INTERACTION IN SYNCHRONOUS COMPUTER-MEDIATED COMMUNICATION: THE EFFECTS OF INTERLOCUTOR, TASK, AND STATE ANXIETY

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 Spanish and Portuguese

By

Christine Bistline-Bonilla, M.S.

Washington, D.C.
June 18, 2020
ABSTRACT

There is an increasingly large body of research that has addressed how interaction via Synchronous Computer-Mediated Communication (SCMC) may support second language (L2) development (see Ziegler, 2016 for a review). Various empirical studies in SCMC have demonstrated that type of interlocutor (e.g., Liu, 2017) and type of task (e.g., Blake, 2000; Yilmaz & Granena, 2010) can significantly impact the amount of negotiation that occurs during interaction. Nevertheless, the number of empirical studies is limited and they vary widely in their methodology. In addition, other researchers have suggested that one of the possible advantages of SCMC over other modes of communication is its potential to reduce L2 learner anxiety (e.g., Abrams, 2003).

The present study aims to contribute to instructed second language acquisition (ISLA) research by employing two types of tasks along with three types of interlocutors in SCMC to address whether 1) production of language-related episodes (LREs) on task, 2) amount of talk on task, and 3) L2 lexical recognition and production scores over time are related to type of interlocutor, type of task, and learners’ state anxiety. It included 82 adult intermediate L2 learners of Spanish who collaborated with either a peer, professor, or native speaker via Zoom instant messaging to complete 1) an information gap task and 2) a decision-making task. After completing each experimental task, they completed a state anxiety questionnaire adapted from Baralt and Gurzynski-Weiss (2011).
Statistical analyses revealed that a combination of Type of interlocutor and Type of task differentially affected the production of LREs and the amount of talk on task. In terms of participants’ lexical recognition and production performance over time, the analyses indicated that while Type of interlocutor played a limited role, participants were significantly more accurate in their recognition and written production of the decision-making task target items. It was also found that interlocutor type and task type were not predictors of participants’ overall state anxiety. Ultimately, these findings provide a better understanding of the roles of interlocutor status and the type of task in text-based SCMC and their impact on subsequent recognition and written production of L2 lexical forms.
DEDICATION

In the words of my late father, Dr. William G. Bistline, “this work is dedicated to my family who endured and persevered until it was done, done.”
ACKNOWLEDGEMENTS

Thank you to my committee members, my husband, my children, my family, my friends, my colleagues, my professors, my students, and everyone else who has been a part of my journey up to this point. I would not be here without all of you.

To Ron, there are not enough words to say how grateful I am for all your personal, professional, and academic guidance over the years. Thank you for taking me under your wing from the very start, and for always believing in me. Thank you for all your support, guidance, and patience with me when I did not even have patience for myself. You helped me learn how to become a teacher, a researcher, and even a language program director. You presented me with opportunities to present and publish when I didn’t know where to look. You wrote my letters of recommendation and counseled me on my job prospects. You told me it was OK to stay an extra year, and even though at the time that made me feel like a failure, things worked out even better than I could have imagined. Thank you for helping me feel confident in my decisions and learning to trust myself. More than anything, I will never forget how you always looked out for me, my husband, and my children during my time at Georgetown. Thank you for that. Thank you for all of it.

To Cristina, thank you for agreeing to serve on my committee, and for being such a wonderful source of guidance and support over these past six years. I have learned so much from you as both a student and a researcher. Thank you for always pushing me to do my best work both in and out of the classroom. I am also incredibly grateful for all the letters of recommendation you wrote me for my job search, and thank you for always being so available for any question or concern I had. Finally, I want to extend a big thank you for always having my
back when it came to my teaching schedule – I never would have been able to be home with Gael for as long as I was without your help. Thank you for making such a difference in our lives.

To Paul, where do I even begin to thank you? I really would not be here writing this dissertation if it weren’t for you. Even though I hadn’t been your student in years, you were still willing to speak with me for hours over the phone about my options for grad school. Georgetown had not even been on my radar when you suggested that I apply. Thank you for that. You changed my whole life, my family’s whole life. I never would have even considered a career in linguistics if it weren’t for you. I still remember my very first undergraduate Spanish class that I took with you, a new professor at Temple who seemed incredibly passionate about what he was teaching, which made me want to feel passionate about it too. You were always so understanding and patient with me when I was having a hard time personally or didn’t put forth my best effort in class. It was incredibly easy for me to talk to you about anything at all. Thank you for always listening, and always understanding, even during my time in graduate school. Your kindness, generosity, and empathy will never be forgotten. Finally, your detailed feedback on my dissertation was truly invaluable and I am forever grateful for the time you invested in it and in me.

To all my other wonderful professors at Georgetown: Dr. Héctor Campos, Dr. Elena Herburger, Dr. Tom Walsh, Dr. Alfonso Morales-Front, Dr. Lourdes Ortega, Dr. Jennifer Sclafani, Dr. Allison Mackey, Dr. Luke Plonsky, Dr. Nic Subtirelu, and Dr. Rusan Chen. Simply put, thank you for teaching me about linguistics and sharing your passion. I really knew very little when I started at Georgetown, and it’s hard to believe how far I’ve come. I might even finally be ready to call myself a linguist. I’m looking forward to passing on the knowledge and experience I’ve acquired from you all through both my teaching and research. Thank you also to
Dr. Molly Borowitz, who invested a great deal of time into helping me prepare my job materials and guiding me through the job market. I could not have achieved so much without your input and support.

To Sakol Suethanapornkul and Nick Pandza, thank you for all your invaluable help with my stats. I could not have finished this dissertation without it. To all my friends and colleagues from Georgetown: Ariel Zach, Pablo Camus, Marisa Filgueras-Gomez, Celia Chomón Zamora, Will Travers, Kevin Murphy, Tim McCormick, Annie Ornelles, Jeong (Eva) Mun, Tris Faulkner, Leah Adelson, Iván Espinosa Orozco, Felipe Toro, Alfredo Poggi, Jafte Robles, Sara Ramirez, Hortensia Gutierrez, Gorka Basterretxea, Leonardo Carvajal, Tyler Bergin, Amanda Ulldemolins, Nohora Arrieta, Ross Karlan, Emma Tierknee, Cecily Raynor, Janire Zalbidea, Colleen Moorman, Yoel Castillo, Zaya Rustamova, and Angela Donate. Gracias por ser parte de este camino y por todos los recuerdos bonitos. Thank you for all your support and guidance.

These have been some of the best years of my life. Thanks for partying with me, Manny, Gael, and Luna. I’ve loved getting to know all of you and spending time with you. Los voy a extrañar mucho y siento mucho que no nos hayamos podido despedir como se debía (stupid COVID-19!).

To Sophie Heller, I am so glad to have met you and honored to have become your friend during our time together at Georgetown. Thank you for always being a wonderful friend, lending a supportive ear, and for always helping me see the other side of a difficult situation. Thank you so much for watching Gael during my dissertation proposal defense, among the countless other times you helped me and Manny out in a pinch. Thank you for loving my babies as if they were your own. I am going to miss all our lunches and our fun nights out together. I can’t wait to bowl and drink craisin water (HAHA) with you again. I am so proud of you for everything you’ve overcome. You are a strong, caring, intelligent, and beautiful person and friend. Never forget it!
To Meagan Driver, thank you for your unwavering support over the years and for being one of the few people who I could always trust to keep it real with me. You’re such a genuine person with a beautiful heart and you always worry about everyone else over yourself. I could not have survived the PhD, particularly this last year, without your support. You were always there to commiserate with me and to ease my anxiety about a million and one things. Manny and I are indebted to you for all your help with watching Gael over the years, especially when our Lunita venía en camino. I know you had a lot of other commitments (like dissertation data collection!!) but you always made time to check in on me and were always ready to run right over when we needed you. Friends like you are few and far between in this life. Te quiero mucho, bichi!

To Gabriela DeRobles, mi comadre. I adore you and could not have done any of this without you. Thank you for being such an awesome and supportive friend (and colleague!) over the past six years. I am so lucky to have you as a friend, and Gael is so lucky to have you as his madrina. I would not have wanted to collect data, present research, and write chapters with anyone but you lol. Thank you for making my experience at Georgetown so bearable, for always making me laugh, and for being a genuine friend to me and having my back when others didn’t have my best interests at heart. Thank you for always responding to my insanely long text rants and entertaining all of my crazy dissertation ideas before they came to fruition. I am so looking forward to us getting out to Colorado and being so close to you! Te quiero mucho, comadre. Gracias por todo tu apoyo.

To Ariel Kersky, my very first and best friend at Georgetown! As someone who can be incredibly socially awkward, you pulled me out of my shell on our very first day of orientation together back in August 2014. From that very moment, we spent so much of the next two years
together studying, reading, laughing, crying, talking, and most importantly, EATING (DQ? Poptarts? Donuts? Chilaquiles? You name it!). I am so grateful to have lived so many amazing experiences with you during this PhD – all the runs to DQ, the dinners, the bars, the parties at my house (where you always came first and left last bc you’re the besttttt!), our joint birthday party, our graduation from the Master’s, our trip to Miami, our day at the orchard, our night together in Juarez after Manny got his green card (DQ? DQ!), and many more. Thank you for always being my #1 cheerleader and ALWAYS having my back in every situation. You've always been able to put words to my thoughts and understand me better than almost anyone. I am so grateful for all the times you watched Gael for me and Manny and gave us a much-needed date night (Let’s never forget #gradschoolprom). It was so hard for me when you left DC and the US but I am really happy that we have been able to maintain our close friendship, and I know we will continue to do so for many years to come. Love you, Tía Ariel!

A la familia Bonilla: Maribel, José, Karina, Gabriel, Édgar, y Ulises. No me podría haber tocado mejores suegros o mejores cuñados. Cuando estoy con ustedes, siempre me siento bienvenida y como si hubiera sido parte de la familia desde siempre. Gracias por siempre estar y por siempre apoyarme y entenderme en todo momento. No habría podido lograr esto sin su apoyo y sus consejos. A mis suegros, muchas gracias por criar al hombre de mis sueños y hacerlo la persona que es. Sé que Emmanuel es así gracias a su ejemplo. Lo siento por robármelo jiji.

To my brothers, Paul and Bill, and my sister, Laura, thank you for always supporting me and being there for all of my school events and activities growing up, and for continuing to lend me a supportive ear even into adulthood. To my mother, Melanie, thank you for being an example of strength and perseverance and for always believing in me. You always told me that I
could do whatever I set my mind to, and I think I heard that from you so often growing up that I started to believe it. When I look back on everything Manny and I have overcome over the past six years, I know that this mentality probably saved us in a lot of tough situations, and helped me push through this last and hardest part of finishing the dissertation, which felt nearly impossible during the height of a pandemic. Thank you for pushing me to always do my best in school, for encouraging my curiosity, and for never saying no when I wanted to study abroad even though you didn't have the money for it. Most importantly, thank you for all your help with surviving and completing this dissertation while having two babies over the past three years. Manny and I are indebted to you for coming down to DC to watch Gael on so many occasions, among the many ways you generously helped us survive this experience.

