>
</tr>
<tr>
<td>Awareness</td>
<td>5.34**</td>
<td>5.02**</td>
<td>4.97**</td>
<td>5.81**</td>
<td>7.25**</td>
<td>8.34**</td>
</tr>
</tbody>
</table>

**p < .001

As Table 53 shows, noticing was not a significant predictor of L2 development on any of the assessments. Reported awareness at the level of Understanding was found to significantly predict L2 development.

A one-way ANOVA was then performed with group (i.e., experimental condition) as the independent variable and reported noticing as the dependent variable. The results revealed no significant differences between groups ($F(3, 66) = 2.21, p = .095$), and are demonstrated below in Figure 25:
Next, a one-way ANOVA was performed with group once again as the independent variable and reported awareness at the level of understanding (i.e., rule statement, rule formation, and/or the explicit naming of the targeted linguistic item) as the dependent variable. Results indicated a significant main effect for group ($F(3, 66) = 3.40, p = .023$). A post hoc Scheffé showed that FTF+C and CMC-C reported awareness significantly more than FTF-C and CMC+C, with the FTF+C (averaging .44) and CMC-C (averaging .47) on the one hand, and FTF-C (averaging .22) and CMC+C groups (averaging .06) on the other hand. In fact, there was no significant difference in reported levels of awareness between the FTF+C and CMC-C groups ($p = .999$). This is demonstrated graphically in Figure 26 below:
Awareness thus appeared to be the centripetal factor that explained the comparable development that took place in FTF+C and CMC-C; treatment in either the FTF+C or CMC-C led to more awareness of the form and there was no statistical difference between these two groups.

Given that production of the form during the treatment was also found to be predictive of L2 development, the researcher ran regression analyses for both production and reported awareness to compare the two. Both were significant predictors, though awareness was slightly more related to L2 development than production of form during the treatment on all assessments except for the CMC immediate posttest, which is demonstrated in Table 54:

**Table 54. Production Compared to Reported Awareness as a Predictor of L2 Development**

<table>
<thead>
<tr>
<th></th>
<th>FTFpost</th>
<th>FTFdpost</th>
<th>CMCpost</th>
<th>CMCdpost</th>
<th>Receppost</th>
<th>Recepdpost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production</td>
<td>.520**</td>
<td>.429*</td>
<td>.571**</td>
<td>.562**</td>
<td>.761**</td>
<td>.779**</td>
</tr>
<tr>
<td>Awareness</td>
<td>5.34**</td>
<td>5.02**</td>
<td>4.97**</td>
<td>5.81**</td>
<td>7.25**</td>
<td>8.34**</td>
</tr>
</tbody>
</table>
7.6.4 Noticing at the Level of Awareness as Predictive of L2 Development

This indicated that both the variables of production of form *during* the treatment and awareness at the level of Understanding (hypothesis testing, rule formation, or naming the targeted linguistic item) was what appeared to most strongly predict L2 development, not uptake or WMC. However, production only distinguished CMC-C; awareness is what explained the similar behavior of FTF+C and CMC-C as well as their advantage over the other groups. In other words, increased task complexity appeared to result in more production of the form and reported awareness, but in the FTF mode only. The opposite proved true in the CMC mode, where the complex task led to the most production and reported awareness of the form.

This leads to the discussion chapter, which will qualitatively explain the results and provide insight as to the unique interaction of task complexity and modality.
CHAPTER 8: DISCUSSION AND CONCLUSION

In this chapter, the results of the present study are discussed in relation to each research question, as are the findings from the additional analyses. A discussion on the qualitative differences from interaction with tasks of more or less cognitive complexity and in the FTF versus CMC mode is provided, with sample transcriptions from dialogues to illustrate these differences. The roles of uptake and individual differences in WMC are then examined. The chapter puts forth implications about the Cognition Hypothesis and the role that increases in task complexity may have in promoting opportunities for awareness. The chapter concludes with a discussion on the importance of doing exploratory analyses, most especially into the concurrent processes in which learners engage during task-based interaction, to gain a better understanding of how SLA is promoted.

8.1 Research Questions 1-4

The first four research questions were:

1. What are the combined effects of task complexity and recasts on L2 development?

2. What are the combined effects of modality (FTF vs. CMC) and recasts on L2 development?

3. Is there an interaction effect between task complexity and modality when recasts are held constant?

4. For learners in the +complex groups (both in FTF and CMC): Are tailor-made items answered more accurately than non-tailor-made items on the assessments?

In addressing Research Questions 1 and 2, it was found that Task Complexity plus recasts did not differentially affect learning, nor did Modality plus recasts. These results were obtained by
looking at Task Complexity and Modality as separate independent variables. However, a triple interaction effect (Research Question 3) between Task Complexity, Modality, and Time was found for all three assessments: FTF production, CMC production, and the multiple-choice reception test. Therefore, examining this triple interaction via an analysis of each of the four groups’ performance (i.e., FTF+C, FTF-C, CMC-C, CMC+C) was the most revealing when interpreting the results. On all three of the assessments, the +complex task resulted in the highest gains in the FTF mode. Contrarily, the +complex task resulted in hardly any development in the CMC mode, and in fact the -complex task in the CMC mode led to the highest amount of development. The CMC-C group significantly outperformed FTF-C and CMC+C in performance on both the FTF and CMC production assessments. The FTF+C group also performed better than the FTF-C and CMC+C groups on the production tests, but the differences were not significant. For receptive development, the FTF+C group performed the highest, followed closely by the CMC-C; both FTF+C and CMC-C performed significantly better than the FTF-C and CMC+C groups on this reception test. These findings therefore confirm the first hypothesis, that more complex tasks plus recasts would lead to the most L2 development, but in the FTF mode only.

In other words, as the Cognition Hypothesis predicts, increases in task complexity did lead to more learning, but differentially so according to the mode in which learners carried out the tasks. Engaging in intentional reasoning combined with recasts in the FTF mode led to greater L2 development of the past subjunctive than not engaging in intentional reasoning. The opposite proved to be true for the CMC mode. Those learners that had to reflect upon the intentional reasons of peoples’ actions in the CMC mode did not develop, while those in the -complex CMC condition outperformed all of the groups on the production assessments.
The similar performances of the FTF+C group and the CMC-C group were remarkable, especially given the differences between the two groups. For example, FTF+C produced significantly more uptake than CMC-C, and also reported significantly higher anxiety and perceived difficulty than the CMC-C group. CMC-C received significantly less recasts than the other three groups.

An examination of the different experimental conditions is illuminating when considering these findings. Below is an excerpt from a complex task in the FTF mode. Participants in FTF-C did not have to reflect upon the intentional reasoning of the characters, as they were given the intentional reasons in the L1 version of the story and in the comic strips. FTF-C learners only had to relay the events of the story in the past tense. (All excerpts are first presented in Spanish and are then translated into English).

### 8.1.1 Example of Interaction in FTF-C Condition

1. Participant 24: Señora Martínez siempre sonríó.
2. Researcher: Siempre sonreía, sí.
4. Researcher: ¿Por qué?
5. Participant 24: Porque estaba contenta por … su trabajo y …
6. Researcher: Estaba contenta de que …

10. Participant 24: Dijo que no fue ella, sí sí. Y entonces ella fue a su cuarto, y …
11. Researcher: Pero primero, ella lloró, verdad?
12. Participant 24: O sí sí es verdad y estaba sorprendida de que Sra. uh Martínez uh um creó que ella
13. Researcher: -Estaba sorprendida de que Sra. Martínez creyera

15. Participant 24: Su cuarto estaba un desastre porque Sra. Martínez … ¿como se dice search?
16. Researcher: Buscar
18. Researcher: Es verdad. Y entonces ¿qué hizo?
19. Participant 24: Um ella tiró la puerta. Estaba molesta, molesta de que buscó por sus cosas.
20. Researcher: Estaba molesta de que buscara por sus cosas.
21. Participant 24: Sí

Translation:

1. Participant 24: Señora Martínez always smiled-PRET.
2. Researcher: Always smiled-IMP, yes.
3. Participant 24: Yes, always smiled-IMP.
4. Researcher: Why?
5. Participant 24: Because she was happy for … her job and…
6. Researcher: She was happy that …
7. Participant 24: She was happy that Srta. Gómez worked-IND for them.
8. Researcher: She was happy that Srta. Gómez worked-SUBJ for them.
10. Participant 24: She said that it wasn’t her, yes yes. And then she went to her room, and …
11. Researcher: But first, she cried, right?
12. Participant 24: Oh yes that’s true and she was surprised that Sra. Martínez uh um thought-IND that she
13. Researcher: -She was surprised that Sra. Martínez thought-SUBJ
14. Participant 24: thought-SUBJ that she um stole the pearls.
15. Participant 24: Her room was a disaster because Sra. Martínez … how do you say to search?
16. Researcher: Buscar
17. Participant 24: Yes because Sra. Martínez searched in her room. And the pearls weren’t there.
18. Researcher: That’s true. And what did she do then?
19. Participant 24: Um she slammed the door. She was annoyed, annoyed that she [Sra. Martínez] searched-IND through her things.
20. Researcher: She was annoyed that she searched-SUBJ through her things.

As can be seen, the participant above easily communicated the intentional reasoning of the main character (turns 7, 12, 19) each time. This is because, as explained in the Methodology Chapter, the intentional reasons were already provided for her in the L1 story and on the comic strip cards she used to retell the story. Turns in the discourse are short, brief, and are complete—there is no scaffolding necessary in order to complete the task. The researcher’s corrective recast followed immediately after each erroneous production of the past subjunctive (turns 8, 13, and 20). Note
too that this participant produced uptake following the first two recasts, and in fact, for the second one, not only uptook the form but continued on, incorporating it into another sentence. The third example was only an acknowledgement: sí (yes) to the recast. Even though she often produced uptake after feedback, this participant did not produce one form of the past subjunctive on either of the FTF or CMC productive posttests or delayed posttests.

8.1.2 Example of Interaction in FTF+C Condition

The second example is from a participant in the +complex group, also in the FTF mode:

1. Researcher: Y qué es lo que Sra. Martínez siempre hacía, en la mañana cuando llegaba la Sra. Gómez?
4. Participant 17: um … puesto que Sra. Martínez um Sra. um …. hace um sabía que Sra. Gómez va a um cómo se dice take care of things?
5. Researcher: cuidar de todo
6. Participant 17: cuidar de todo, sí
8. Participant 17: estaba feliz y like ... happy
9. Researcher: alegre
10. Participant 17: sí estaba feliz y alegre que -
11. Researcher: de que -
12. Participant 17: de que Sra. Gómez trabajó para ellos
13. Researcher: Estaba feliz y alegre de que Sra. Gómez trabajara para ellos
14. Participant 17: trabajara … ok.