A mis niños, Gael y Luna. Gracias por darme el honor de ser mamá. Más que nada, mis años en Georgetown han sido especiales (de hecho, excepcionales!!!) porque llegaron ustedes a mi vida y a la vida de su papá. Ya saben que los adoro y tienen mi corazón. Gracias por su paciencia conmigo en todos los momentos difíciles, y por quererme siempre como soy. Lo siento que no haya podido estar con ustedes en todo momento durante estos años, pero les aseguro que siempre han sido y siempre serán mi primera prioridad en la vida. Todo lo que hago, lo hago por ustedes. Son lo mejor que me ha pasado, y siempre lucharé por ustedes. Los amo mucho mucho mucho! Gracias, mi Gaelito, por siempre darme los mejores abrazos y besos del mundo cuando más los necesito. Gracias, mi Lunita, por tu bella sonrisa que siempre me levanta el ánimo. El amor que me dan todos los días es el único sustento que necesito para seguir adelante. Sé que ustedes no se acordarán de esta etapa bonita en la vida de nuestra familia, pero no se preocupen, seguro que su papá y yo les estaremos contando las mismas historias de nuestro tiempo en DC por muchos años más –me imagino que hasta se van a cansar de escucharlas jaja.
A Emmanuel, mi esposo querido. Sé que no te gusta que te diga lo que siento por ti en frente de la gente, pero ni modos, esta vez te tendrás que aguantar. A través de estas seis páginas, he agradecido a muchas personas por muchas razones, pero sin duda, has sido la persona más integral para que este sueño se hiciera realidad. La gente siempre me pregunta cómo logré terminar la tesis con dos niños pequeños, que me parezco a Superwoman, pero de lo que algunos no se dan cuenta es que tengo a Superman a mi lado. Yo sé que no te ves así y que a veces te cuesta creer en ti mismo, pero te pido que pienses bien en todo lo que hemos logrado juntos en estos últimos seis años. Empezaste tu tiempo en DC sin trabajo, sin papeles, sin nada. Un año después, conseguiste un trabajo estable como pintor, y al año siguiente ya habías terminado tu GED. Luego, llegó nuestro Gaelito, lograste sacar tu tarjeta verde a pesar de las dificultades que has vivido en este país, llegó nuestra Luni Luna, y empezaste tu propio negocio de pintura. Y durante todo este tiempo, siempre cuidaste de tu familia emocional y económicamente. Siempre encontrabas la manera de sacarnos adelante – te aguantaste costillas rotas y semanas seguidas sin ningún día de descanso, y mientras todo, me brindaste apoyo infinito en el proceso del doctorado. Cuando necesitaba trabajar (o hasta cuando necesitaba simplemente descansar), te encargabas de los niños sin ninguna queja, a veces por varios días o semanas seguidas. Nunca estabas demasiado cansado para cuidarlos. Así que encima de todo lo que has logrado tú solito en estos últimos años, me apoyaste a mí y apoyaste a esta familia en cada momento sin falta. Tú eres el que merece todos los cumplidos, no yo. Este doctorado es tanto tuyo como mío. Gracias por todo tu apoyo y amor incondicional. Te amo mucho. Estoy contigo hasta que la muerte nos separe.

Many thanks,
Chrissy Bistline-Bonilla
TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION .......................................................................................... 1
    The Statement of the Problem .................................................................................. 1

CHAPTER 2: INTERACTION IN SECOND LANGUAGE ACQUISITION (SLA) ........... 5
    Theoretical Underpinnings of the Interactionist Approach .................................. 5
        Input .................................................................................................................. 7
        Output .............................................................................................................. 7
        Negotiation and Feedback .............................................................................. 9
    Sociocultural Theory .............................................................................................. 13
    Empirical Studies on Interaction .......................................................................... 15
        Face-to-face (FTF) Interaction-based Research ............................................. 18
        Interaction-based Research in Synchronous Computer-Mediated Communication ........ 21
    Second Language Vocabulary Learning ............................................................... 24

CHAPTER 3: KEY VARIABLES IN THE PRESENT STUDY AND INDICATIONS OF
    PREVIOUS EMPIRICAL RESEARCH .................................................................. 33
    Type of Interlocutor .............................................................................................. 33
    Type of Task ........................................................................................................ 48
    State Anxiety ...................................................................................................... 58
        Empirical Studies on Anxiety in Synchronous Computer-Mediated Communication ...... 62

CHAPTER 4: RESEARCH DESIGN AND METHODOLOGY .................................. 69
    Pilot Study ........................................................................................................... 69
    Pilot Methodology ............................................................................................... 70
        Participants ...................................................................................................... 70
        Materials ......................................................................................................... 73
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure</td>
<td>77</td>
</tr>
<tr>
<td>Coding and Analysis</td>
<td>79</td>
</tr>
<tr>
<td>Pilot Results</td>
<td>82</td>
</tr>
<tr>
<td>Research Question 1</td>
<td>82</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>91</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>96</td>
</tr>
<tr>
<td>Research Question 4</td>
<td>99</td>
</tr>
<tr>
<td>Pilot Modifications</td>
<td>102</td>
</tr>
<tr>
<td>Main Study</td>
<td>107</td>
</tr>
<tr>
<td>Methodology</td>
<td>109</td>
</tr>
<tr>
<td>Participants</td>
<td>109</td>
</tr>
<tr>
<td>Materials</td>
<td>111</td>
</tr>
<tr>
<td>Procedure</td>
<td>118</td>
</tr>
<tr>
<td>Coding and Analysis</td>
<td>120</td>
</tr>
<tr>
<td>CHAPTER 5: RESULTS</td>
<td>124</td>
</tr>
<tr>
<td>Research Questions 1 and 2</td>
<td>124</td>
</tr>
<tr>
<td>RQ1a and RQ2a: Language-Related Episodes</td>
<td>124</td>
</tr>
<tr>
<td>RQ1b and RQ2b: Talk on Task</td>
<td>127</td>
</tr>
<tr>
<td>RQ1c and RQ2c: L2 Lexical Recognition</td>
<td>130</td>
</tr>
<tr>
<td>RQ1d and RQ2d: L2 Lexical Production</td>
<td>136</td>
</tr>
<tr>
<td>Summary of Findings for Research Questions 1 and 2</td>
<td>143</td>
</tr>
<tr>
<td>Research Question 3</td>
<td>143</td>
</tr>
<tr>
<td>RQ3a: Language-Related Episodes</td>
<td>144</td>
</tr>
</tbody>
</table>
APPENDIX C: PILOT OPEN-ENDED ASSESSMENT TASK ................................................. 220
APPENDIX D: PILOT FILL-IN-THE-BLANK ASSESSMENT TASK ............................... 221
APPENDIX E: PILOT LEARNER STATE ANXIETY QUESTIONNAIRE .......................... 223
APPENDIX F: DECISION-MAKING TASK .................................................................. 225
APPENDIX G: INFORMATION GAP TASK ................................................................. 228
APPENDIX H: RECOGNITION ASSESSMENT TASK .................................................. 230
APPENDIX I: WRITTEN PRODUCTION ASSESSMENT TASK .................................. 232
APPENDIX J: TARGET LEXICAL ITEMS AND DISTRACTORS .................................. 233
APPENDIX K: LEARNER STATE ANXIETY QUESTIONNAIRE .................................. 234
REFERENCES ........................................................................................................... 236
LIST OF FIGURES

Figure 1. Number of language related episodes by type of interlocutor.................................84
Figure 2. Word count per participant.........................................................................................85
Figure 3. Total word count..........................................................................................................85
Figure 4. Peer interlocutor group...............................................................................................88
Figure 5. Professor interlocutor group........................................................................................88
Figure 6. Peer interlocutor group...............................................................................................88
Figure 7. Professor interlocutor group........................................................................................88
Figure 8. Word count by type of task........................................................................................92
Figure 9. Information gap task. ................................................................................................94
Figure 10. Decision-making task................................................................................................94
Figure 11. Information gap task.................................................................................................94
Figure 12. Decision-making task...............................................................................................94
Figure 13. Time 2 gains (Open-ended) ......................................................................................98
Figure 14. Time 2 gains (Fill-in-the-blank)................................................................................98
Figure 15. Time 3 gains (Open-ended) ......................................................................................99
Figure 16. Time 3 gains (Fill-in-the-blank)................................................................................99
Figure 17. Overall state anxiety in relation to type of interlocutor. .........................................101
Figure 18. Overall state anxiety in relation to type of task.......................................................102
Figure 19. Estimated marginal means of number of language-related episodes, by interlocutor group ..................................................................................................................127
Figure 20. Estimated marginal means of participant word count, by interlocutor group ........129
Figure 21. Participant performance on lexical recognition assessment tasks over time...........133
Figure 22. Participant performance on lexical production assessment tasks over time............139
Figure 23. Correlation between interlocutor anxiety and amount of talk on the decision-making task .......................................................... 146

Figure 24. Correlation between overall anxiety and gains from Time 1 to Time 2 in the recognition of targeted lexical items from the information gap task.............................................. 149

Figure 25. Correlation between overall anxiety and gains from Time 2 to Time 3 in the recognition of targeted lexical items from the information gap task.............................................. 149

Figure 26. Correlation between overall anxiety and gains from Time 1 to Time 2 in the production of targeted lexical items from the information gap task.............................................. 152

Figure 27. Correlation between overall anxiety and gains from Time 2 to Time 3 in the written production of targeted lexical items from the information gap task.............................................. 152
LIST OF TABLES

Table 1. Pilot study’s number of language-related episodes by type of interlocutor ..................83
Table 2. Pilot study’s word count by type of interlocutor..........................................................84
Table 3. Pilot study’s word count per participant – Peer interlocutor group .............................86
Table 4. Pilot study’s word count per participant – Professor interlocutor group......................86
Table 5. Pilot study’s L2 lexical development – Peer interlocutor ...........................................87
Table 6. Pilot study’s L2 lexical development – Professor interlocutor .....................................87
Table 7. Pilot study’s L2 performance on distractor items - Peer interlocutor ............................90
Table 8. Pilot study’s L2 performance on distractor items - Professor interlocutor ....................90
Table 9. Pilot study’s number of language related-episodes by type of task ............................92
Table 10. Pilot study’s word count by type of task ....................................................................92
Table 11. Pilot study’s L2 lexical development – Information gap task .....................................93
Table 12. Pilot study’s L2 lexical development – Decision-making task ....................................93
Table 13. Pilot study’s gain scores at Time 2 (T2 – T1) by type of task .....................................95
Table 14. Pilot study’s gain scores at Time 3 (T3 – T1) by type of task .....................................95
Table 15. Pilot study’s number of language-related episodes in relation to state anxiety ...........97
Table 16. Pilot study’s word count in relation to state anxiety ..................................................97
Table 17. Pilot study’s gains at Time 2 (T2 – T1) in relation to state anxiety .............................98
Table 18. Pilot study’s gains at Time 3 (T3 – T1) in relation to state anxiety .............................99
Table 19. Pilot study’s losses at Time 3 (T3 – T2) .....................................................................99
Table 20. Pilot study’s state anxiety in relation to type of interlocutor ......................................101
Table 21. Pilot study’s state anxiety in relation to type of task ................................................102
Table 22. Number of language-related episodes, by interlocutor group and type of task ........125
Table 23. Tukey’s post hoc test for number of language-related episodes.......................... 126
Table 24. Amount of talk on task, by interlocutor group and type of task ......................... 128
Table 25. Tukey’s post hoc test for amount of talk on task ............................................. 130
Table 26. Descriptive data for recognition assessment task at Time 1, 2, and 3 (in percentages)......................................................................................................................... 131
Table 27. ANOVA summary table for within-subjects effects for recognition assessment data ........................................................................................................................................ 132
Table 28. Paired t-test results for recognition assessment data, separated by type of interlocutor .......................................................................................................................... 134
Table 29. Paired t-test results for recognition assessment data, separated by type of task .... 135
Table 30. Descriptive data for written production assessment task at Time 1, 2, and 3 (in percentages).................................................................................................................. 137
Table 31. ANOVA summary table for within-subject effects for written production assessment data ........................................................................................................................................ 139
Table 32. Paired t-test results for written production assessment data, separated by type of interlocutor .......................................................................................................................... 141
Table 33. Paired t-test results for production assessment data, separated by type of task ...... 143
Table 34. Participants’ self-reported state anxiety levels......................................................... 145
Table 35. Descriptive data for recognition assessment task gains (in percentages) ......... 148
Table 36. Kendall’s tau-b correlations between participants’ self-reported anxiety and gain scores on the recognition assessment task ......................................................................................... 149
Table 37. Descriptive data for written production assessment task gains (in percentages) .... 151
Table 38. Kendall’s tau-b correlations between participants’ self-reported anxiety and gain scores on the written production assessment task ......................................................................................... 152
Table 39. Significant Kendall’s tau-b correlations between participants’ overall state anxiety and gain scores on the assessment tasks ......................................................................................................... 154
CHAPTER 1: INTRODUCTION