-----

1. Researcher: Y ¿entonces qué hizo Sra. Gómez?
2. Participant 17: Sra. Gómez fue a su cuarto que era … un desastre!
3. Researcher: Totalmente, sí! Y ¿por qué era un desastre?
5. Researcher: y entonces qué hizo Sra. Gómez?
6. Participant 17: Sra. Gómez tiró la puerta, porque …
7. Researcher: ¿Por qué? Cuál fue su intención?
8. Participant 17: Sí, sí, tiró la puerta porque estaba enojada
9. Researcher: de qué?
10. Participant 17: de qué um la familia o Sra. Martinez buscó-IND por sus uh posesiones
11. Researcher: Estaba enojada de que Sra. Martínez buscara-SUB por sus posesiones
12. Participant 17: sí
Translation

1. Researcher: And what is it that Sra. Martínez always did, in the morning when Srta. Gómez arrived?
3. Researcher: Excellent, she always smiled. Now, think, why did she always smile? In that moment, that specific action - smiling when Srta. Gómez arrived - what emotion did she have? What caused that action?
4. Participant 17: um … given that Sra. Martínez um Sra. um does um knew that Srta. Gómez will um how do you say take care of things?
5. Researcher: cuidar de todo
6. Participant 17: cuidar de todo, yes
7. Researcher: Good and what emotion did that give her? Sra. Martínez .. what?
8. Participant 17: was happy and like … excited
9. Researcher: excited
10. Participant 17: yes she was happy and excited that-
11. Researcher: that-
13. Researcher: She was happy and excited that Srta. Gómez worked-SUBJ for them.

----

1. Researcher: So then what did Srta. Gómez do?
2. Participant 17: Srta. Gomez went to her room that was .. a disaster!
3. Researcher: Yes, totally! And why was it a disaster?
4. Participant 17: Because Sra. Martínez searched in her room.
5. Researcher: and so then what did Srta. Gómez do?
6. Participant 17: Srta. Gomez slammed the door, because …
7. Researcher: Why? What was her intention?
8. Participant 17: Yes yes, she slammed the door because she was angry
9. Researcher: [that]?
10. Participant 17: That um the family or Sra. Martínez searched-IND through her possessions
11. Researcher: She was angry that Sra. Martínez searched-SUBJ through her possessions.
12. Participant 17: Yes

An immediate glance as the excerpt shows longer discourse moves in FTF+C, to include scaffolding, negotiation of meaning, and the participant asking the researcher how to say feelings or other vocabulary items in Spanish (i.e., turns 4 and 8). Where only 9 turns were necessary to set up a recast for the first past subjunctive opportunity in the FTF-C condition (Example 1), 14 turns were necessary in this example. Unlike Participant 24 in the FTF-C condition, this
participant did not have the intentional reasons given to him in either the L1 story or comic strip; he had to come up with the intentions/mental states himself. The researcher asked him to reflect on the intentional reasons in turns 3 and 7 (i.e., “Why did she smile?” “Think of the emotion that caused her action at that moment.” “What was her intention?”). As with the erroneous production in the FTF-C, recasts were provided immediately after the error, setting up an immediate juxtaposition of the two utterances (erroneous with the correct one) so that a cognitive comparison of the forms could be carried out. All uses of the past subjunctive were mandated by the participant’s own intentional reasons; his intentional reasons thus served as some of the main points of the story. Note too that there are many more pauses (i.e., uh, hmm, etc.) in the complex condition than in the -complex condition. It seemed to be the case that more thinking, engaging, and reasoning were going on with this task. This was observable in person too; the participant would often look up, tap his forehead, try to say something but then reformulate it, struggled at times, demonstrating overall a major effort. Included in this effort were gestures of frustration at times. After this task, the participant said to the researcher: “That was hard,” but that he also had really enjoyed it, and wished he had more opportunities to converse one-on-one while having to relay information in Spanish.

Also notable is that uptake was produced by this participant after the first recast (line 14), but was not after the second (line 12). Though uptake was not always produced, this excerpt comes from a learner who demonstrated L2 development of the past subjunctive on both the FTF and CMC posttests, and even higher development on the delayed posttests. He achieved perfect scores on the immediate and delayed receptive tests.
8.1.3 Example of Interaction in CMC-C Condition

The next excerpt come from a complex condition in the CMC mode. An inserted image from the actual CMC chat file is shown, with the translation in English to follow:

( Participant 46 is blue, researcher is white)
Example 3 Translation:

1. Participant 36: she waited for sra. gomez and sra. martinez smiled when sra. gomez arrived
2. Researcher: Why?
3. Participant 36: because she was happy that sra. gomez worked-IND for them
4. Researcher: She was happy that sra. gomez worked-SUBJ for them
5. Participant 36: yes
6. Researcher: alright, let’s go to the next one!

------

7. Participant 36: srta. gomez went to her room and saw that it was a disaster
8. Researcher: Yes!
9. Researcher: because Sra. Martinez had looked there
10. Researcher: and then what happened?
11. Participant 36: Srta. Gomez left the house and slammed the door really hard because she was angry that they looked-SUBJ through her things

In looking at the interaction from CMC-C, it is obvious that overall few turns were necessary to communicate the main points of the story. The amount of discourse per turn is also
quite short in length. The juxtaposition of corrective recasts immediately after sentences with erroneous production are ideally set up, because the error sentence and the recast sentence are identical except for the past subjunctive form. Because this communication was in the CMC mode, the corrective feedback was visually written out for the participant. In addition, the screen-recorded videos of the -complex condition in CMC showed that for the most part, participants were waiting for a response from the researcher before typing out their next message. Participants in CMC-C were not processing other messages, thoughts, or typing when they received the corrective recast, creating an ideal setting for the provision of feedback. Note that after receiving a recast, Participant 36 above did not produce uptake (turn 5). Then, in the next turn (turn 11), she attempted to actually produce the form: *buscabaran*. The participant marked the past tense indicative form of the very buscar (*buscaba*) with what she appeared to have noticed as the subjunctive morphology, -ran (creating *buscabaran*). Production of the form early on in the treatment was often observed for participants in CMC-C, and empirically confirmed later (recall that participants in this group produced the past subjunctive form during the treatment significantly more than the other groups). This excerpt came from a participant who demonstrated high L2 development on the productive and receptive assessments.

**8.1.4 Example of Interaction in CMC+C Condition**

The next excerpt comes from a participant who had to intentionally reason in the CMC mode. The first excerpt is notably longer, with 10 turns before the researcher had the opportunity to recast a sentence containing an intentional reason and then a dependent clause that needed the past subjunctive. The translation is provided after the excerpt:

(Participant is blue, researcher is white):
Porque las mujeres habían sido amigas. Ok, pero qué emoción se le dio ese hecho?

Piensa en la emoción

4:35 PM

Sra. Martínez se sentía muy alegre tener una de mejor trabajadoras en su vida

Entonces, enfocándonos en la Srta. Gómez ... Sra. Martínez se sentía muy alegre de que Srta. Gómez ... qué?

Lo siento, no entiendo la pregunta

Quiero que me digas la frase entera. Solo hay que relacionarlo con la acción de Srta. Gómez

Sra. Martínez se sentía muy alegre de que Srta. Gómez ... qué?

Sra. Martinez se sentía muy alegre de que Sra. Gomaz porque ella no hacía todo el trabajo en la casa

:-/

4:40 PM

Ah, entonces, Sra. Martínez se sentía muy alegre de que Sra. Gómez hiciera todo el trabajo en la casa.

sí
Translation:

1. Researcher: because the women had been friends. Ok, but what emotion did this fact cause her?
2. Researcher: Think of the emotion
3. Participant 57: Sra. Martinez felt very happy to have one of the best workers in her life
5. Participant 57: I’m sorry, I don’t understand the question
6. Researcher: I want you to tell me the entire phrase. You simply have to relate it [Sra. Martinez’ feelings] to Srta. Gomez’ action
8. Participant 57: Sra. Martinez felt very happy that Srta. Gomez because she did not do all of the work in the house
9. Participant 57: :-/
11. Participant 57: yes

Turn 2, where the researcher asked the participant to think of the emotion that caused the character’s action, was the third time an intentional reason had been requested by the researcher. Once the participant came up with “she felt very happy”, attempts to get the participant to relate that feeling to the actions of Srta. Gomez (the second character in the story) proved to be extremely challenging. (Recall that the past subjunctive requires a change in referent in the second dependent clause). There was finally a breakdown in communication by Turn 5, where the participant apologized and said that she did not understand what she had to do. After the researcher tried to clarify, the participant attempted to relate Sra. Martinez’s mental state with Srta. Gomez’s action, but then sent the researcher an emoticon message: :-/. This possibly indicated that she was confused and felt bad that she was not, in her opinion, performing the task well. Unlike a recast in the CMC-C group, the trouble trigger and subsequent negotiation repair were not so contingent in time in the CMC+C condition. The recast by the researcher in this example was delayed; while she was typing out the correction, the participant sent the emoticon message. In addition to being delayed, another problem with recasts in the CMC+C group was
that they were often not a perfect mirror of what the participant had said. So much negotiation work was required in order to get the participant to reflect on the intentional reasons and then relate them to characters’ actions that utterances were often long and contained many more errors than just the past subjunctive form. As shown in the example above, a comparison of the erroneous utterance with its correction required the participant to notice that (1) the second conjunction had been eliminated, (2) the referent in the dependent clause had been switched to Srta. Gomez, and (3) the verb was now marked with subjunctive morphology. Also, because they were so concentrated in meeting the demands of the task, learners in the CMC+C group often missed the recast entirely. The screen-recorded videos showed that while participants in the CMC-C group were observed waiting for a message from the researcher before typing out their next message, participants in the CMC+C group were almost always typing their next message when a recast arrived from the researcher. Turn taking in CMC+C was also problematic, because both the participants and the researcher were always typing at the same time as receiving messages. This serves as an example of why carrying out the more complex tasks were experienced so differently in FTF versus CMC.

8.1.5 Customizing Assessments to Measure Learning with Cognitive Complexity

After learners carried out the task, the use of customized and standardized assessments to measure L2 development proved to be revealing in this study. It was found that learners in the two complex groups answered tailor-made items on the post assessments significantly better than the critical items. This was a very important finding, because it shows that, as Nuevo (2006) contented, customized assessments may be better for measuring learning because they acknowledge what the learner contributed to the task. More studies are clearly needed that
measure learning when operationalizing cognitive complexity with both customized and standardized assessments.

8.2 Research Question 5

The fifth research question asked:

5. Does uptake mediate L2 development? In other words, if a treatment effect is observed, is it because of the production of uptake during interaction? If so, is the effect different under different experimental conditions?

Contrary to the predictions of the Cognition Hypothesis, the present study revealed that uptake did not predict learning, a finding that is congruent with the results of Révész, Sachs and Mackey (in press). In fact, it was found that Modality, not Task Complexity, is what significantly affected the production of uptake. (The fact that mean rate of uptake was identical for CMC-C and CMC+C is particularly illustrative of this finding). Regardless of level of task complexity, interacting in the CMC mode resulted in significantly less uptake than in FTF mode. This finding empirically confirms Smith’s (2005) recent suggestion that uptake has much less of a presence in the CMC environment.

Uptake has played an important role in SLA research because it is premised to serve as evidence of what learners notice (e.g., Chaudron, 1977; Lyster, 1998; Lyster & Ranta, 1997; Mackey, Gass, & McDonough, 2000; Sheen, 2004, 2006). But what was very surprising is that in the present study, uptake was often found to be negatively related to learning, especially for learning in the FTF+C mode. Hypothetically speaking, such a negative relationship implies that more production of uptake would lead to less L2 development.
While surprising, what this finding underscores is the fact that while uptake may indicate that learners have paid attention to feedback, any association of uptake with learners’ processing of feedback is at best discretionary. Only a few researchers have acknowledged that uptake can be a mere parroting of feedback on behalf of the learner, without really understanding what the feedback was for (Gass, 2003; Egi, 2010). A deeper probe into learners’ online processes, such as indications of noticing and awareness, may be more revealing for how learners’ process feedback and secondly, how task complexity mediates this outcome. This certainly seemed to be the case in the present study.

A possible explanation for the negative relationship between uptake and learning in this study is the way in which uptake was operationalized. Following Révész, Sachs & Mackey (in press), both partial and entire incorporations of the feedback were counted as uptake. Robinson (2007a) did a more fine-grained coding of uptake, looking at exact versus partial uptake. He found that when learners had to engage in intentional reasoning (+complex), they produced significantly more partial uptake, but not exact uptake. It may be the case that an exact incorporation of the recast is more of an imitation of the feedback, whereas partial uptake is more indicative of assimilation of the feedback. If learners repeat the recast exactly, it is difficult to know if they are engaging in real analytic thinking about what the recasted form means, perhaps due to temporary expedient circumstances. Partial uptake, on the other hand, could be more indicative of learners breaking up the feedback, trying to assimilate the content that has been corrected.

At the same time, partial versus exact uptake might make a difference depending on the linguistic item being investigated. For example, learning the Spanish past subjunctive involves recognizing that the semantic value of the main clause mandates mood marking in an embedded
clause. If the learner repeats the entire utterance (i.e., a main clause with an attitude-based verb, then the conjunction, then the dependent clause with a subjunctive), might that assist in acquiring the lexical classes that require the subjunctive? Or would partial uptake be more evident of feedback assimilation with the subjunctive? As Collentine (2010) highlighted, the Cognition Hypothesis has the potential to contribute to what is known about how learners acquire the Spanish subjunctive. Clearly, the construct of uptake needs to be looked at in much more detail, with different task types and linguistic items. A distinction of exact versus partial uptake also deserves greater attention in future research.

8.3 Research Question 6

The sixth research question was:

6. Does working memory capacity moderate L2 development? In other words, if a treatment effect is observed, is the effect different for learners of high or low WMC? If so, is the effect different under different experimental conditions?

Contrary to the predictions of the Cognition Hypothesis, WMC did not, for the most part, moderate learning in this study. In fact, the only group that showed a significant correlation between WMC and development were the high WMC learners in the FTF-C group, and this was for the immediate receptive test only. This meant that their higher WMC helped them on the post receptive test.

Applying these findings to Baddeley’s Model, a plausible explanation for the significant correlation between WMC and receptive development in the FTF-C group is that their higher WMC allowed them to notice the form more in the FTF mode, which assisted them in the receptive multiple-choice immediate posttest. These learners did not engage in intentional
reasoning, and might or might not have noticed (at the level of Understanding) the form as the past subjunctive. But due to their higher WMC, the auditory input experienced auditory processing in the phonological loop (one of the slave systems in WM). On the receptive test where they were shown the forms and had to choose which one was correct, the prior phonological processing assisted them in choosing the right form. Therefore, higher WMC led to higher receptive *recognition* of the form in the FTF-C group. It may be that the forms did not go through the phonological output buffer, as learners in FTF-C showed lower production rates of past subjunctive on the post-tasks. This is further exemplified with the fact that WMC was not correlated with L2 productive development in this group. One observes too that this only occurred for learners of higher WMC in the -complex task FTF condition, not in the CMC condition. Therefore, the extent to which WMC plays a role in -complex tasks versus +complex where learners are provided with feedback is worthy of more empirical investigation. Based on the findings of the current study, it may be the case that awareness at the level of Understanding, *not* WMC, is what is most related to L2 development in more cognitively complex tasks. It is an empirical question whether increases in the cognitive complexity of tasks promote awareness, and that this construct, as opposed to WMC, most predicts learning in +complex tasks; WMC may be more relevant for incorporation of forms in -complex tasks, where deeper processing is not so necessary in order to meet the demands of the task.

WMC did not moderate L2 development differentially for learners of high versus low WMC in any of the other groups. For the majority of learners in the CMC groups, WMC was in fact negatively correlated with development. In fact, for high WMC learners in the CMC-C group, WMC was *significantly* and *negatively* related to development. The negative correlations for the CMC groups indicate that WMC does not play a role in the CMC mode at all, for low or
for high WMC learners, contrary to what the literature suggests (Payne & Whitney, 2002). Much
more research is needed that looks at the effects of WMC in the FTF versus CMC environments.
Arguably, the findings on WMC reported here are robust because the spans, methodological
procedures for administration, and grading scheme for measuring WMC are highly attested in the
field of psychology (i.e., Conway et al., Engle et al., 1992; Engle, Tuholski et al., 1999; Turner
& Engle, 1989; Unsworth et al., 2005; Unsworth et al., 2009). At the same time, it is possible
that minimal effects for WMC were found because of the population of learners used in this
study. Participants in the present study were undergraduate university students, all young adults
at the “peak” of their cognitive control abilities (e.g., Craik & Bialystok, 2006). It is not known if
a ceiling effect took place, which could explain why WMC was not related to learning. A future
analysis of the current dataset presented in this study should examine variance in participants’
composite WMC scores, and also look at the standard deviations of the scores.

8.4 Research Question 7

The seventh research question was:

7. How does the combination of task complexity and modality (i.e., FTF+C, FTF-C,
CMC+C, CMC-C) affect learners’ reported independent measures of cognitive
complexity (anxiety/perceived difficulty questionnaire, time judgments)?

Participants’ time judgments were most revealing in regards to cognitive complexity; the
anxiety and perceived difficulty questionnaire reflected an interaction between cognitive
complexity and mode. That is, learners in the two +complex groups judged the task as taking
significantly more time than the time that had actually passed. Learners in the -complex groups
judged the tasks as taking significantly less time that had passed. The measure of anxiety and
perceived difficulty via questionnaires seemed to be most affected by mode, with the FTF groups reporting higher levels than the CMC groups; significant differences were only found between the FTF+C and CMC-C groups.

The time judgment data from this study implies that the complexity conditions were delivered and in both modes—more complex tasks were judged as taking significantly longer in the FTF and CMC environment. These findings indicate that subjective time estimations may be a robust way to predict complexity. When operationalizing cognitive complexity, it is crucial to measure the construct independently (Norris & Ortega, 2009). This study is the first to use time judgments to measure cognitive complexity; more research that uses time judgments as well as other independent measure of the construct are needed to see if the finding is replicated.

One other innovation in this study to show the effects of cognitive complexity was the use of box plots to visually demonstrate the variance that occurred in productive development. The outliers (values outside the threshold of the upper quartile) shown by box plots revealed that there was a great deal of variation in performance in the complex groups, irrespective of mode. Because statistical analyses such as ANOVAs compare means, box plots help to show differences in individual performance; in this study, they showed that individual performance was much more varied in the complex than the simple tasks. Larson-Hall and Harrington (2009) have recently argued that SLA researchers should employ more box plots specifically due to their robustness to outliers; few studies have, however. This study serves as one example of how box plots not only assisted in interpreting the results, but also confirmed the hypothesis that more complex tasks lead to greater variation in performance.
8.5 Noticing and Awareness: Exploring Learners’ Concurrent Processes

In consideration of the above findings, one of the main questions is why the CMC-C group did so well, especially given that learners in this group produced hardly any uptake. In addition, what explained the similarity of CMC-C with the FTF+C condition? Recall that FTF+C did second best on the FTF and CMC production tasks, and was the only group to continue in performance at the delayed posttest. FTF+C also had the highest scores on the receptive posttest. The fact that uptake and WMC were not predictive of L2 development rendered follow-up analyses necessary.

One of the original goals of the researcher was to tabulate all comments about learners’ perceptions of the tasks in order to differentiate +/-complexity (i.e., comments such as “this is difficult” or “that wasn’t so bad”). As the experiment progressed however, it was observed that many participants made comments indicative of the fact that they had paid attention to the form, were processing it, and were formulating hypotheses. As explained in the Results chapter, comments such as “tuviera… tuvie …RA… is that the subjunctive?” were all recorded and coded for. The finding that awareness was significantly correlated with L2 development showed that awareness appeared to be what most predicted learning. It was also the variable that statistically demonstrated the similarity between FTF+C and CMC-C: both were significantly higher in promoting a higher level of awareness than the FTF-C and CMC+C groups.

Stemming from Schmidt’s noticing hypothesis, where awareness at the level of Noticing is hypothesized to be the “necessary and sufficient condition for converting input to intake” (1990, p. 129), the construct of awareness has been highly attested in the field of SLA (e.g., Hama & Leow, 2010; Leow, 1997, 2000; Rosa & Leow, 2004; Rosa & O’Neill, 1999;
Sachs & Suh, 2007). These studies have found that noticing of the L2 forms facilitates SLA, and that higher levels of awareness are more strongly associated with L2 development.