The Statement of the Problem

In recent years, research on computer-mediated interaction has grown tremendously (see Ziegler, 2016 for a review). As Gonzalez-Lloret and Ortega (2014) note, this is likely because internet-connected devices and digital technologies have come to form an integral part of the life and learning processes of our newest generation of students (Ito et al., 2009). For most middle-class families in the United States, internet access is no longer considered a luxury, but an essential utility, not unlike water or electricity. The young adult participants found in much of today’s empirical SLA research are members of our newest generation, Generation Z, and have grown up with the ubiquitous presence and convenience of technology that includes cell phones, tablets, laptops, instant messengers, social media, and Google. These technologies have not only shaped their existence, but also their skills, literacies, and cognitive and learning processes (Rosen, 2010; Thorne, 2013). As Leow (2007) suggested, “it may be advantageous to incorporate technology, in both research designs and instructional exposure, by means of carefully designed activities founded on theoretical SLA underpinnings to help learners process L2 structures and establish form-function relationships.” (p. 47).

Beginning in the early 2000s, a number of researchers began to incorporate these technologies into their investigations of foreign language learning. Many studies that have empirically addressed interaction via computer-mediated communication (CMC) have considered the opportunities for language learning it provides when compared with the face-to-face (FTF) mode. A recent meta-analysis of these studies conducted by Ziegler (2016) revealed an advantage for interaction in synchronous CMC over FTF on overall second language (L2) learning outcomes, and more specifically, productive and written measures, albeit with low
effect sizes, but no significant differences between the two modes of communication for L2 development resulting from interaction. As a result, Ziegler and Mackey (2017) have concluded that L2 learners interacting via SCMC will have similar or potentially improved opportunities for L2 learning. Given these findings, and combined with the extensive research that has already been conducted on task-based interaction in both the FTF mode and FTF versus SCMC, the field seems to be gravitating towards more empirical research that focuses solely on interaction in SCMC, the modality employed in the present study.

Two of the most widely researched variables within the interaction strand of instructed second language acquisition (ISLA) research are type of interlocutor and type of task. The former refers to the speaker with whom an L2 learner interacts, who can be a peer, teacher, native speaker, nonnative speaker, and so on. Although it has been empirically demonstrated in SCMC that peer interaction, when examined on its own or in comparison with native speaker-nonnative speaker (NS-NNS) interaction, can be facilitative of L2 learning (Blake, 2000; Pellettieri, 2000; Smith, 2003; Yilmaz and Granena, 2010; Yilmaz, 2011; Zeng, 2017; Kim, 2017), the potential benefits of peer interaction when compared to interaction involving other types of interlocutors are less clear. Only a handful of studies have empirically compared the effects of learner-learner versus learner-teacher interaction (e.g., Van den Branden, 1997), and no empirical studies have been conducted with learner-teacher, learner-learner, and learner-native speaker dyads within the same research design. The comparison of these three types of interlocutors, and more specifically, the incorporation of teacher interlocutors into research designs, is one that deserves attention in task-based interaction in SCMC given that teachers can function as ‘more knowledgeable’ partners who encourage instances of negotiation for form and modified output (Avermaet, Colpin, Van Gorp, Bogaert, & Van den Branden, 2006).
Furthermore, although native speakers and non-native speakers have often been empirically compared as interlocutors in ISLA research, the native speakers are also often teachers, and in some cases, the participants’ teachers. Sometimes these participants are defined as teacher interlocutors and sometimes they are defined as native speaker interlocutors, but their identity as both is often not considered by researchers when interpreting empirical findings. As a result, teachers’ and native speakers’ position of linguistic authority and its ability to affect learners’ performance and emotions on task needs to be considered.

One of the advantages of text-based SCMC is its self-paced setting, which researchers have argued (e.g., Long & Robinson, 1998; Smith, 2003) is likely to further facilitate the noticing of L2 forms and may potentially result in subsequent learning. This added time to process linguistic input and plan output can help foster a ‘safe space’ for anxious learners (Beauvois, 1992), who often fear judgment based on linguistic insecurities such as imperfect pronunciation (Kern, 1995). A limited number of studies have empirically investigated the effects of modality on learners’ state anxiety with varying results (Baralt & Gurzynski-Weiss, 2011; Cote & Gaffney, 2018); however, at no point have the participants of any empirical study reported higher state anxiety levels while interacting in SCMC than while interacting face-to-face, suggesting that the SCMC mode may offer a similar or slightly improved environment for anxious learners. In summary, although SCMC seems to offer an environment that is similar or less anxiety-provoking than the FTF mode, very few studies have investigated affective factors in SCMC.

Only a small number of studies have explored the effects of task type on amount of negotiation in text-based SCMC contexts (e.g., Blake, 2000; Smith, 2003; Yilmaz, 2011; Yilmaz & Granena 2010), and even fewer studies have empirically examined the effects on L2 learning
outcomes (Baralt, 2013). With the exception of Keller-Lally’s (2006) dissertation, all other CMC studies in ISLA have reported that tasks can affect the amount of negotiation that occurs in dyadic interaction. However, the variety of tasks employed in researchers’ experimental designs have resulted in highly variable amounts of negotiation for meaning and negative feedback, ranging from 3% to 30% of turns (Ortega, 2009). Furthermore, given the potential impact of learner characteristics, context, mode, and so on, it is difficult to identify a single task design as being uniformly ideal for eliciting negotiation. For instance, while Blake (2000) found jigsaw tasks to be superior to one-way information gap and decision-making tasks, Smith (2003) reported that his decision-making tasks resulted in more negotiation than jigsaw tasks. In addition, while Yilmaz (2011) reported that dictogloss tasks were superior to jigsaw tasks, Yilmaz and Granena (2010) found opposite patterns, with the jigsaw task resulting in more accurately solved language-related episodes (LREs) than the dictogloss task. Ultimately, given the divergent findings of a limited number of studies, the present study will also address the effects of task type in text-based SCMC.

Given the potential impact of interlocutor type on the characteristics of interaction and participant performance, along with the divergent findings on task effects, combined with the scant research on affective factors in SCMC, the present study offers an investigation examining the effects of type of interlocutor and type of task on 1) the amount of negotiation on task, 2) the amount of talk on task, 3) learners’ recognition and production of L2 lexical forms over time, and 4) the relationship between type of interlocutor and type of task and learners’ foreign language state anxiety. The research motivating these variables is discussed in more detail below, beginning with the theoretical underpinnings of interactionist research.
CHAPTER 2: INTERACTION IN SECOND LANGUAGE ACQUISITION (SLA)

This chapter first offers an overview of the theoretical underpinnings of the interactionist approach and its major components: input, output, negotiation and feedback. This is followed by a review of literature in the interaction strand, starting with empirical studies conducted in the Face-to-Face (FTF) mode, followed by a summary of interaction-based studies that have compared FTF interaction with Synchronous Computer-Mediated Communication (SCMC) or that have examined interaction in SCMC alone.

Theoretical Underpinnings of the Interactionist Approach

The interactionist approach is a common framework in second language acquisition (SLA) research that stems from Long’s (1981) Interaction Hypothesis. The Interaction Hypothesis posits that during an interaction there are breakdowns in communication that result in negotiation for meaning among learners, which results in the interlocutors providing input that is more accessible to the learner (Gass, Mackey & Pica. 1998; Long 1981, 1983). Long (1980, 1981, 1983) first investigated native speaker-nonnative speaker (NS-NNS) interactions in an attempt to build on the work of Hatch (1978), who had investigated NS-NNS interactions involving naturalistic child and adult L2 learners. Based on her observations, Hatch suggested that interaction with more advanced interlocutors had a direct effect on the acquisition of L2 grammatical forms. Long’s early work led him to propose the concepts of negotiation and interactional modifications.

Much of the earliest interaction research focused on the potential of negotiation for second language development. However, the first published study to empirically address the relationship between L2 development and interaction did not come until 1994 when Ellis,
Tanaka, and Yamazaki addressed the effects of modified interaction on vocabulary acquisition and comprehension with two classroom studies of 79 and 127 high school students. Since the publication of Ellis et al. (1994), many studies in interaction research have empirically demonstrated that interaction facilitates L2 learning (see Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010; Ziegler, 2016 for reviews). However, what constitutes interaction that is facilitative of L2 learning? Interaction exposes learners to linguistic information (input), provides them with opportunities to produce language (output), and offers them positive and negative feedback on their production (Mackey, Ziegler & Bryfonski, 2016). As Mackey et al. (2016) highlight, “Although input is a key component of L2 acquisition, the interaction approach suggests that it is the ways in which learners interact with input and their interlocutors through negotiating for meaning […] that are most likely to support L2 development” (p. 108). Negotiating for meaning is a pivotal element of interaction given that through the provision of feedback, learners are given opportunities to identify gaps between their interlanguage and the target language and produced modified output (Long, 1996; Swain, 1985, 1995, 2005). However, while input and interaction with said input are pivotal to L2 development, opportunities to practice and produce the L2 are equally as important. Production aids learners in moving from processing meaning to processing forms, since production requires that learners attend to the linguistic forms they are producing instead of simply comprehending the input they are exposed to (Swain, 1985). Thus, though it may come as no surprise that communication is vital to language learning, the interaction approach posits that it is important that this communication possess certain characteristics in order to facilitate the ultimate outcome: acquisition. These characteristics and their foundations in (I)SLA theory are discussed further below.
Input

Input, or linguistic information that a learner is exposed to through reading and listening, is commonly understood as a vital component of language learning. Access to comprehensible input, which is input made to be understood by learners, is needed in order for language acquisition to take place, and a lack of input will lead to little or no language development (Long, 1983). Although Long’s (1981) Interaction Hypothesis posits that comprehensible input, output and positive and negative feedback are all needed for language development, not all SLA researchers agree. For instance, Krashen’s Input Hypothesis (Krashen, 1981, 1982, 1985) posited that language acquisition is achieved solely from comprehensible input, and more specifically, exposure to language that is slightly beyond the current level of competence \((i + 1)\) (Krashen, 1982). However, Krashen specifies that “natural, communicative, roughly-tuned, comprehensible input has some real advantages over finely-tuned input that aims directly at \(i + 1\), in other words, classroom exercises that aim to teach the structure of the day.” (p. 25). Furthermore, as part of his Affective Filter Hypothesis, Krashen suggests that in addition to comprehensible input, learners must not be forced to produce language, given that this could cause them considerable anxiety, resulting in a high ‘affective filter’, and disrupting and ultimately preventing the acquisition of the target language (Krashen, 1982). Proponents of the modern-day interactionist approach maintain that although Krashen is correct in suggesting that input is key to interaction, it is not sufficient on its own for fostering L2 development (Mackey, 2012).