The awareness data in this study seems to explain the uniqueness about CMC-C and FTF+C. The data also afforded important insights into the way in which language is processed in CMC, which is different than FTF, and how, consequently, learning might be different in the two modes. This is an important issue to investigate, especially as more language courses incorporate technology and CMC environments for tasks. The CMC mode allows for more control of one’s own intentions, especially when having to meet the demands of cognitive complex tasks. CMC obviates the need to respond immediately to the interlocutor, giving learners in the CMC-C group more time to reflect. In addition, the very nature of the CMC environment combined with the complex task seemed to have made correctional feedback more salient than in the FTF mode. The CMC-C participants clearly got to the form first before all other groups. This is evidenced by their early production of the form during the treatment (recall that CMC-C produced the past subjunctive significantly more during the treatment than everyone else). The screen recordings of CMC-C interaction are also very revealing. One participant in the CMC-C group did the following: 38

\begin{verbatim}
tuvo
tuvieron
\end{verbatim}

This participant wrote the past tense indicative and singular form of the verb tener (to have): *tuvo* (he had). He then erased it, and typed out the plural verb form, still indicative: *tuvieron* (they had). He then erased the indicative morphological ending -*ron*, and added -*ra*, marking the verb with subjunctive morphology. Before sending the message, the participant scrolled up quickly to

38 This example is very similar to the previous example with the verb *invitar* (to invite) from the examples of demonstrated awareness in CMC.
view a past subjunctive form (from a previous recast) for comparison. Upon seeing that it was correct, he scrolled down again, added the plural morpheme -n, and then sent the message to the researcher. This is a form of hypothesis testing in CMC, and is demonstrative of a high level of awareness. The concurrent data from the video recording showed that the learner formed a hypothesis about the form and then scrolled up to confirm his hypothesis. The CMC-C condition promoted high levels of awareness because the task did not pose high demands on the learner; the task and the environment allowed for a visual comparison of forms because learners in the CMC-C group had more time and more attentional resources available.

What is very interesting to consider is that, based on this study, increasing the cognitive demands of tasks in the CMC may completely mitigate those benefits of CMC that seem to promote awareness. In the CMC mode, engaging in intentional reasoning resulted in cognitive overload and confusion. The juxtaposition of the recasts after erroneous production was problematic and delayed. Cognitive complexity clearly has different effects according to the mode in which task-based interaction takes place.

Very different to CMC, having to engage in intentional reasoning in the FTF mode did promote more opportunities to become aware. Having to reflect upon the intentional reasons that caused the characters’ actions pushed learners in FTF to be more receptive of corrective feedback, processing it at a deeper level than the FTF-C condition. Some learners in the FTF+C group repeated the linguistic form slowly, seemingly processing the new subjunctive morphology that replaced their own indicative mood marking: “tuviera … tuverie … RA. Is that the subjunctive?” The fact that the intentional reasons were the participant’s own reflections may have also been a promoter of development — all interaction and negotiation work during the task was based upon what the learner herself came up with, making the task more learner-relevant.
Therefore, learners engaging in intentional reasoning in the FTF mode spent more time processing the correctional feedback they received because the demands of the task pushed them to do so, and much of the communicative goals were based around the learners’ own exchange. Learners in the FTF+C group formulated hypotheses, which included questions such as “is that the subjunctive?” or even production of new forms: “tengaban?” (this form is a creation of the past subjunctive with the present subjunctive root plus past tense morphology; the correct past subjunctive is *tuvieran*). The corrective feedback they received during the interaction assisted in confirming their hypotheses, which led to awareness at the level of Understanding; this was rarely observed in the FTF-C group. Language, therefore, may be processed quite differently in FTF versus CMC, and the cognitive complexity of tasks mediates this outcome.

### 8.6 Summary of Findings

The data from this study provide empirical evidence for Robinson’s Cognition Hypothesis, but in the FTF mode only. In FTF, engaging in a more cognitively complex task promoted a higher level of awareness, which, like in this study, has been empirically proven to lead to more L2 development. In CMC, the more complex task did not work, most likely due to potential cognitive overloading as revealed in the discourse samples. Contrarily, the -complex task in the CMC mode led to high L2 development, and this may have been due to the very nature of the CMC environment.

In this study, the production of uptake and WMC were not predictive of L2 development. Rather, awareness at the level of Understanding appeared to be most associated with learning. These findings support the strand of research showing that awareness predicts L2 development (Leow, 1997, 2000; Rosa & Leow, 2004; Rosa & O’Neill, 1999). The construct of awareness
seemed to most robustly explicate the findings and importantly, helped to explain the similarities of FTF+C and CMC-C. In revisiting the concurrent processes in which learners were engaging, the follow-up post analysis was able to provide a relatively clear picture of what was going on while learners were processing the L2 input. This study therefore highlights the critical importance of collecting concurrent data to elucidate the results.

8.7 Conclusion and Implications

8.7.1 Conclusion

Five main findings can be extracted from this study. First, the dissertation provides empirical support for Robinson’s Cognition Hypothesis in that increases in the cognitive complexity of tasks were found to lead to more L2 development, but it is possible that this only holds for the FTF mode. In this study, increases in task complexity in FTF appeared to promote deeper processing and subsequent higher level of awareness, which was found to significantly predict L2 development. This is supported by the findings that FTF+C performed better than FTF-C on all of the assessments.

Second, and related to the above point, task complexity and modality interact in unique ways for SLA. While the +complex task in FTF led to more L2 development, the opposite was found for the CMC mode, with the -complex task in CMC demonstrating the most development. Increasing the cognitive complexity of tasks induces greater cognitive effort on behalf of the learner in FTF. Doing so in the CMC had adverse and potentially unfavorable effects on L2 development, given that many participants apologized and reported confusion during the CMC+C interaction. One of the suggestions of the Cognition Hypothesis is to increase
sequentially the cognitive complexity of tasks for learners, which helps to approximate the real-world L2 tasks they will encounter. This finding may render the CMC mode limited in a sense, as carrying out more complex tasks alongside feedback provision in this mode do not appear to work. At the same time, the CMC mode might be incorporated into Robinson’s Triadic Componential Framework as a task design feature that is simple, one that could be carried out before advancing to more complex tasks.

Third, the important issue of transferability of L2 development from one mode to the other was attested in this study, but only partially. Specifically, learning as a result of task-based interaction in the CMC mode can be extended to the FTF mode, but this may depend on the cognitive complexity of the task. A quick comparison of development shown on the immediate and delayed post FTF and CMC productive assessments demonstrates this. The FTF+C and FTF-C groups achieved very similar scores on the FTF and CMC post assessments, showing a clear transfer of development from FTF to CMC. This finding held for the CMC-C group as well; mean scores were 4.88 and 4.79 on the CMC immediate and delayed posttest, respectively. For their FTF post assessments, the CMC-C group’s mean scores were 4.59 and 4.06, indicating a cross-modal transfer effect. The CMC+C group, however, achieved a mean score of 1.74 and 1.12 on the CMC immediate and delayed posttests, respectively; on the FTF productive assessments, however, mean scores were 1.35 and only 0.58 on the delayed post. The transferability aspect of L2 development in FTF and CMC must be emphasized, because if CMC+C only achieved a mean of half a point on the FTF delayed posttest, it is not clear that learning in this group did transfer to the FTF mode. Until now, the extent to which CMC learning translates into FTF performance abilities, and to a lesser extent, vice versa, has not been well established in the field. In order to test cross modal transfer effects, it is critical that more
studies employ assessment tasks in both modes, also looking at the effects of task design features as mediating these outcomes.

Fourth, neither uptake nor WMC were found to be related to L2 development in this study. The assumptions in the literature that higher WMC might be necessary to benefit from feedback (Mackey et al., 2002) and that the CMC mode benefits learners with lower WMC (Payne & Whitney, 2002), were therefore not supported. The predictions of Robinson’s Cognition Hypothesis that more cognitively complex tasks will lead to more uptake, and that task-based performance will be increasingly affected by cognitive variables such as WMC, were also not supported.

Finally, a more fine-grained operationalization of what constitutes uptake after feedback may be needed in future research that takes this variable into account. The concept of uptake as predictive of L2 development may need to be reconsidered given that manifestations of uptake do not appear to reflect robustly learners’ internal processes. Rather, this study found that demonstrations of hypothesis testing and awareness at the level of Understanding gathered in concurrent data were superior in significantly predicting L2 development, thereby underscoring the importance of collecting concurrent data to explicate learners’ processes.

8.7.2 Implications

8.7.2.1 Implications for Feedback and Uptake

This study has implications for the L2 literature on corrective feedback (i.e., Long, 1996). Corrective feedback may work best in the CMC mode with simple tasks that do not pose a cognitive overload. In the FTF mode, feedback in the form of recasts may work best congruently with more cognitively complex tasks, which, as shown in this study, pushed learners to form and test hypotheses more than do simple tasks. For the most part, the learners in the FTF-C group
were not compelled to process the recasts they received at a deeper level of processing, while those in the FTF+C were. Contrarily, recasts alongside cognitively more complex tasks in the CMC mode may be too much for learners to process. The cognitive complexity of tasks alongside recasts may have very different implications according to the mode in which the interaction is carried out.

Another important implication to this study is that feedback in and of itself may not predict learning. Instead, it is what the learner does with that feedback, and how he/she internalizes it. It may also be the case that uptake is not representational of internalization of correctional feedback. Therefore, it is possible that researchers in SLA are examining uptake with too many assumptions regarding its role. How learners perceive and process input is something done internally, and demonstrations of awareness at the level of Understanding of that input seems to be the best way to capture internalization, not the production of uptake. This notion seemed to be supported in the present study.

8.7.2.2 Theoretical Implications

The results from this study also have important theoretical implications, in particular for the Interaction Hypothesis (i.e., Long, 1996), TBLT, and the Cognition Hypothesis (Robinson, 2001, 2002, 2005, 2007, 2010; Robinson & Gilabert, 2007). This results from this study provide support for the Interaction Hypothesis because it was shown that interaction and feedback promoted learning in this study. The main premise of TBLT is that interaction and learning is successfully achieved through the use of tasks. This study showed however that the design of the task, particularly the different cognitive demands of the task, differentially mediate learning

---

39A relatively similar observation was made by Leow (1997) who pointed out that external manipulation of input, in this case, input enhanced to draw learners’ attention to targeted linguistic items, may not have its desired effect on learning due to several variables that include learners’ internal processes.
outcomes. The discourse environment of the task (FTF versus CMC) also played a significant role in how learners’ processed feedback and subsequently learned. This study therefore contributes to the strand of research on task design features that best promote interaction and learning.

This study also offers evidence for the Cognition Hypothesis in that task complexity (having to intentionally reason or not) alongside recasts, affected L2 development, though differentially so according to the mode in which the interaction was carried out. To revisit some of the claims of the Cognition Hypothesis, increasing the cognitive complexity of tasks will:
- “promote interaction and negotiation work, and heightened attention to, noticing of, and incorporation of forms made salient to the input; and
- individual differences in cognitive abilities (e.g. working memory) and affective factors (e.g. anxiety) will increasingly affect task-based performance and learning as tasks increase in complexity” (Robinson, 2001b, p. 56)

Indeed, task complexity did lead to more “noticing of … and incorporation of forms made salient in the input.” However, in this study, it seemed to be the case that increases in task complexity promoted more opportunities for awareness at the level of Understanding. Production of the form with a high level of awareness and deeper processing provided the most robust evidence of this relationship of task complexity and learning. It therefore may be the case that Robinson’s reference to “incorporation of forms” in his Cognition Hypothesis should be made more explicit by alluding to learners’ high level of awareness. If the findings of the present study are replicated, the Cognition Hypothesis might be elaborated by referring to the role of high levels of awareness, that is, more cognitively complex tasks promote opportunities to become more aware, and that this higher level of awareness leads to more learning. Awareness in this
study appeared to contribute significantly to what learners took in for further processing and to their subsequent L2 development (as also reported in previous studies), and more awareness occurred in the more cognitive complex task in the FTF mode.