Output

Equally important to the L2 acquisition process is output, which interactionists operationalize as opportunities for practicing the language. Swain (1985, 1995) demonstrated
this through her work with children in a French immersion program. She discovered that the
children’s productive skills in L2 French were lacking when compared to their L1 French peers,
despite possessing native-like receptive skills. She posited that this was caused by a lack of
opportunities to practice speaking the L2. Ultimately, this led her to formulate the Output
Hypothesis, which proposed that in order for L2 development to take place, learners need to
produce output. It is important to note, however, that Swain was not contradicting Krashen’s
hypothesis, but rather complementing it: “Sometimes, under some conditions, output facilitates
second language learning in ways that are different from, or enhance, those of input.” (Swain &

According to Swain, there are three functions of output: the noticing function, the
hypothesis-testing function, and the metalinguistic function. During opportunities for output,
learners encounter a gap in their knowledge of the L2, thus they notice what they do not know.
As a result, they produce utterances that test their underlying hypotheses about the L2. When
they receive feedback from their interlocutor, they are “pushed” to make adjustments and modify
their output in order to be understood. These opportunities have the potential to promote fluency
and automaticity, a critical process in L2 development.

Early interpretations of the Output Hypothesis initially equated the concepts of pushed
output and comprehensible output. Pushed output, or modified output, was defined as the
learner’s reformulation of an utterance in response to self-monitoring or feedback from an
interlocutor (Mackey, 2012). The concept of comprehensible output, on the other hand, was later
modified to include all utterances that the interlocutor could comprehend (Van den Branden,
1997). Modified output promotes L2 development by pushing learners to make their output more
target-like (Swain, 1985, 1995; Shehadeh, 2002; McDonough, 2005; McDonough & Mackey,
2006) and providing opportunities for corrective feedback, which helps learners attend to mismatches between the L2 and their own production, facilitating noticing (Izumi, Bigelow, Fujiwara, & Fearnow, 1999), and potentially subsequent L2 learning. However, noticing mismatches alone is not a sufficient condition for producing output at this stage – they need to be processed beyond noticing for the linguistic information to be internalized, which may then lead to modified output (Leow & Driver, in press). Furthermore, as Mackey (2012) highlights, whether or not the learner’s modified input equates with producing the targeted L2 forms correctly, modifying output is still a beneficial process for language development. Ultimately, while both the input and output hypotheses are valid to a certain extent, interaction-based empirical research over the past few decades has demonstrated that they both represent only a portion of the bigger picture in SLA.

**Negotiation and Feedback**

Up to this point, some of the major theories in SLA that have influenced the current interactionist approach have been discussed. These hypotheses have essentially focused on either input or output, neglecting to address the third key component, negotiation, and failing to recognize these three constructs as inherently intertwined in the process of L2 development, effectively three cogs in the same wheel. Around the same time that the Input hypothesis was proposed and a few years before the Output hypothesis was proposed, Long proposed his initial Interaction Hypothesis (Long, 1981), which was inspired by the research of Hatch (1978) and Long’s subsequent research into NS-NNS interaction (Long, 1980, 1981, 1983). Hatch’s research on NS-NNS interaction led to her suggestion that “language learning evolves out of learning how to carry on conversations” (Hatch, 1978, p. 63), which, combined with his own
research, led Long to the notions of interactional modifications and negotiation. Long argued that when more advanced learners or native speakers of the L2 are interacting with less advanced learners, they make adjustments in their speech in order to provide the less advanced learners with more comprehensible input and to repair any communication breakdowns. These opportunities for comprehensible input and negotiation for meaning are enough to facilitate L2 development.

According to Long (1983), types of negotiation for meaning can include confirmation checks, clarification requests, and comprehension checks. Confirmation checks are designed to “elicit confirmation that the utterance has been correctly heard or understood by the speaker” (Long, 1983, p. 137) and comprehension checks are meant to confirm the learner understood the interlocutor through an attempt to “anticipate and prevent a breakdown in communication (Long, 1983, p. 136). Clarification requests are utterances “designed to elicit clarification of an interlocutor’s preceding utterance” (Long, 1983, p. 137).

In Long’s later version of the Interaction Hypothesis (Long, 1996), he included a number of key changes. First, he incorporated Swain’s (1985) Output Hypothesis by emphasizing pushed output, and the opportunities it provides for fluency, automaticity, hypothesis testing, and feedback, as vital elements of interactionist theory. Second, citing Tomlin and Villa’s (1994) model for the role of attention in SLA, Long highlighted the increased opportunities for learners to focus on form. He argued that it is more likely that the learner will be alert during conversation since they may be interested in seeing how their utterances affect her interlocutor. Furthermore, because they already know what they have attempted to communicate to the interlocutor, their processing resources are freed, allowing them to pay attention to the form of the interlocutor’s response.
Most importantly, Long’s (1996) update underscored the importance of negative evidence, which occurs when interlocutors offer learners direct or indirect information on their use of ungrammatical forms. In other words, learners receive negative evidence through interlocutors’ feedback on their utterances, which often occurs while negotiating for meaning. Ultimately, exposure to negative evidence helps learners notice the gap between their non-targetlike L2 production and the targetlike form, thus connecting input, internal learning capacities, and output in productive ways, promoting L2 development (Long, 1996, pp. 451-452). Negative evidence, or corrective feedback, can vary in the following ways: 1) its degree of explicitness, 2) whether a targetlike model is provided (i.e., positive evidence), and 3) whether the type of feedback prompts the learner to modify his or her output (i.e., modified output). In addition, although Long initially suggested that in order for interaction to promote L2 development, the interlocutor must be a more advanced speaker of the L2, there has also been quite a bit of research conducted addressing the L2 developmental benefits of peer interaction (e.g., Lyster, Saito, & Sato, 2013; McDonough, 2004; Philp & Iwashita, 2013; Sato & Lyster, 2012; Toth, 2008), with most researchers agreeing that peer interaction has the potential to facilitate L2 development.

More recently, the construct of negotiation has been expanded to include forms of implicit and explicit feedback, such as recasts and metalinguistic feedback (Mackey, 2012). To clarify, recasts are a type of feedback that can be classified either as implicit or explicit, and involve the interlocutor repeating the learner’s erroneous utterance and replacing the non-targetlike form with the grammatical one. They have been shown to facilitate L2 development given that they increase the salience of target features and allow learners to compare their incorrect utterance with the interlocutors’ reformulation (e.g., Long, 2007; Goo & Mackey,
Nevertheless, recasts’ effectiveness can depend on a myriad of different factors, including learner proficiency (Mackey & Philp, 1998), task characteristics (Revesz, Sachs, & Mackey, 2011), instructional context (e.g., Lyster & Mori, 2006; Toth & Garritano, 2008), and individual differences (Mackey, 2012), among others. Although few studies have found limited SLA benefits stemming from the provision of recasts (e.g., Lyster, 2004), most research has empirically demonstrated that recasts are beneficial to L2 development (see Mackey & Goo, 2007 for a review).

In summary, Long’s Interaction Hypothesis (1996) suggests that successful interaction possesses the following three characteristics: linguistic information (input), opportunities to produce language (output), and positive and negative feedback (interaction). These key interactional features have guided empirical ISLA research on interaction for various decades since Long’s initial proposal in 1981.

---

1 It should be noted that the earliest investigations into corrective feedback were born out of researchers’ goal to understand the role of errors in language learning (e.g., Corder, 1967; Selinker, 1972). Since then, corrective feedback has been widely investigated in (I)SLA, and this research has not only been founded on the Interaction Hypothesis. Multiple other underpinnings have also served as the theoretical foundation of corrective feedback studies, such as the Noticing Hypothesis (Schmidt, 1990), Skill-learning Theory (DeKeyser, 2015), and the Model of the L2 learning process in ISLA (Leow, 2015) (see Leow & Driver, forthcoming for further elaboration). Several meta-analyses (e.g., Brown, 2016; Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006) of these empirical studies have all concluded that corrective feedback is beneficial to L2 learning.
Sociocultural Theory

Although the present study is positioned within the cognitive-interactionist framework and the researcher utilizes this framework to explicate the majority of the findings, the somewhat limited empirical research conducted within an interactionist framework regarding the effects of type of interlocutor, particularly in terms of expert-novice collaboration, calls for a review of empirical studies within the sociocultural framework. Therefore, the purpose of this section is to offer a brief overview of sociocultural theory and a comparison of the two frameworks.

Sociocultural theory was first proposed by psychologist Lev Vygotsky, who posited that knowledge is socially constructed, and relatively similar to Long’s Interaction Hypothesis (1996), that learning results from interacting with knowledgeable others. In other words, he theorized that social interaction and language are key components of learning (Vygotsky, 1978). Vygotsky also proposed the Zone of Proximal Development (ZPD), which represents the distance between what a learner is capable of accomplishing on her own and what she can achieve with the help of a more competent interlocutor. In the past, learning within ZPD has been perceived as occurring in expert-novice dialogic interaction, where an expert (e.g., a teacher) offers contingent and graduated assistance to a novice (e.g., a student) (Aljaafreh and Lantolf, 1994). As Donato (1994) explains:

Social interaction is a mechanism for individual development, since in the presence of a more capable participant, the novice is drawn into, and operates within, the space of the expert’s strategic processes for problem solving. This collaboration is a way for the novice to extend current competence. The experienced individual is often observed to guide, support, and shape actions of the novice who, in turn, internalizes the expert’s strategic processes. (p. 37)

Many socioculturists have also empirically addressed opportunities for language learning in terms of scaffolding (e.g., Donato, 1994; Heritage, 2012). Scaffolding is a concept with origins in cognitive psychology and L1 research that says that during social interaction, an expert
participant can foster supportive linguistic conditions in which the novice may participate, giving them the opportunity to reach for a higher level of competence (Greenfield, 1984; Wood, Bruner, & Ross, 1976). The characteristics of successful scaffolding may include: 1) recruiting interest in the task, 2) simplifying the task, 3) maintaining pursuit of the task, 4) marking critical features and discrepancies between what has been produced and the ideal solution, 5) controlling frustration during problem solving, and 6) demonstrating an idealized version of the act to be performed (Wood, Bruner, and Ross, 1976). Although it was initially posited that scaffolding, and therefore language learning, only occurs between an expert (teacher) and a novice (student), more recent empirical research has demonstrated that the asymmetry in knowledge between interlocutors required for language learning is present in almost all types of interaction (Heritage, 2012). For this reason, scaffolding may also be referred to as collaborative dialogue, which Swain (2000) defines as “dialogue in which speakers are engaged in problem solving and knowledge building” (p. 102). This also ties into intersubjectivity, a characteristic of interaction that is commonly addressed by sociocultural research. Intersubjectivity is when L2 learners mutually construct a scaffold through negotiating contexts of shared understanding. This shared understanding of task goals and means allows them to coordinate activity during collaboration (Rommetveit, 1985).