8.7.2.3 Methodological implications

This study also has implications for research methodology in the field of SLA. First and foremost, this study highlights the importance of collecting concurrent data (i.e., recordings from the FTF interactions, screen recordings of the CMC interactions, a tabulation of qualitative comments by participants during the task). Without concurrent data, the follow-up analyses that revealed exemplars of hypothesis testing and awareness at the level of understanding would not have been possible. The data served as the crucial window in explicating the findings of this study in relation to the independent variables.

Second, collecting more than one type of independent measure of cognitive complexity is of utmost importance when investigating this construct. Cognitive complexity is not something determined by the researcher, but rather the learner, and how s/he perceives the task. Asking learners to report their anxiety and perceived difficulty of the task with questionnaires has so far been the only method used to gain insight into how learners perceive the task. This study was the first to collect learners’ time estimations of tasks, comparing them to real task time. Collecting learners’ retrospective estimations of time on task was a robust way to tap the construct of cognitive complexity, showing that complex conditions were delivered irrespective of mode. More methods are needed to measure how learners’ perceive and interact with the task in order to gain further insight into cognitive complexity and task design.

Lastly, this study contributes to applied statistical methods in SLA by using box plots to show the variance that took in the two complex groups. Box plots are a robust way to capture
outliers and visually demonstrate variation in performance. The Cognition Hypothesis predicts that more variance will occur in performance as tasks increase in cognitive complexity; the box plots visually confirmed this. More studies should use box plots to visually summarize SLA data, as has also recently been argued by Larson-Hall and Harrington (2009).

8.7.2.4 Pedagogical Implications

Though this study was conducted in a laboratory, its findings in regards to task design and interaction environment can be extended to the classroom, to include the online classroom. This study showed that tasks of differing degrees of cognitive complexity along Robinson’s resource-directing dimensions affect interaction and L2 development. For regular classrooms, the incorporation of more cognitively complex tasks, or ideally, a sequenced order of simple to complex tasks, may assist in promoting more interaction and learning. For online classrooms, carrying out tasks one-on-one in the CMC mode can result in L2 development, but more research is needed that extends TBLT research to CALL. It is hoped that this study generates more research on this important topic.

8.8 Limitations and Future Research

8.8.1 Limitations

This study did not go without limitations. First, care must be taken when interpreting the results on awareness. The purpose of this study was not to explore noticing and awareness, and participants were not asked to think aloud in this study, in either FTF or CMC. Therefore, while the think-aloud data (and CMC movements) that were obtained was very insightful and revealing, it may not be representative of the whole participant pool. The concurrent processing
data gathered here was used for a follow-up analysis to explicate the original findings. Clearly, more research is needed that explores the relationship between task complexity and awareness.

Another limitation may be with the assessments used to measure learning. Recall that the FTF and CMC production assessments contained ten items, while the multiple-choice reception test contained 15. The question remains as to whether 10 items on an assessment are indeed enough to measure change. Arguably, more than 10 items on the production tasks should been used, as well as other types of assessments. At the same time, the one-on-one interaction with the researcher, especially in the CMC+C condition, took so much time that the assessments used in this study were probably all that the learners could have handled. For example, on Day 3 of the present study, after carrying out the task, all participants had to do yet another story retell in FTF, and then another story retell in CMC (or vice versa), followed by a multiple-choice test. Indeed, coming up with assessments to properly measure L2 development is a difficult challenge for researchers, especially when claims want to be main in regards to a cross modal transfer (rendering two productive assessments in both FTF and CMC necessary).

The linguistic item chosen for this study also poses unique challenges. The Spanish past subjunctive is a difficult form to acquire from conversational interaction alone. Collentine’s recent review on the acquisition of the subjunctive reviewed a study on the frequency of the subjunctive in over two million words in the Corpus del español. The subjunctive, both present and past, comprised a little more than 7% of all verb forms in the analysis (Collentine, 2010). What’s more, the past subjunctive form chosen for this study occurs in the dependent clause position when a change of referent has occurred, and the primary clause must contain a verb expressing volition or emotion, as opposed to factual-perceitional denotations. It may therefore be difficult to extrapolate the findings from this study to other linguistic items given its difficulty
level. Related to this is the fact that critical items on the tasks and assessments were equally
divided according to the three classes of verbs that exist in Spanish: half were -ar verbs (where
the subjunctive morphology is -ara) and half were -er/-ir verbs (where the subjunctive
morphology is -iera). It is true that a majority of Spanish verbs come from the first -ar class. It is
therefore unknown whether or not learners were more tuned to the -ar verbs and respective
subjunctive morphology because of this. More research is needed to shed light on how Task
Complexity and modality affect L2 development in relation to other types of linguistic items.

**8.8.2 Future Research**

Future research should examine the acquisition of different linguistic and pragmatic
features and in different languages; the only two languages that have so far been explored in
relation to task complexity and learning are ESL/EFL and Spanish. More research is needed on
different task types as well, and how other task design features affect interaction and learning.
For example, the extent to which increases in complexity on research-dispersing features is
currently underexplored. In terms of resource-directing, most studies have looked at the +/- here-
and-now or +/-causal reasoning demands. The variables of spatial reasoning or perspective
taking have yet to be operationalized. The other two factors in Robinson’s Triadic Componential
Framework, Task Conditions and Task Difficulty, equally merit further exploration. These refer
to the interactive and learner factors, and there are some participant and learner-affective
variables that have never been researched, variables which arguably play a role in L2/FL
classrooms. These include +/-same proficiency, +/-same gender, +/-shared cultural knowledge,
and high or low willingness to communicate, an individual difference attested in SLA literature
(MacIntyre, Baker, Clement & Conrod, 2001). Other propensities that should be explored
include learner beliefs (c.g., Horwitz, 2987), learning strategies, and personality types (Dewaele & Furnham, 1999).

Only two modes for interaction were explored in this study. It would be interesting to see how other modes, such as video Skype (reported as used often by many of the participants in this study) play in task-based interaction. This is especially in light of the increasing demand for online language courses and computerized mediums to promote and teach SLA. That will be one of the next steps of the present study.

Also, deeper probing into the constructs of uptake and awareness is clearly warranted, especially as they relate to task complexity and modality. In this study, uptake did not mediate learning, and was hardly ever produced in the CMC. Why does uptake have significantly less of a presence in CMC? What does that mean for learning in this mode? In light of the results discussed above, one of the outcomes of this study will be to code uptake according to partial and exact, and to see if WMC plays a role in uptake type.

More studies are also needed that examine how different task designs mediate opportunities for the promotion of awareness, with mode as a potentially intervening variable. This is at present an unexplored area but clearly linked to robust designs that include concurrent or qualitative data in addition to quantitative data. In addition, the construct of WMC is just beginning to be explored in the field, and our understanding of the role it plays in SLA is still incipient. How WMC affects L2 development with different task types and different modes is undoubtedly in need of more empirical investigation. As has been argued in this study, however, a serious limitation has so far been the way in which studies have measured WMC. Researchers in the field of SLA must refer to findings, WM spans, administrative methods, and grading
schemes that are attested in the field of psychology for purposes of generalization and comparability.

Finally, as stated at the beginning of this dissertation, studies on CMC have thus far been tangential to task-based research. Investigations of what types of tasks work best in the CMC mode are crucial so that researchers and teachers who use CMC can maximize learning in this mode. This does not only apply to online classes; some FTF classes may chose to have external CMC-based tasks as additional out-of-class tasks for learners. It may also be the case that CMC has benefits over FTF for certain linguistic items; Skyes (2005), for example, reported that CMC was better than FTF in promoting pragmatic development.

8.9 Summary

This dissertation has extended the study of task design, and in particular, the cognitive complexity of tasks, to the CMC environment. The study has contributed to the literature showing that task complexity plays an important role in promoting interaction, language learning, and internalization of feedback, but poses some important questions about how different modalities (FTF versus CMC), as well as WMC and the production of uptake, modulate these outcomes. More research is needed to extend these findings, examining how task complexity differentially affords opportunities for SLA and the role it plays for current SLA theory.
Bibliography


Révész, A. (*in press, b*). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*.


Robinson, P. (2005b). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and


Internet to promote conversation and writing in intermediate level German language courses. *CALICO Journal, 15*, 129-43.


The Martínez family had a house cleaner, Srta. Gómez. Srta. Gómez always arrived on time, cleaned the house well, cooked good food, and cared for the children. The family really liked her and had trust in her. Srta. Gómez recently mentioned that the economic situation of her own family was bad, but overall, she was a positive and hard working person. Every morning, Sra. Martínez would wake up and make coffee as she waited for Srta. Gómez. When Srta. Gómez arrived, Sra. Martínez always smiled. She was happy that Srta. Gómez worked for them.
One day, after Srta. Gómez arrived, Sra. Martínez went upstairs to get her pearls for work. When she opened her jewelry box, she saw that her fine jewelry and the pearls were missing. Sra. Martínez asked the children if they moved the pearls. All of the kids said no. The family looked through the house, and even Sr. Martínez helped. Someone had definitely stolen Sra. Martínez’ pearls. Sr. Martínez said, “It must be Srta. Gómez. No one else has access to our room.” Sra. Martínez thought of how long Srta. Gómez had been with their family. “No,” she said, looking at the wall. She doubted that Srta. Gómez robbed them.
Sra. Martínez thought about who entered the house since yesterday morning. No one else came except for Srta. Gómez. Sra. Martínez started to panic. She ran downstairs to Srta. Gómez’ room. It angered her that Srta. Gómez would do such a thing. Sra. Martínez looked all throughout Srta. Gómez’s room. There wasn’t much to search though. She realized that Srta. Gómez did not have much compared to what the Martínez family had. She stopped and looked down at what she was doing. She felt sad that Srta. Gómez didn’t have any other option.
At the same time, the only possible person that stole the jewelry was Srta. Gómez. But if Srta. Gómez needed help, she could have asked. They all knew each other well enough for that. No … a person that stole was not somebody to be trusted. When Sr. Martínez arrived home to have lunch, Sra. Martínez informed him that he had to fire Srta. Gómez that afternoon. “Are you sure?” he asked. “¡Sí!” But when he left, Sra. Martínez started to cry. She felt very bad that they were going to fire Srta. Gómez.
After lunch, Sr. Martínez called Srita. Gómez outside. They talked for a long time, during which he told her she was fired. Srita. Gómez started to cry, and said that she’d never do such a thing. She was surprised that Sra. Martínez believed she stole the pearls. “But there is no way it could have been someone else,” Sr. Martínez explained. Srita. Gómez went slowly to her room to get her things. The room was a disaster! Srita. Gómez, left the house, shaking her head and then slamming the door. It bothered her that they looked through her things.
In the house, Sra. Martínez was looking at her jewelry box. She thought of the broken trust. “Five years!” she said aloud, throwing her hand up. She hated that they had trusted in someone so much. Just then, the youngest Martínez child came into her room. “Mamá, look what I found.” It was a small traveler’s bag with the pearls inside. Sra. Martínez packed it the week before for a trip, and completely forgot about it. “Oh my God, what have I done?” she said. She could not believe that she forgot about the jewels. She ran outside of the house, screaming, “Srta. Gómez!” She wanted Srta. Gómez to come back!
The Martínez family had a house cleaner, Srta. Gómez. Srta. Gómez always arrived on time, cleaned the house well, cooked good food, and cared for the children. The family really liked her and had trust in her. Srta. Gómez recently mentioned that the economic situation of her own family was bad, but overall, she was a positive and hard working person. Every morning, Sra. Martínez would wake up and make coffee as she waited for Srta. Gómez. When Srta. Gómez arrived, Sra. Martínez always smiled.

---

---
One day, after Srta. Gómez arrived, Sra. Martínez went upstairs to get her pearls for work. When she opened her jewelry box, she saw that much of her fine jewelry and the pearls were missing. Sra. Martínez asked the children if they moved the pearls. All of the kids said no. The family looked through the house, and even Sr. Martínez helped. Someone had definitely stolen Sra. Martínez’ pearls. Sr. Martínez said, “It must be Srta. Gómez. No one else has access to our room.” Sra. Martínez thought of how long Srta. Gómez had been with their family. “No,” she said, looking at the wall.
Sra. Martínez thought about who entered the house since yesterday morning. No one else came except for Srta. Gómez. Sra. Martínez started to panic. She ran downstairs to Srta. Gómez’ room. Sra. Martínez looked all throughout Srta. Gómez’s room. There wasn’t much to search though. She realized that Srta. Gómez did not have much compared to what the Martínez family had. She stopped and looked down at what she was doing.
At the same time, the only person possible that stole the jewelry was Srta. Gómez. But if Srta. Gómez needed help, she could have asked. They all knew each other well enough for that. No … a person that stole was not somebody to be trusted. When Sr. Martínez arrived home to have lunch, Sra. Martínez informed him that he had to fire Srta. Gómez that afternoon. “Are you sure?” he asked. “¡Sí!” But when he left, Sra. Martínez started to cry.
After lunch, Sr. Martínez called Srta. Gómez outside. They talked for a long time, during which he told her she was fired. Srta. Gómez started to cry, and said that she’d never do such a thing. “But there is no way it could have been someone else,” Sr. Martínez explained. Srta. Gómez went slowly to her room to get her things. The room was a disaster! Srta. Gómez left the house, shaking her head and then slamming the door. 😞
In the house, Sra. Martínez was looking at her jewelry box. She thought of the broken trust. “Five years!” she said aloud, throwing her hand up. Just then, the youngest Martínez child came into her room. “Mamá, look what I found.” It was a small traveler’s bag with the pearls inside. Sra. Martínez packed it the week before for a trip, and completely forgot about it. “Oh my God, what have I done?” she said. She ran outside of the house, screaming, “Srta. Gómez!”
Appendix C. Story and Comic Cards for Treatment Session 2: Intentional reasoning

Alejandro and Luis were best friends. They had much in common, especially their love for soccer. Every day they practiced together. To Alejandro, soccer was everything. For example, Alejandro often talked of going to the major leagues, both he and Luis together. “Luis, one day we’ll play for our country!” He wanted them to go far. Luis didn’t take it so seriously, rather, he just enjoyed the game. But to Alejandro, soccer was life.
One day a scout for the city team saw the boys playing. The man approached Luis and Alejandro and asked them if they wanted to play for the city of Maracaibo. “Who, us??” replied Luis, almost laughing. He was surprised that the man took notice of them. Alejandro glanced angrily at Luis. He couldn’t believe Luis responded that way. “Señor, that would be our dream. Let us try out!” he said to the scout. The scout invited them to the team practice the next day. The boys jumped up in the air and ran home. They were happy that the scout invited them.
The next day, Alejandro left for Luís’ house extra early. When he arrived there, Luís was still sleeping! Alejandro threw his soccer ball at him. It angered him that Luís did not take the invitation seriously. “Luís, we’re going to be late! Let’s go!” Luis grabbed his clothes and his soccer ball and they left together.
When they got to the city stadium, they saw a sign that said “INVITADOS ENTER HERE”. There was a line of about 15 men. Some soccer players laughed when they saw Alejandro and Luis go to the line. “Alejandro … are .. are you sure about this?” Luis stopped walking. “I don’t know about this man.” He doubted that they could do this. “Luis, what’s wrong with you!? This is our chance! And this is my dream!” “I don’t want to do this” said Luis. Alejandro grabbed Luis’ shirt and rolled his eyes. He hated that Luis give up his dreams like this.
All of the guys in the line were bigger than Alejandro and Luis. A couple more guys joked that Luis was too little to be at a professional practice. Luis looked down at the ground. He felt bad that they made fun of him. Finally, the formal introduction and session began. After that, they were all on the field practicing. Alejandro did excellent, but it was too rough for Luis. He kept getting pushed and hit with the ball. After the third time, he threw the ball and walked off the field. “Luis, where are you going?” said Alejandro. “This is your fault!” Luis said. Luis was mad that Alejandro insisted on this.
Alejandro watched as his best friend left the field. It looked like Luis had a bloody nose too. “Luis! Luis!!” But Luis continued walking. A coach yelled to Alejandro. “Let’s go!” Alejandro covered his face with his hands. He felt sad that Luis had a bad experience. He looked one last time at the stadium, but knew what he had to do. He left and ran to catch up with his friend.
Appendix D. Story and Comic Cards for Treatment Session 2: +intentional reasoning

Alejandro and Luis were best friends. They had much in common, especially their love for soccer. Every day they practiced together. To Alejandro, soccer was everything. For example, Alejandro often talked of going to the major leagues, both he and Luis together. “Luis, one day we’ll play for our country!” Luis didn’t take it so seriously, rather, he just enjoyed the game. But to Alejandro, soccer was life.
One day a scout for the city team saw the boys playing. The man approached Luis and Alejandro and asked them if they wanted to play for the city of Maracaibo. “Who, us??” replied Luis, almost laughing. Alejandro glanced angrily at Luis. “Señor, that would be our dream. Let us try out!” he said to the scout. The scout invited them to the team practice the next day. The boys jumped up in the air and ran home.
The next day, Alejandro left for Luis’ house extra early. When he arrived there, Luis was still sleeping! Alejandro threw his soccer ball at him. 😴“Luis, we’re going to be late! Let’s go!” Luis grabbed his clothes and his soccer ball and they left together.
When they got to the city stadium, they saw a sign that said “INVITADOS ENTER HERE”. There was a line of about 15 men. Some soccer players laughed when they saw Alejandro and Luis go to the line. “Alejandro … are .. are you sure about this?” Luis stopped walking. “I don’t know about this man.” “Luis, what’s wrong with you!? This is our chance! And this is my dream!” “I don’t want to do this” said Luis. Alejandro grabbed Luis’ shirt and rolled his eyes.
All of the guys in the line were bigger than Alejandro and Luis. A couple more guys joked that Luis was too little to be at a professional practice. Luis looked down at the ground. Finally, the formal introduction and session began. After that, they were all on the field practicing. Alejandro did excellent, but it was too rough for Luis. He kept getting pushed and hit with the ball. After the third time, he threw the ball and walked off the field. “Luis, where are you going?” said Alejandro. “This is your fault!” said Luis.
Alejandro watched as his best friend left the field. It looked like Luis had a bloody nose too. “Luis! Luis!!” But Luis continued walking. A coach yelled to Alejandro. “Let’s go!” Alejandro covered his face with his hands. He looked one last time at the stadium, but knew what he had to do. He left and ran to catch up with his friend.
Appendix E. Cards for CMC Productive Task A

Computer-Mediated Communication

Re-telling a story in Spanish

For this task you will read a brief story in English, and your job is to re-tell the story to the researcher in Spanish via iChat. During the task, the researcher will give you some words to assist you in telling the story. If you have any questions, just ask the researcher on iChat. ¡Buena suerte!

PART 1
Marta and José were a married couple that lived in Florida. One night, José gave Marta a surprise. He bought her earrings and made reservations at an expensive restaurant. ¡Marta was so surprised that José planned such a spectacular date! It was cold that night, so Marta brought her special sweater. The restaurant was excellent and was next to the sea. Marta and José ordered the special, mahi mahi, and drank wine. Marta could not believe that they had so much fun. They concluded their date by walking to the house. Marta was very happy that they went out together.

PART 2
When they arrived home, Marta realized that she left her sweater at the restaurant. She wanted them to go back to look for the sweater. As it was 9:30 p.m., José said no. He was annoyed that she felt it was necessary to do this. Nevertheless, Marta doubted that she could find it the next day. She went to the restaurant without José. José felt bad that she went alone, so decided to accompany her.

PART 3
When they arrived to the restaurant, the kitchen was closed, but the workers were still there cleaning. Marta asked if they could look under the table. The sweater was not there. Marta looked everywhere, but it was gone. Marta cried. She hated that they did not find it. José was so confused! Marta looked at him and left, still crying. She was sad that he did not understand. On the other hand, it angered José that Marta behaved this way. It was just a sweater. He scratched his head and followed his wife home.
Appendix F. Cards for CMC Productive Task B

Computer-Mediated Communication

Re-telling a story in Spanish

For this task you will read a brief story in English, and your job is to re-tell the story to the researcher in Spanish via iChat. During the task, the researcher will give you some words to assist you in telling the story. If you have any questions, just ask the researcher on iChat. ¡Buena suerte!

**PART 1**
Susana and Carla were roommates in college. Carla was also from Texas, and had a similar family. It was perfect because they had similar schedules and lifestyles. Susana liked Carla a lot, and wanted them to be friends for life. The next week, the girls had midterm exams, but Susan wasn’t worried. One the other hand, Carla hated that they had exams. She was so stressed! Susana tried to help Carla. She felt bad that Carla internalized her stress so much.