Foster and Ohta (2005) offer a detailed comparison of the cognitive-interactionist and sociocultural frameworks. While cognitive theory posits that the language learning process involves the acquisition of systems of knowledge (e.g., lexical, grammatical, phonological) in the L2, sociocultural researchers see L2 development as a process occurring in social interaction and in particular settings and social contexts. Cognitive-interactionist research favors quantitative research methods where dependent variables are operationalized and controlled, and participants
are subjected to a variety of treatments, the impact of which are measured statistically with the aim of producing findings that are generalizable. The process of acquiring the L2 system during interaction is often measured through a quantification of learners’ fluency, accuracy, and linguistic complexity. On the other hand, sociocultural researchers believe that this sort of research “masks fundamentally important mechanisms for L2 development and reduces the social setting to an opportunity for input crunching” (Donato, 1994, p. 34). According to Foster and Ohta (2005), “This risks sacrificing the richness of the interaction that occurs, eliminating the subjectivity of both researcher and study participant. Selection of categories for quantification is viewed as sacrificing the whole for the sake of a partial picture that may not apply to any real-world situations” (p. 403). Instead, they believe that language development should be studied by examining how a learner uses the L2 during interaction with others. Among other techniques, this is often done through microgenetic analysis or discourse analysis. Socioculturists are often concerned with attention and memory, and how these processes are manifested during interaction. However, a shortcoming of this type of research is that the findings cannot produce generalizable results (Foster & Ohta, 2005).

In summary, although the theories behind the cognitive-interactionist and sociocultural approaches differ greatly and the current study takes a mostly cognitive-interactionist approach to the data, past empirical research from the sociocultural perspective, particularly regarding the social relationships between interlocutors, serves to fill some of the gaps not widely addressed in cognitive-interactionist research.

**Empirical Studies on Interaction**

The interaction approach attempts to “enable learners to move beyond their current receptive and productive capacities when they need to understand unfamiliar language input or
when they are required to produce a comprehensible message” (Kumuravadivelu, 2006, p. 69).

While initial research on interaction examined the effects of conversational adjustment on L2 development (e.g., Gass & Varonis, 1994; Mackey, 1999), more recent research has addressed a wide range of variables, including the effects of different types of implicit and explicit corrective feedback (e.g., Ellis, 2007; Loewen & Philp, 2006; Oliver & Mackey, 2003; Sato & Loewen, 2018), overt focus on form through language-related episodes (LREs) (e.g., Bueno-Alastuey, 2013), social factors, including peer attitudes and relationships (e.g., Philp & Tognini, 2009), and cognitive factors, such as working memory capacity (e.g., Mackey, Adams, Stafford, & Winke, 2010; Mackey & Sachs, 2012). This research has been conducted both in the lab and in the classroom with varying participant populations and modes of interaction (e.g., face-to-face vs. computer-mediated), and meta-analyses of these studies have concluded that, overall, interaction is beneficial to L2 development (Keck, Iberri-Shea, Tracy-Ventura, & Wa-Mbaleka, 2006; Mackey & Goo, 2007; Ziegler, 2016).

Many interaction-based studies have considered the effects of collaborative tasks in terms of the occurrence of language-related episodes (LREs). LREs are defined as ‘any part of dialogue where the students talk about the language they are producing, question their language use, or correct themselves or others’ (Swain & Lapkin, 1998, p. 326). This includes recasts and uptake, in addition to more elaborate languaging. Empirical research on LREs has suggested that collaborative tasks such as jigsaw, dictogloss, or text reconstruction tasks promote ‘languaging’, where learners discuss the language they are using during the task, and they can also draw learners’ attention to vocabulary and grammar. LREs can focus on any aspect of language, including grammar, lexis, pronunciation, or sociolinguistic appropriateness. Task-based research, from both the interactionist and sociocultural perspectives, has shown that meaning-oriented
tasks, such as jigsaw or dictogloss tasks, tend to result in more lexical than grammatical LREs (Alegría de la Colina & García Mayo, 2007; García Mayo, 2002; Philp, Walter, & Basturkmen, 2010; Storch, 2001a; Swain & Lapkin, 2001). This can be contributed to the type of tasks employed in these studies, but lexis is also often more salient to learners, making it more likely for them to interact overtly with lexical structures. Overall, it would seem that lexical LREs tend to be more common than grammatical LREs (Fernández Dobao, 2012; Williams 1999, 2001), but factors such as learners’ proficiency level (Leeser, 2004; Williams, 1999), pair dynamics (Kim & McDonough, 2008; Storch, 2002; Watanabe & Swain, 2007), and the mode of communication (Shekary & Tahririan, 2006; Zeng & Takatsuka, 2009) can impact the frequency of both types of LREs.

During lexical LREs, learners discuss the meaning of a word or try to choose the correct lexical form for a certain context. Empirical research from a sociocultural perspective has demonstrated that the production of lexical LREs during interaction between two learners or between a native and nonnative speaker can help learners retain these lexical items, potentially facilitating learning (Fernández Dobao, 2012; Tocalli-Beller & Swain, 2005). Given that LREs are thought to facilitate learning, several sociocultural studies have also addressed whether type of interlocutor can affect the frequency of LREs (e.g., Fernández-Dobao, 2012; Leeser, 2004; Storch, 2002; Watanabe & Swain, 2007). Overall, these studies have reported that different interlocutor characteristics can impact the pattern of exchanges in dyadic interaction. For instance, Fernández Dobao (2012), addressed the production of lexical LREs with learner-learner and NS-NNS dyads, and she found that LREs were more frequent and often more successfully resolved in NS-NNS dyads than in learner-learner dyads.
As explained by Gass and Mackey (2007), the role of interaction in second language development can be studied through the use of any one of four data collection methods: 1) naturalistic data (e.g., Tarone & Liu, 1995), 2) prompted production (e.g., Iwashita, 2003), 3) prompted responses (e.g., Ellis, Tanaka, & Yamazaki, 1994), or 4) introspective data (e.g., Mackey, 2002). Interaction-based research most commonly employs prompted production, which refers to the use of tasks that promote language production. This is also the data collection method of choice for the present study. The use of tasks in empirical interaction-based research has been extensive, with Mackey and Goo (2007) meta-analyzing 28 different studies from 1990 to 2006. Since then, task-based interaction research has only grown exponentially, likely because it not only lends opportunities for input, negotiation, and output, but when the task is carefully designed, it also allows researchers to manipulate said input, negotiation, and output in order to promote the use and learning of targeted linguistic forms. Furthermore, according to Foster and Skehan (2009), given that tasks always have a beginning, middle, and end, task-based interaction has the potential to ease the language processing burden, allowing learners to focus on producing accurate and fluent language. In the sections below, a review of interaction-based research is divided into sections based on the experimental medium employed, culminating in the medium of the current study: synchronous computer-mediated communication, or SCMC.

**Face-to-face (FTF) Interaction-based Research**

Early research on interaction aimed to describe interactional modifications in NS-NNS dyads and the resulting effects on learners’ L2 comprehension (e.g., Gass & Varonis, 1985; Long, 1983; Pica, 1988). Although the logical next step was to empirically investigate the relationship between interaction and L2 development, one of the first studies to do so within the
cognitive-interactionist framework was not published until 1994, when Ellis, Tanaka, and Yamazaki investigated whether L2 learners exhibited greater vocabulary retention and comprehension after receiving interactionally modified input when compared with learners who received premodified or unmodified input. The researchers found that the group receiving interactionally modified input generally outperformed the other two groups on the task, a post-test, and two delayed post-tests. However, they also reported that the participants who actively participated in the negotiation of meaning did not acquire more vocabulary or comprehend more than those who were simply exposed to modified interaction. Nevertheless, the researchers concluded that opportunities for interactionally modified output had the potential to promote comprehension and lexical development (Ellis et al. 1994).

A study by Gass and Varonis (1994), which investigated the roles of modified input and interaction in task performance and L2 comprehension, produced similar results. Participants were 16 nonnative speakers of English with varying L1s and 16 native speakers of English studying at the same university. Each NS-NNS dyad interacted to complete two direction-giving tasks, a type of information gap task. Based on their observations, the researchers reported a significant relationship between modified input from NS interlocutors and 1) task performance, 2) immediate comprehension gains, and 3) delayed effects on production.

While these two empirical studies were some of the first to demonstrate a beneficial relationship between interaction and L2 development, many other subsequent studies have also supported their findings (e.g., Iwashita, 2003; Leeman, 2003; Long, Inagaki, & Ortega, 1998; Mackey, 1999; McDonough, 2005; Muranoi, 2000; Philp, 2003, but see Mackey & Goo, 2007 for a review). L2 development has often been operationalized as the learning of lexical items, grammatical items, or both, with many studies finding that lexical items facilitate more
negotiation than grammatical items (Williams, 1999; Mackey, Gass, & McDonough, 2000; Mackey & Goo, 2007).

In spite of the countless studies that have reported that negotiation facilitates L2 learning, not all researchers have produced findings that support this relationship. For example, Newton (2013) utilized two opinion gap tasks and two information gap tasks with two groups of four ESL learners. He found that only a small number of the total unfamiliar words in the tasks were negotiated for meaning, and although learners were more likely to show improvement in the assessments on negotiated words, they also showed improvement on many words that had not been negotiated at all, bringing into question the role of negotiation in facilitating L2 learning. Nevertheless, it must be noted that this study included a very small sample size.

Many researchers who have empirically investigated interaction within a sociocultural framework have argued that the quantification of negotiation alone during interaction may not necessarily determine a task’s ability to provide learners with language learning opportunities (Foster & Ohta, 2005). For example, Nakahama, Tyler, and Van Lier (2001) used a two-way information gap task and a conversation task to measure negotiation for meaning. The researchers found that the analysis of negotiation for meaning did not accurately capture the learning opportunities created by the two tasks: Although the conversational task produced fewer instances of negotiation for meaning, it was also more challenging and provided learners with more opportunities to consider the broader discourse. Ultimately, the researchers argued that analyses should go beyond negotiation for meaning when determining the usefulness of a task for language learning.