**PART 2**
Fortunately, the girls had the same chemistry class. Every night they read one chapter together. Susana was content that they studied together this way. Carla, however, was not so confident. She doubted that she could pass the exam. On the day of the exam, Susana and Carla sat together. Ten minutes later, Susana saw Carla look at a piece of paper. Susana was shocked that Carla used notes! She could not believe that Carla was a cheater. It made her angry that Carla lied. Then, in that moment, the chemistry professor arrived at their desks.

**PART 3**
“Show me what you have in your hand,” the professor said. He was annoyed that a student deceived him. Susana’s heart almost stopped. Shaking, Carla gave him the paper. “Come with me.” All of the students watched as Carla followed the professor out of the classroom. Susana felt so sad that everyone watched her in that way. What was going to happen?
Appendix G. Cards for CMC Productive Task C

Computer-Mediated Communication

Re-telling a story in Spanish

For this task you will read a brief story in English, and your job is to re-tell the story to the researcher in Spanish via iChat. During the task, the researcher will give you some words to assist you in telling the story. If you have any questions, just ask the researcher on iChat. ¡Buena suerte!

PART 1
The Badell family had a large farm in Maracaibo, Venezuela. Sr. Badell was very content that all of his family lived on the same land. Everyone worked hard each day. Everyone except for the oldest daughter, Marisa. She hated that they were on a farm! It angered her that she had to work each day with the animals. She wanted to be with her friends instead and do normal teenager activities.

PART 2
The next month was an important harvest date. Sr. Badell wanted all of the family to help. He gave everyone a task and they were all excited. Marisa, however, told her father that she had more important things to do. Sr. Badell was surprised that his daughter said that. He couldn’t believe that Marisa was so selfish. But fine, he said. “Do what you feel is best Marisa.” Sr. Badell went to prepare the machines and feed the animals. He had much work to do. Marisa doubted that her father allow her to leave, but went anyway.

PART 3
On the first day of the harvest, everyone was working. Mr. Badell even hired temporary farm workers. The family was annoyed that Marisa did not participate. “Why do we have to work, and she doesn’t?” said the youngest sister. In secret, Mr. Badell was sad that Marisa did not desire to help.

Two hours later, Marisa arrived. “Dad, I’m here! I was at the commercial center and realized that family – and our business – was more important. Please forgive me.” Marisa felt bad that her actions hurt her dad. She said that she learned a lesson – that family comes first.
Appendix H. Cards for FTF Productive Task A

Task: Retelling a story in Spanish

For this task, you will read a short story in English and then will be asked to retell it to the researcher as best as you can in Spanish. Sometimes the researcher will help you with vocabulary items to help you along. ¡Buena suerte!

**Part 1**
Ana and Carlos were a couple and had been together for two years. They were fun and did many activities together. To Ana, Carlos seemed like the perfect man. He was affectionate and caring, and told Ana things that made her feel good about herself. She knew he was the person for her, and she was happy that they were together. In fact, Ana wanted Carlos and her to be a couple forever.

**Part 2**
The only thing was, Carlos seemed to be friendly with all women. He gave lots of complements to all of his female friends, sometimes in front of Ana. She told herself that this was just his personality, and accepted that he was nice with everyone. Until one evening, when she saw Carlos’s phone. While Carlos was in the shower, he received a text message. It was a message from another girl, asking when she could meet with Carlos again. The girl’s name was Marta. Ana was surprised that Carlos received messages from another woman. No, actually, she couldn’t believe that Carlos and Marta communicated like this! Ana didn’t know Carlos went out with another woman. It angered her that he cheated on her. Ana looked more at Carlos’s phone. She read more text messages that Carlos had sent to Marta. She hated that Carlos sent messages to this other person. Some of these messages he sent to Ana too!

**Part 3**
Ana stood up and threw Carlos’s phone against the wall. She doubted that Carlos did this, but the evidence was clear. She grabbed some paper and a pen, and wrote Carlos a note that said “Have a great time with MARTA.” She left the apartment. As she walked outside to her car, Ana started to cry. She turned around to go back home, but then stopped herself. She felt bad that she broke Carlos’ phone. But at the same time, she was annoyed that Carlos ended their relationship in this way. At that moment, Carlos tried to call her on her cell phone, but Ana did not answer. She was sad that the relationship could not continue.
Appendix I. Cards for FTF Productive Task B

Task: Retelling a story in Spanish

For this task, you will read a short story in English and then will be asked to retell it to the researcher as best as you can in Spanish. Sometimes the researcher will help you with vocabulary items to help you along. ¡Buena suerte!

Part 1
Raquel García played the piano very well. She participated in some competitions and always earned the best prize. Mrs. García was so proud of her daughter, and told everyone that Raquel was the most talented player ever. Raquel hated that her mother did this. She didn’t want her mother to talk this way. But at the same time, she knew that her mom was proud, so didn’t say anything. Still, while she enjoyed playing the piano, she was annoyed that her mother made it into such a big deal.

Part 2
One weekend Raquel had a piano competition. She was happy that all of her friends were there. However, Mrs. García seemed more excited than she did. She reminded Raquel each night: “You have to practice!” Raquel felt bad that her friends heard this. On Saturday they went to the competition. Raquel saw her friends and started talking. “Raquel, you have to go prepare!” said Mrs. García. It angered Raquel that her mother was so obsessed with winning.

Part 3
Mrs. García saw that Raquel was wasting time. She could not believe that Raquel passed time with her friends instead of practicing. Suddenly, the competition began. Raquel started to play. She did a fantastic job, but another girl won. Mrs. García felt sad that Raquel did not win. “But, my daughter is the best!” Raquel heard her mother’s conversation. She was shocked that her mother bragged in this way. Why didn’t her mom act normal, like other moms? “Mom, I need to talk to you. Let’s go outside.” Mrs. García didn’t understand, but went. Raquel knew that some things had to change. However, she doubted that her mother understand.
Appendix I. Cards for FTF Productive Task C

Task: Retelling a story in Spanish

For this task, you will read a short story in English and then will be asked to retell it to the researcher as best as you can in Spanish. Sometimes the researcher will help you with vocabulary items to help you along. ¡Buena suerte!

---

**Part 1**

Samuel and Rafael were friends at school. They both loved to play basketball. Rafael was happy that they were friends, but on the court, Sam was a different person. He was such a show off! Rafael hated that Sam did this. For example, Sam never passed the ball to the other players. Sam also yelled at Rafael in front of everyone when he made a bad pass. **Rafael was annoyed that Sam called him in front of everyone.** But Rafael just continued to play and ignored Sam during a game.

---

**Part 2**

One day, Sam went too far. During a game against the school rival, Rafael missed a basket. **It angered Sam that Rafael missed it.** He went crazy! He pushed Rafael on the court and yelled --- all in front of everyone. Rafael could not believe that Sam changed that way! **Rafael wanted that Sam know that he could not keep doing this.** He threw the basketball really hard at Sam’s head. **Sam was surprised that Rafael reacted!** Then, it happened – the boys got in to a physical fight, right in the middle of the court.

---

**Part 3**

Three coaches had to stop Sam and Rafael’s fight. They left the court and were told that they were suspended. They had a meeting with the school director. Rafael had to explain why he threw the ball at Sam. **He felt bad that they caused a fight during the game.** At the same time, **he felt sad that the administrators suspended them.** Rafael admitted his involvement too. He apologized to Sam for his actions. Initially, the **director doubted that the boys admit their mistake.** But because they did, he said he would reconsider their suspension. Sam and Rafael were relieved. They promised to never fight again.
Appendix J. Background and Biodata Questionnaire

**Background Questionnaire**

Name: _______________________________  NetID: __________  [Research code: ___________________]

Gender: _____Male _____Female  Age: ______

First Language(s): ______________________

Have you studied any other languages besides Spanish?  : ______________________

How old were you when you first started learning Spanish? ________

Where have you studied Spanish before?  For how long?

<table>
<thead>
<tr>
<th>Year</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kindergarten</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1) What are your reasons for studying Spanish here at Georgetown? (examples: fulfill the language requirement, to learn another language, I want to study abroad, etc.)

2) Have you studied Spanish in any other way? (For example, private tutor, nanny, language school, etc.) If so, in what context and for how long?

3) Have you ever visited a Spanish-speaking country before? If so, for how long? __________________________

4) What Spanish course(s) are you taking right now? Do you plan on continuing with Spanish next semester? If so, which classes?

**Spanish Usage**

5) How many hours per week would you say you do the following activities in Spanish?

<table>
<thead>
<tr>
<th>Activity</th>
<th>0</th>
<th>1-2</th>
<th>3-4</th>
<th>5-6</th>
<th>7-8</th>
<th>9 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speak Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read in Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listen to Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6) In what contexts do you practice speaking Spanish?

<table>
<thead>
<tr>
<th>Context</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>In Spanish class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With family</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With friends</td>
<td></td>
<td>Other?</td>
</tr>
</tbody>
</table>

362
Technology Usage

7) How many hours per week would you say you use the computer to do the following activities?

<table>
<thead>
<tr>
<th>Activity</th>
<th>0</th>
<th>1-2</th>
<th>3-4</th>
<th>5-6</th>
<th>7-8</th>
<th>9 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet Use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook/ MySpace (social sites)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chat Online</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Chat type (please circle all that apply): Instant Messenger / iChat / Social site chat / Google chat / other _________

<table>
<thead>
<tr>
<th>Activity</th>
<th>0</th>
<th>1-2</th>
<th>3-4</th>
<th>5-6</th>
<th>7-8</th>
<th>9 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Video Chat Online</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Video type (please circle all that apply): OoVoo, sending videos through Facebook, other _________

<table>
<thead>
<tr>
<th>Activity</th>
<th>0</th>
<th>1-2</th>
<th>3-4</th>
<th>5-6</th>
<th>7-8</th>
<th>9 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listen to Music</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watch videos, movies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other?:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8) Have you have ever used the following technologies in Spanish (or another language)? If so, which ones and how often?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Yes/No</th>
<th>Other foreign language</th>
<th>How often? (i.e., once, a couple of times, 4 hours a week, etc.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Networking</td>
<td>Y/N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text Messaging</td>
<td>Y/N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet browsing</td>
<td>Y/N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Watch videos &amp; movies</td>
<td>Y/N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chat online</td>
<td>Y/N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Video chat online</td>
<td>Y/N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any others?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9) Do you have a cell phone or other kind of mobile device? Y/N  device: ____________

Do you use your cell or mobile device to send texts messages and to access email and/or internet?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Yes/No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text messaging</td>
<td>Y/N</td>
</tr>
<tr>
<td>Check/write email</td>
<td>Y/N</td>
</tr>
<tr>
<td>Check internet</td>
<td>Y/N</td>
</tr>
<tr>
<td>Other?:</td>
<td></td>
</tr>
</tbody>
</table>

If so, do you ever do so in Spanish or any other foreign language? Yes/ No

(Language other than Spanish?: __________________________)

10) Do you think technology is important for learning another language? What do you think are some of the pros and/or cons of technology in language learning? __________________________________________________________

                                                                                      __________________________________________________________

11) Have you used computers or other technologies to help you learn Spanish? __________________________________________________________
12) Does your teacher use any kind of technology in your Spanish class? What do you think about it?