Interaction-based research has also examined the relationship between L2 development, interaction, and a number of different variables such as type of feedback (e.g., Ellis, 2007;
Loewen & Philp, 2006), learner uptake (e.g., Loewen, 2005; McDonough & Mackey, 2006), individual differences, including working memory (e.g., Mackey & Sachs, 2012; Kim, Payant, & Pearson, 2015) and anxiety (Robinson, 2007), context (classroom vs. laboratory) (e.g., Gass, Mackey, & Ross-Feldman, 2005), medium (FTF vs. SCMC) (e.g., Baralt & Gurzynski-Weiss, 2011; Baralt, 2013; Kim, 2017), type of interlocutor (e.g., Leeser, 2004), and type of task (e.g., Baralt, 2013; Newton, 2013). A majority of interaction-based empirical studies has been carried out in the laboratory setting, although some researchers have also documented the benefits of interaction in the classroom (Ellis, 2000; Mackey & Goo, 2007; Oliver, 2000; Samuda, 2001; Williams, 1999). In the past, some researchers (e.g., Foster, 1998; Nunan, 1991) have argued that interaction may not be as effective for L2 learning in the classroom setting. However, upon direct comparison of the classroom and laboratory contexts, other studies have shown that similar features of interaction appear, with variables such as type of task having a greater impact on the amount and type of negotiation (Gass et al., 2005). Overall, identifying which type of task has the potential to produce the greatest amount of negotiation has proved difficult for researchers in the interactionist strand. Some researchers have argued that there is no one-size-fits-all when it comes to task, since individual differences, communicative needs among learner populations, and highly variable conditions of context can influence the efficacy of any task (Long, 2015; Slimani-Rolls, 2005). This is an argument that must be considered when undertaking any empirical investigation of task type effects.

**Interaction-based Research in Synchronous Computer-Mediated Communication**

There is an increasingly large body of research that has addressed how and why interaction in the synchronous computer-mediated communication (SCMC) mode may support
L2 development (e.g., Chapelle, 2009; Kern, Ware, & Warschauer, 2008; Sauro, 2011; Ziegler, 2016). SCMC has been defined as ‘real-time, synchronous conversation that takes place online’ (Baralt & Leow, 2016, p. 200) and a hybrid between written and oral interaction with the potential to facilitate L2 learning (Beauvois, 1992; Lamy & Hampel, 2007; Pellettieri, 2000). Chun (2016) highlights the mode’s hybridity by asking a pivotal question: “What counts as speaking and what counts as writing? In today's world, so much communication is done via writing, typing, and texting rather than speaking that the lines between writing and speaking are becoming blurred” (p. 105).

When compared to the oral mode, interaction in SCMC provides relatively similar opportunities for interaction, negotiation for meaning, and corrective feedback (Blake, 2000; Fernández-Garcia & Martínez-Albelaitz, 2002; Lee, 2004; Pellettieri, 2000; Sauro, 2009). Researchers have posited that SCMC might provide learners with additional opportunities for noticing, and therefore increased opportunities for L2 development than traditional face-to-face (FTF) interaction, due to the possibility of greater saliency, additional opportunities to review input and output, and more time for processing and planning production (Pellettieri, 2000; Smith, 2004, 2005; Warschauer, 1996, 1997; Yuksel & Inan, 2014). This is especially true for text-chat, given that learners are offered a written record of the interaction, which can provide added opportunities for learners to attend closely to form and meaning (Ziegler & Mackey, 2017), facilitate accuracy (Salaberry, 2000), and notice gaps in their interlanguage (Kelm, 1992; Gass, 1997). Furthermore, given that intonation, facial expressions, and gestures that often guide visual interaction are absent in text-based SCMC, learners are encouraged to use language (e.g., pragma-linguistic forms) or emojis to express themselves (Sykes, 2005) as they encounter and attempt to resolve gaps in their interlanguage (Swain, 2005).
Many researchers have argued (e.g., Long & Robinson, 1998; Smith, 2003) that not only the visual salience of written discourse, but also the self-paced setting of text-based SCMC, are likely to further facilitate the noticing of L2 forms and may potentially result in subsequent learning. This added time to process linguistic input and plan output can help foster a ‘safe space’ for anxious learners (Beauvois, 1992), who often fear judgment based on linguistic insecurities such as imperfect pronunciation (Kern, 1995). A limited number of studies have empirically investigated the effects of modality on learners’ state anxiety, which is anxiety in response to a specific stimulus, contrasting with trait anxiety which is used to describe someone who has an overall tendency toward anxiety. These empirical studies have produced varying results (Baralt & Gurzynski-Weiss, 2011; Cote & Gaffney, 2018); however, at no point have the participants of any empirical study reported higher state anxiety levels while interacting in SCMC than while interacting face-to-face, suggesting that the SCMC mode offers a similar or slightly improved environment for anxious learners. In addition, SCMC can result in increased participation and improved learner attitudes (Chun, 1994; Kern, 1995).

Numerous empirical studies in interaction-based SCMC have investigated L2 development, more specifically L2 lexical development (de la Fuente, 2003; Smith, 2004, 2005), unfocused grammatical items (Shekary & Tahririan, 2006), and focused grammatical items (Baralt, 2013; Sachs & Suh, 2007; Shintani & Aubrey, 2016; Yilmaz, 2012; Yilmaz & Yuksel, 2011). In line with studies investigating FTF interaction, these studies have generally reported more instances of negotiation with lexical items than with grammatical items (e.g., Pellettieri, 2000; Blake, 2000; Smith, 2003). Other SCMC-based studies have also addressed the relationship between interaction, L2 development, and a number of different variables such as awareness (Sachs & Suh, 2007; Gurzynski-Weiss, Al Khalil, Baralt, & Leow, 2016; Bistline-
Bonilla, DeRobles, & Xu, 2019), uptake (e.g., Smith, 2010), type of task (e.g., Baralt, 2013; Blake, 2000), anxiety (e.g., Arnold, 2007; Baralt & Gurzynski-Weiss, 2011), type of interlocutor (e.g., Bueno-Alastuey, 2013; Liu, 2017), and pronunciation (Loewen & Isbell, 2017).

When comparing SCMC and other mediums for interaction, Ziegler (2016) found a small advantage in terms of effect sizes for interaction in SCMC over FTF on overall L2 learning outcomes, and more specifically, productive and written measures, but no significant differences between the two modes of communication for L2 development resulting from interaction. As a result, Ziegler and Mackey (2017) have concluded that L2 learners interacting via SCMC will have similar or potentially improved opportunities for L2 learning. In addition, Lin, Huang and Liou’s (2013) meta-analysis of 10 studies found a small but positive overall effect ($m = .33$) of text-based SCMC on second language acquisition when compared with other modes of communication. As a result, they suggested that text-based SCMC could make a larger difference in L2 development than other means of communication. Given these findings, and combined with the extensive research that has already been conducted on task-based interaction in both the face-to-face mode and face-to-face versus SCMC, the field seems to be gravitating towards more empirical research that focuses solely on interaction in SCMC, the modality for the present study.

**Second Language Vocabulary Learning**

One of the most commonly investigated aspects of L2 development is vocabulary learning. As explained by Nation (1990), there are three main aspects of vocabulary knowledge: 1) knowledge of form (spoken, written, and word parts), 2) knowledge of meaning, and 3) knowledge of use (grammatical function, collocations, and constraints on use). Vocabulary
learning can take the form of either recognition of lexical forms, production of lexical forms, or both. Receptive and productive lexical knowledge are not a dichotomy, but a continuum, and only a portion of our receptive vocabulary knowledge will ever become productive knowledge (Henriksen, 1999). In terms of written vocabulary, research has demonstrated that concrete lexical items are learned and remembered more effectively when they are presented to learners via both L2 text and images rather than text alone (Mayer 2002; Mayer & Moreno 2003) or than L1 glosses, even in the case of true beginners (Morett, 2019).

Over the last few decades, countless SLA researchers have measured the effects of a variety of types of tasks, interlocutors, instruction, feedback, modalities, and more on the development of their participants’ L2 vocabulary (e.g., Ellis, Tanaka & Yamazaki, 1994; Fernández-Dobao, 2014; Kim, 2008; Rott, 2005; Sasaki & Takeuchi, 2010; Sippel, 2019; Tocaimaza-Hatch, 2016; Morett, 2019). Interaction-based empirical research in particular has demonstrated that learners pay more attention to vocabulary during interaction than to other linguistic features (Plonsky & Loewen, 2013; Williams, 1999) and that lexical items are the focus of more negotiation than grammatical items (Williams, 1999; Mackey, Gass, & McDonough, 2000; Mackey & Goo, 2007). As Ellis, Tanaka, and Yamazaki (1994) highlight, “Whereas learners may not be aware of a grammatical source of their incomprehension, they are much more likely to recognize a lexical source and therefore to seek clarification of its meaning” (p. 456-457). In other words, lexical accuracy is much more essential to effective interpersonal communication than grammatical accuracy. Some empirical studies have also demonstrated that collaborative tasks may result in more vocabulary learning opportunities than individual tasks (e.g., Kim, 2008). Overall, researchers have found that interaction, either between learners (e.g., Adams, 2007; Newton, 1993) or between a learner and a teacher (e.g., de la Fuente, 2002; Ellis
& He, 1999; Ellis et al. 1994; Gass & Alvarez Torres, 2005), has the potential to facilitate vocabulary learning.

Generative models (e.g., Wittrock, 1974), which investigate the effects of generative processing, argue that learning and retention are improved when learners make connections between prior knowledge and new knowledge by using, reformulating, and elaborating the new information. This argument has been supported by numerous empirical studies (Joe, 1995; Joe, 1998; Zaki & Ellis, 1999) that have found that more generative processing leads to increased word gains. Building on these theories, Laufer and Hulstijn (2001) introduced the construct of the “involvement load”, which refers to the cognitive and motivational aspects of a task which promote the establishment and retention of form-meaning connections. A perceived need for a word, combined with a search for its definition and an evaluation of whether the correct syntactic function and word meaning were assigned to the word are the three cognitive processes that are most likely to result in learning and retention of lexical forms.

Looking beyond task design, when it comes to lexical and grammatical L2 development, learners’ depth of processing and level of awareness should also be taken into consideration. Researchers and educators alike can create the ideal task to promote L2 vocabulary learning, but if learners do not minimally pay attention to the new information presented to them and ideally process the information further, learning is unlikely to occur (Leow, 2015). Depth of processing is defined as “the relative amount of cognitive effort, level of analysis, and elaboration of intake together with the usage of prior knowledge, hypothesis testing, and rule formulation employed in decoding and encoding some grammatical or lexical items in the input” (Leow, 2015, p. 204). The concept of depth of processing in SLA comes from Craik and Lockhart’s (1972) levels of processing depth theory, which posited that the possibility that a new grammatical or lexical
form will be stored in long-term memory is determined by the shallowness (e.g., phonemic or orthographic features) and depth (e.g., semantic-associative features) of the learner’s processing. Hulstijn (1992, 2001) has argued that greater depth of processing (also referred to as “elaborate processing”), through inferring and hypothesis-testing, results in improved lexical retention. As learners attend to more features of a lexical item (e.g., orthographic, prosodic, semantic), depth of processing increases. Many empirical studies that have investigated vocabulary learning have corroborated these claims (Fraser, 1999; Rott, 2005; Wesche & Paribakht, 2000). According to Leow (2015), the higher the depth of processing, the higher the potential level of awareness; however, learners only achieve awareness at the level of understanding (the highest level of awareness) when the correct form-meaning connection for the novel lexical item is made.

One common way of measuring depth of processing is through think-aloud protocols, in which participants are asked to verbalize their thoughts and the reasoning beyond their decisions while completing the experimental tasks. Rott (2005) used think-alouds to investigate why certain vocabulary interventions facilitate more lexical development than others. 10 L1 English learners of German read a text enhanced with either multiple-choice glosses (MCGs) or single-translation glosses (STGs). Each target word was glossed once and then appeared three more times in the text. The researcher found that participants who read the text enhanced with MCGs retained notably more lexical knowledge than participants in the STG condition. Form-meaning connections that resulted from MCGs led to the encoding of more word aspects, notably the ability to use the word in a sentence and to recognize its meaning. MCGs also led to ‘more robust entries in the mental lexicon’ since participants in this treatment group continued to evaluate their initial meaning assignments and searched for more semantic clues, thus strengthening the form-meaning connection. Ultimately, her findings supported Laufer and
Hulstijn’s (2001) Involvement Load Hypothesis, which claims that different intervention tasks induce a different involvement with a new word.

In terms of interaction-based vocabulary research, Kim (2008) investigated the effects of peer interaction vs. working alone on vocabulary learning and the production of LREs. 32 L2 learners of Korean completed a dictogloss task with 15 targeted lexical items, half of them in pairs and half of them alone while thinking aloud. The researcher found that participants who worked in pairs produced twice as many LREs as learners working alone and thinking aloud. The participants who worked in pairs also performed significantly better on both immediate and delayed vocabulary posttests. Thus, through negotiation, peer interaction facilitated lexical knowledge.

Ellis, Tanaka, and Yamazaki (1994) also found a positive effect for interaction on vocabulary learning. They investigated the effects of modified interaction on comprehension and lexical development with two classroom studies of 79 and 127 L1 Japanese high school students of English. The researchers found that participants who were exposed to interactionally modified input outperformed the participants who were exposed to premodified input in their recognition and production of the targeted lexical items. They maintained this advantage over a two-week time period (i.e., at the time of the delayed posttest). However, the researchers also found that, in terms of lexical recognition and production, the participants who engaged in active meaning negotiation in the classroom did not significantly outperform those participants who just listened but did not engage with the interactionally modified input. Nevertheless, they noted that some of their findings may have stemmed from participants studying the vocabulary words at home, which they discovered had happened by speaking with the students later on. They ultimately concluded that interactionally modified input facilitates vocabulary learning, but it is not the only
way that new words can be learned. Lexical knowledge can also be acquired through study and memorization (Nation, 1990).

From a sociocultural perspective, Kim (2017) addressed intersubjectivity and appropriation during interaction between 1 native speaker and 2 nonnative speakers of English using conversation analysis. She reported many examples of appropriation, which is when a learner (in this case, the nonnative speaker of English) borrows a linguistic expression used by another party (the native speaker) to fit his own communicative purpose. Thus, through observing interactional practices used by another participant and employing it for his own purposes, the learner expands his communicative repertoire through interaction. Kim states that “this showcases the very practice whereby participants in interaction learn a language by using it. The practice of ‘appropriation’ illustrates how tightly interwoven language learning and use are” (p. 312). However, just as empirical research has demonstrated that not all lexical LREs result in L2 lexical development, Kim explains that not all new lexical items in context are appropriated, given that vocabulary learning sometimes takes place over a more extended period of time. Finally, she argues that

The common sense belief that interaction is a primary site where language learning takes place should be predicated on the understanding of how interaction is constantly driven by an effort to achieve and maintain mutual understanding. Knowledge check questions and definition talk that ensues them are interactional practices motivated by the preference for achieving intersubjectivity in talk-in-interaction (Sacks and Schegloff 1979). Learning happens while participants collaboratively strive to achieve mutual understanding as they estimate each other’s knowledge status and check if mutual understanding is not at risk moment-by-moment (p. 323).

Tocaimaza-Hatch (2016) also investigated L2 vocabulary learning via NS-NNS interaction from a sociocultural perspective. Nine L1 English participants enrolled in an advanced Spanish language course engaged in five oral conversations with a native speaker for a class assignment. These conversations were recorded, transcribed, and analyzed, and participants
were administered a post-test tailored to the LREs that were produced during their individual interactions. The researcher reported that lexical mediation that occurred during interaction resulted in lexical recall at the time of the posttest. However, the researcher also noted that participants’ interactional behavior varied depending on the dyad in terms of their level of engagement and the quantity and quality of mediation. This ultimately impacted participants’ learning outcomes. Tocaimaza-Hatch’s study thus provides support for the potential of L2 vocabulary learning during interaction as well as the positive role of native speaker interlocutors on said learning.

Investigating the effects of peer interaction combined with metacognitive instruction on vocabulary learning, Sippel (2019) recruited 77 L2 learners enrolled in third-semester German language courses. Citing Sato and Lyster (2012), the researcher argued that peer interaction alone was not enough for the development of learners’ L2 vocabulary knowledge. Therefore, form-focused vocabulary instruction was used to direct participants’ attention to a specific set of concrete lexical items before and during peer interaction. In one of the groups, participants were also trained to correct each other’s errors on the targeted lexical items. Sippel’s findings supported her initial argument: both conditions that included form-focused vocabulary instruction promoted vocabulary acquisition, while peer interaction on its own was not effective in supporting vocabulary development. However, it could be argued that the type of tasks and type of interaction utilized might have influenced these findings. For example, on the first day of classroom interaction, the peer instruction group was exposed to the targeted lexical items through discussion in pairs and as a class of a movie and a related reading, while the other treatment groups were given very lexically-specific tasks. These treatment groups completed one activity where they received 15 cards with pictures and 15 cards with the corresponding nouns
and had to match the pictures and the words, and a guessing game where they received 15 cards and had to explain each of the cards to their partner through acting, drawing, or explaining without using the actual vocabulary words.

Many researchers have argued (e.g., Long & Robinson, 1998; Smith, 2003) that not only the visual salience of written discourse, but also the self-paced setting of text-based SCMC, are likely to further facilitate the noticing of L2 forms and may potentially result in subsequent learning. Thus, numerous empirical studies within both the cognitive-interactionist and sociocultural frameworks have also addressed the effects of interaction in SCMC on L2 lexical development (de la Fuente, 2003; Smith, 2004, 2005). For instance, from the cognitive-interactionist perspective, Blake (2000) used jigsaw, one-way information gap, and decision-making tasks with 50 learners of intermediate Spanish divided into dyads. In addition to finding that the jigsaw task resulted in the most negotiation, he noted that the 75% to 95% of the negotiations were lexis-based. In another study, Smith (2003) found that the target items, which in this case were also lexical, might have played a bigger role in the amount of negotiation that occurred during decision-making and jigsaw tasks than the design of the tasks themselves. However, although there is strong empirical evidence supporting the role of interaction in L2 development, both of these studies did not address the development of participants’ receptive or productive vocabulary knowledge.

De la Fuente (2003) investigated the vocabulary knowledge of 20 L1 English L2 Spanish participants who were assigned to either an oral interaction group or a virtual chat group (i.e., text-based SCMC) to complete two timed one-way information gap tasks. 14 concrete nouns were used as the lexical target forms, 7 on each task, and all of the target items were food-related. Overall, both groups improved significantly over time in both their receptive and
productive knowledge on both oral and written vocabulary measures, but both groups also performed better on the receptive measures than on the production measures over time. However, she also noted that the oral interaction group outperformed the virtual chat group on oral and written productive and receptive measures, but the differences between groups was not significant. Nevertheless, it might have played a role that the oral interaction group was allowed to draw pictures while completing the information gap tasks – the virtual chat group only typed to communicate. In addition, it should be noted that the oral and written receptive and productive pretests were conducted only one day prior to the experiment.

From a sociocultural perspective, Sasaki and Takeuchi (2010) investigated vocabulary learning through imitation during interaction via email among L2 learners of English. Their data included a pretest and a posttest, email logs, and a questionnaire. The researchers reported that the imitation that occurred during interaction was not sufficient for vocabulary learning. Furthermore, although participants were introduced to novel lexical forms via interaction, the exposure was not sufficient for lexical retention.

In summary, these empirical studies demonstrate that interaction, whether with a peer, native speaker, or teacher, can facilitate L2 vocabulary learning in a variety of contexts, such as in the classroom, laboratory, face-to-face, or online.
CHAPTER 3: KEY VARIABLES IN THE PRESENT STUDY AND INDICATIONS OF PREVIOUS EMPIRICAL RESEARCH

This chapter offers a review of empirical research that has investigated the main variables addressed in the present study: interlocutor, task, and state anxiety.

**Type of Interlocutor**

In task-based interaction, the interlocutor may be a peer, teacher or researcher, nonnative speaker of the L2, native speaker of the L2, more proficient learner, or less proficient learner, among other possibilities. Over the past few decades, a sizeable body of literature has addressed various interactional patterns resulting from interlocutor characteristics (Gass & Varonis, 1985; Long, 1983, 2007; Mackey, 2012; Mackey, Oliver, & Leeman, 2003; Philp, 2003; Rashidi & Rafieerad, 2010; Wang, 2011). While some studies have addressed interactional differences resulting from interpersonal familiarity (e.g., Lee & Son, 2019), others have examined interlocutor proficiency (e.g., Mackey, Oliver & Leeman, 2003).

In the earliest days of cognitive-interactionist research, studies often described the interactions between native (NS) and nonnative (NNS) speakers (e.g., Hatch, 1978; Long, 1980). This research resulted in Long’s (1981) Interaction Hypothesis, which stated that the interlocutor must be a native or more proficient speaker of the L2 in order for instances of negotiation and subsequent L2 learning to occur. Vygotsky’s (1978) sociocultural theory is premised on a similar idea – for language learning to occur in interaction, the interlocutor must possess knowledge that the learner does not. This means that effective mediation can come from an adult or “a more capable peer” (p. 86). In spite of this, research has progressively diversified over the past few decades, with many studies from both the cognitive-interactionist and sociocultural approaches comparing NNS-NNS and NNS-NS dyads (e.g., Shehadeh, 2003; Varonis & Gass, 1985), and...
also examining peer interaction with no native speaker interlocutor component (Buckwalter, 2001; Gass, Mackey, & Ross-Feldman, 2005; Kasanga, 1996; Kim, 2009; Mackey, Oliver, & Leeman, 2003; Philp, Walter, & Basturkmen, 2010). Although student-teacher interaction has not been widely researched from a cognitive perspective, there have been many sociocultural studies that have addressed this interlocutor relationship (e.g., Dao & Iwashita, 2018; Lantolf & Thorne, 2006; Lantolf, 2012; Nassaji & Swain, 2000; Poehner & van Compernolle, 2011; Sato & Ballinger, 2016). Ultimately, empirical research has demonstrated that L2 learners’ conversations can differ in many ways depending on the interlocutor (e.g., Plough & Gass, 1993).

Many studies that have empirically investigated peer interaction have reported that it does have the potential to facilitate L2 development (Lyster, Saito, & Sato, 2013; McDonough, 2004; Philp & Iwashita, 2013; Sato & Lyster, 2012; Toth, 2008), given that learners feel comfortable providing feedback and expressing when there is a breakdown in communication. Philp, Adams and Iwashita (2014) emphasize that peer interaction is “generally felt to be less stressful than teacher-led interaction, precisely because it will not be carefully monitored” (p. 198). As a result, peer interaction may provide more opportunities for modified output (Shehadeh, 2001, 2003), and thus more learning opportunities (Sato & Lyster, 2007). On the other hand, Mackey et al. (2016) argue that “learners interacting with other learners might not feel that they have the metalinguistic knowledge to provide positive evidence when giving corrective feedback, potentially limiting learners’ exposure to interactions with instructors or learners with higher proficiency levels.” (p. 109). Mackey et al.’s (2016) argument has basis in empirical studies such as Slimani-Rolls (2005), who found that peer interaction resulted in limited negotiation for meaning because the participants did not feel comfortable providing feedback to peers.
Seedhouse (1999) also expressed concerns regarding task-based peer interaction from a sociocultural perspective, noting that there is a tendency for learners to minimize their linguistic output and produce only the language that is necessary for completion of the task.