________________________________________________________________________________________

13) What is your opinion on the use of computerized chat or video chat to practice in Spanish?

________________________________________________________________________________________

14) What would you think of an online Spanish or foreign language course that utilized chat or video chat?

________________________________________________________________________________________

15) What is your opinion about speaking the language in person in order to practice Spanish? How does this relate to your opinion about any kind of technology use to practice Spanish?

________________________________________________________________________________________

________________________________________________________________________________________

16) Lastly, is there any kind of activity that you would like to see incorporated (more) into your language class here at Georgetown, technology related or not? Yes/ No and why?

________________________________________________________________________________________

________________________________________________________________________________________

________________________________________________________________________________________

Thank you for your feedback. 😊
Appendix K. Anxiety and Perceived Difficulty Questionnaire

Name: ___________________________  NetID: ________________  [Research Code: ____________ ]

Please tell me how much you agree or disagree with the following statements by circling a number from 1 to 6:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Slightly disagree</th>
<th>Slightly agree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

1. This task was difficult.  
2. I felt like I didn’t have enough time to think before I had to respond.  
3. I felt confident in my ability to learn new things in Spanish.  
4. The task I had to do with my partner wasn’t hard.  
5. I was relaxed and comfortable completing the task.  
6. I felt rushed during the task.  
7. It was tough to communicate the main points of the story in Spanish.  
8. This task was stressful for me.  
9. I enjoyed communicating with my partner during this task.  
10. What my partner asked me to do in this task wasn’t so difficult.  
11. Sometimes I struggled during this task.  
12. This task made me feel less anxious than I feel in class.  
13. The goals of the story I had to tell were easy.  
14. It did not bother me when I did not understand everything my partner was saying.  
15. I felt tense having to communicate with my partner.
Appendix L. Exit Questionnaire

Final Questionnaire

Please complete the following questionnaire and let the researcher know when you are finished. Thank you!

1) What do you think were the goals of this study?
__________________________________________________________________________________________
__________________________________________________________________________________________

2) Did you learn anything from this study?
__________________________________________________________________________________________
__________________________________________________________________________________________

3) Did you focus on any specific grammatical form in this study?
__________________________________________________________________________________________
__________________________________________________________________________________________

4) Did you look up any forms or information during the weeks you partook in the study outside of the sessions? If so, which ones?
__________________________________________________________________________________________
__________________________________________________________________________________________

5) What is your opinion about Spanish conversation practice in person and via chat on the computer?
__________________________________________________________________________________________
__________________________________________________________________________________________

6) Do you think one platform (in person versus the computer) is harder to communicate in than the other? Or, do you think they’re about the same? Why?
__________________________________________________________________________________________
__________________________________________________________________________________________

7) In your opinion, might one platform (computer vs. face-to-face) for practicing language conversation have more benefits over the other? Or are they the same? Why or why not?
__________________________________________________________________________________________
__________________________________________________________________________________________

8) Do you have any other comments for the researcher?
__________________________________________________________________________________________

Thank you for your participation!
Appendix M. Informed Consent Form

GEORGETOWN UNIVERSITY
CONSENT TO PARTICIPATE IN RESEARCH

PROJECT TITLE: Task complexity, the Cognition Hypothesis, and interaction in CMC and FTF environments

PRINCIPAL INVESTIGATOR: Melissa Baralt
PHONE: (202) 687-5302
E-mail: mlb65@georgetown.edu

SPONSOR: Georgetown University, Department of Spanish and Portuguese

The Georgetown University Institutional Review Board (IRB) has approved this research project. For information on your rights as a research subject, call the Institutional Review Board office at 202-687-1506.

INTRODUCTION
You are invited to consider participating in this 4-part research study to investigate how language is learned. This form will describe the purpose and nature of the research, its possible risks and benefits, other options available to you, and your rights as a participant in the study. Participation in this study is entirely voluntary. Your decision not to participate will in no way negatively affect your grade. You have the right to leave the study at any time. Should you decide to leave the study, please let the researcher know (mlb65@georgetown.edu). If you decide to participate, please sign and date the last line of this form.

WHY IS THIS RESEARCH STUDY BEING DONE AND WHAT IS INVOLVED IN THE STUDY?
In this research study, I am interested in learning more about how students learn Spanish. As part of the study, I will ask you to carry out tests of working memory, as it has been suggested to affect how humans learn a language. I will also ask you to carry out two decision-making tasks with me in Spanish, as well as one pretest and two posttests. When you do the task in Spanish, you will be asked to talk with me in Spanish in person and to talk with me in Spanish via messenger on the computer. This study will involve four sessions (approximate times below).

<table>
<thead>
<tr>
<th>Session</th>
<th>Procedure</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-test</td>
<td>60 minutes</td>
</tr>
<tr>
<td></td>
<td>Working Memory Tests</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Experimental Session 1: Task in Spanish</td>
<td>15-30 minutes</td>
</tr>
<tr>
<td>3</td>
<td>Experimental Session 2: Task in Spanish Posttest</td>
<td>30-60 minutes</td>
</tr>
<tr>
<td>4</td>
<td>Delayed Posttest</td>
<td>15-30 minutes</td>
</tr>
</tbody>
</table>
HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?
About 200 people will take part in this study.

HOW LONG WILL I BE IN THE STUDY?
You will be in the study for approximately four to five weeks.

As participation is voluntary, you may stop participating at any time. If you do decide to stop participating, you are encouraged to let the researcher know.

WHAT ARE THE RISKS OF THE STUDY?
There are no risks involved with this study.

ARE THERE BENEFITS TO TAKING PART IN THE STUDY?
You will not experience any personal, direct benefits as a result of participating in this study. However, others may benefit in the future from the information we obtain in this study.

WHO CAN PARTICIPATE IN THE STUDY?
This study is designed for intermediate-level students of the Spanish language.

WHAT ABOUT CONFIDENTIALITY?
Your name will NOT be used when data from this study are published.

Every effort will be made to keep your research records and other personal information confidential. However, we cannot guarantee absolute confidentiality. It is possible that individuals from the Georgetown University IRB, other Georgetown University offices, Federal regulatory agencies may look at records related to this study, both to assure quality control and to analyze data. Your name and any material that could identify you will remain confidential except as may be required by law.

We will take the following steps to keep information about you confidential, and to protect it from unauthorized disclosure, tampering, or damage: participants’ names will not be disclosed – all participants will be assigned a number in this study. All data (your tests, recordings of the conversational tasks, working memory results) will be kept on a password-protected computer or in a locked box in the researcher’s home. Only the principal investigator will have access to the data.

COMPENSATION FOR YOUR PARTICIPATION
Upon completion of all four parts of this study, you will be awarded six (6) extra credit points to be added on to any one quiz in your Spanish class this semester. If you choose not to participate but would like to earn the 6 extra points, you may write a 5-page, double-spaced paper in Spanish, researching one of the cultural components you have talked about in class. If you would like to choose this option, please contact the researcher for details.

WHAT ARE MY RIGHTS AS A RESEARCH PARTICIPANT?
Participation in this study is entirely voluntary at all times. You have the right not to participate at all or to leave the study at any time. Deciding not to participate or choosing to leave the study will not result in any penalty or loss of benefits to which you are entitled, and it will not harm your relationship with Georgetown University or any of its employees.

**WHOM DO I CONTACT IF I HAVE QUESTIONS OR PROBLEMS?**

Call Melissa Baralt at 202-687-5302 during regular business hours if you have questions about the study, any problems, unexpected physical or psychological discomforts, etc. Please also call Mrs. Baralt if you are interested in knowing what the results are of the study, as well as your results on the working memory and aptitude tests (recall that taking the aptitude test is optional in this study).

Call the Georgetown University IRB Office at 202-687-1506 with any questions about your rights as a research participant.

**Statement of Person Obtaining Informed Consent**

I have fully explained this study to the subject. I have discussed the study’s purpose, its procedures, the possible risks and benefits, the alternatives to participation, and the voluntary nature of participation. I have invited the subject to ask questions and have answered any questions that the subject has asked.

____________________________

Signature of Person Obtaining Informed Consent

____________________________

Date

**Consent of Subject**

I have read the information provided in this Informed Consent Document.

My questions were answered to my satisfaction.

I voluntarily agree to participate in this study.

____________________________

Signature of Subject

____________________________

Date

____________________________

Printed Name of Subject
Appendix N. Instructions for the Treatment Task: -intentional reasoning

INSTRUCTIONS

You are about to read a story in English. Your task is to retell the story as best as you can to the researcher in Spanish. The researcher will help you - you and she will work to reconstruct the story together. Be sure to tell the story in the past tense. To make the task a little easier, you will read the story in sections so that you do not have to remember so much. After each section, you will be given picture prompts in the form of a comic narrative strip. The comic strips serve as visual guides to help you retell the story.

If you need any help or want to know how to say something, just ask the researcher during the conversation. Remember to tell everything in the past!

¡Diviértete y buena suerte! 😊
Appendix O. Instructions for the Treatment Task: +intentional reasoning

INSTRUCTIONS

You are about to read a story in English. Your task is to retell the story as best as you can to the researcher in Spanish. The researcher will help you - you and she will work to reconstruct the story together. Be sure to tell the story in the past tense. To make the task a little easier, you will read the story in sections so that you do not have to remember so much. After each section, you will be given picture prompts in the form of a comic narrative strip. The comic strips serve as visual guides to help you retell the story.

After some of the actions done by characters (marked with a (1)), you will see a yellow empty thought bubble in the comic strip (marked with a “2”), like this:

Whenever you see this thought bubble, your job is to think about and say the intentional reason that caused him/her to do the action that just took place (1). In other words, besides communicating the events of the story in the past, you are asked to reflect upon the emotional and mental states of the character that caused him or her to do something. These bubbles will also be in the English story, showing where you need to reflect on the personal emotions/thoughts that caused the previous action.

Below, is an example for you in English:

A character throws a ball at his big brother. (1) (action)

⇒ (think of the intentional reason/mental state that caused him to do this action)

*Juan doesn’t like that his brother receives all of the attention.* (2)

This is just an example – your job is to come up with your own reasoning of the mental states behind people’s actions. If you need any help or want to know how to say something, just ask the researcher during the conversation. ¡Diviértete y buena suerte! ☺