In spite of these concerns, the general consensus is that the benefits of peer interaction for L2 learners are still apparent when it is directly compared to interaction with a native speaker. Cognitive-interactionist research that has compared NS-NNS and NNS-NNS (peer) interaction has investigated four main interactional features: input modifications, corrective feedback, output modifications, and self-initiated modified output (Loewen & Sato, 2018). For example, Varonis and Gass (1985) employed an open-ended conversation task with learner-learner \((n = 14)\), learner-NS \((n = 4)\), and NS-NS \((n = 4)\) dyads. The NNS participants were adult L1 speakers of Spanish or Japanese and L2 speakers of English studying at the same university. None of the participants in any of the groups knew each other prior to the experiment. The researchers instructed the participants to introduce themselves and find out more about their partner, and these informal interactions were recorded. They reported that negotiation for meaning was most frequent in peer interaction, which they credited to the learners’ perceptions of shared incompetence due to their various first language (L1) backgrounds. However, the small sample sizes for the NS-NS and learner-NS dyads mean that these results should be interpreted with caution.

In Shehadeh (2001), 35 adult participants \((8 \text{ NSs}, 27 \text{ NNSs})\) with varying L1 backgrounds were separated into NS-NNS or NNS-NNS dyads to complete three tasks face-to-face: picture description, opinion exchange, and decision-making. A comparison of the types of dyads demonstrated that peer interaction resulted in significantly more modified output. Similarly, Sato and Lyster (2007) investigated the interactional moves of 8 L1 Japanese learners
of L2 English and 4 speakers of L1 English. NNS participants completed tasks with both the NS participants and with each other. The results revealed that peer interaction resulted in more modification moves, and as a result, created more learning opportunities. Finally, Sato (2015) employed two picture description jigsaw tasks with learner-learner and learner-NS dyads. Participants were 8 L1 Japanese learners of L2 English and 4 speakers of L1 English. He analyzed the interactions for density and complexity and found that learners and NSs produced similar amounts of output, but learners produced longer utterances with more vocabulary and verb types in the peer interaction context. Overall, when interacting with another learner, it would seem that although learners are exposed to less modified input, there are more opportunities for them to modify their output, thus facilitating L2 development (Sato & Lyster, 2007; Sato, 2015; Shehadeh, 2001, 2003; Varonis & Gass, 1985).

In another study that addressed learner proficiency, Mackey, Oliver, and Leeman (2003) addressed negotiation for meaning and corrective feedback with 2 native English speakers, 2 L1 Chinese L2 English speakers with high proficiency, and 2 L1 Chinese L2 English speakers with low proficiency. These participants were separated into 14 different dyads and a pretest-treatment-posttest design was utilized. The researchers found that the most successful dyads were the mixed proficiency peer dyads, with one high proficiency NNS and one low proficiency NNS. Furthermore, participant errors decreased from pretest to posttest, particularly with the low proficiency NNS. Nevertheless, they also reported that the native speaker interlocutors provided more feedback than the nonnative speaker interlocutors. This observation supported the findings of Oliver (1995), who also asserted the critical role of native speaker interlocutors in the provision of negative feedback. Ultimately, Mackey et al. (2003) argued that these findings offer support for the role of corrective feedback and negotiation for meaning in L2 development.
There have also been numerous empirical studies that have addressed peer interaction outside of the cognitive-interactionist framework. Overall, these studies have suggested that NNS-NNS interaction is beneficial to learners, given that they tend to negotiate for meaning with other learners more often than with native speakers, subsequently benefitting their language development (Philp, Adams, & Iwashita, 2013; Porter, 1986; Varonis & Gass, 1985). Storch (2002) identified four patterns of dyadic interaction among peer interlocutors in the classroom: 1) collaborative, 2) dominant/dominant, 3) dominant/passive, and 4) expert/novice. Her analysis showed that the collaborative pattern was the most conducive to language development. On the other hand, the dominant/dominant and dominant/passive patterns did not facilitate language learning. In one of her more recent studies (Storch, 2017), she found that in tasks focusing on form, mixed proficiency pairs have the potential to improve the interlanguage of both members of the dyad, but only when they form an expert/novice relationship.

Donato (1994) was one of the first empirical studies to address whether scaffolding—traditionally attributed to expert-novice dyads—occurred in peer interaction. 3 L1 English learners of French worked together in a group to complete an open-ended classroom task. Prior to completing the task, they spent one hour preparing for it. Donato analyzed this hour-long interaction for instances of scaffolding, and found that “collaborative work among language learners provides the same opportunity for scaffolded help as in expert-novice relationships in the everyday setting” (p. 41).

In a study that addressed amount of talk on task, Poupore (2016) examined interaction during 15 different interactive tasks among 10 L1 Korean L2 learners of English with intermediate to high intermediate proficiency. The main purpose of the study was to introduce a measurement of participants’ group work dynamic (which the researcher measured during
interaction using a points system), and to examine the relationship between this variable and the amount of talk produced on task. He found a significant correlation between participants’ group work dynamic and the amount of talk produced on task; thus, groups who had a better group work dynamic produced more output. Citing Swain (2005), Poupore argues that “while producing more language does not necessarily entail a direct link with acquisition, it may increase its chances” (p. 731).

Foster and Ohta (2005) took both a cognitive-interactionist and sociocultural approach to examining the interactions of learners using two sets of data. The first data set came from 20 L2 learners of English with varying L1 backgrounds, and the second data set came from 19 L1 English learners of L2 Japanese. From a cognitive-interactionist standpoint, instances of negotiation for meaning were very low. However, a qualitative, sociocultural analysis of the data indicated that, during the periods where no negotiation for meaning took place, learners were actively helping each other to complete the task via co-construction and prompting. In the absence of communication breakdowns, learners both sought and provided assistance and initiated self-repair of their utterances. Ultimately, maintaining a supportive and friendly interaction for environment was more important to the participants than the comprehensibility of the input. This analysis of participants’ interaction lends support for socioculturalists’ arguments that quantitative methods, often employed in the cognitive-interactionist framework, can obscure important interactional features. From a cognitive-interactionist perspective, it is also argued that peer interaction offers learners a greater comfort level, which can help them notice their interlocutor’s errors, correct those errors, and modify their own errors upon receiving interactional feedback. This feeling of comfort may also help learners to produce more linguistic output (Loewen & Sato, 2018).
Many empirical studies within the sociocultural framework have also addressed whether type of interlocutor can affect the frequency of LREs (e.g., Fernández-Dobao, 2012; Leeser, 2004; Storch, 2002; Watanabe & Swain, 2007), often finding that different interlocutor characteristics can impact the pattern of exchanges in dyadic interaction. However, a variable not often considered is interlocutor familiarity. Lee and Son (2019), an empirical study situated within the cognitive-interactionist framework, is one of very few studies that has addressed this variable by examining 71 transcribed interactions in both learner-NS and learner-learner dyads in relation to the frequency of lexical LREs. The participants had varying degrees of familiarity with each other in the classroom setting. The analyses indicated that the number of LREs between familiar and unfamiliar interlocutors was similar in all conditions. This led the researchers to conclude that “there may be little influence of interlocutor – learner or NS, familiar or unfamiliar – on the frequency of LREs, which suggests that learners negotiate meaning in similar ways regardless of whom they interact with” (p. 163). This conclusion contradicts the findings of previous studies on interlocutor familiarity (e.g., Cao & Philp, 2006; Gass & Varonis, 1984) who found that learners are more willing to communicate, are more confident, and better comprehend each other when interacting with familiar peers. These discrepancies in results may perhaps suggest that not all tasks between interlocutors who either know or don’t know each other are the same, and perhaps not all social relationships of knowing or not knowing one’s peers are the same. Ultimately, Lee and Son (2019) brought into question the influence of type of interlocutor on the features of interaction, and this merits further investigation.

Peer interaction has also been widely investigated in the SCMC mode (e.g., Blake, 2000; Smith, 2003; Yilmaz & Granena, 2010), and it seems to offer similar opportunities for L2
learning as the FTF mode, but the amount and types of negotiation can depend on factors such as type of task, type of targeted linguistic item, participants’ L1, and patterns of negotiation (e.g., expert-novice) (Baralt, 2013; Chun, 1994; Liu, 2017; Yilmaz & Granena, 2010; Yilmaz, 2011; Ziegler, 2016). For instance, Bueno-Alastuey (2013) investigated the number and type of LREs, the kind of LRE signals, and the amount of modified output with 42 dyads: 14 NNS-NNS sharing the same L1, 14 NNS-NNS with different L1s, and 14 NNS-NS. The nonnative participants were learners of English with varying proficiency levels. The results revealed that the NNS-NNS dyads with different L1s exhibited the greatest number of LREs and modified output, and the NNS-NNS dyads with the same L1 exhibited the lowest number of LREs and modified output. In consequence, Bueno Alastuey (2013) concluded that “the typical arrangement in FL contexts (NNS-NNS same L1) seems to contribute the least to language acquisition, at least in two-way information exchange tasks” (p. 555). Although Bueno Alastuey’s (2013) study did have small sample sizes, the varying proficiency levels of the participants were not considered when interpreting her results. Nevertheless, these findings bring into question the conclusions drawn from other studies where participants had varying L1 backgrounds, such as Shehadeh (2001, 2003), since the effects of participants’ linguistic backgrounds could have potentially been more significant than their status as peer interlocutors.

In another SCMC study, Liu (2017) investigated the frequency of instances of negotiation, successfully resolved instances, and interactional strategy in NNS-NS and NNS-NNS dyads. This study was particularly unique in that it addressed the effects of not only native versus nonnative interlocutors, but also the proficiency level of the nonnative interlocutors. NNSs were native speakers of Mandarin learning English, and the NSs hailed from various parts of the English-speaking world. NNS-NNS dyads were divided into five H-H pairs and five L-L
pairs, with H and L representing high- and low-proficiency learners. Similarly, NNS-NS dyads were made up of five H-N pairs and five L-N pairs, with N being native speakers. The dyads collaborated via Facebook chat to complete reading tasks synchronously. The results revealed that negotiated instances, successfully resolved instances, and interactional strategy use took place most frequently in H-H pairs. Furthermore, while H-H-, H-N and L-N dyads all employed interactional moves such as clarification requests and confirmation checks, the L-L dyads most often asked their partners for assistance. Ultimately, Liu (2017) found that negotiation was more frequent in peer interaction, which contrasted with the findings of Bueno-Alastuey (2013) but paralleled the findings reported in many FTF studies (e.g., Varonis & Gass, 1985).

From the sociocultural perspective, Zeng and Takatsuka (2009) investigated peer-peer collaborative dialogue in text-based SCMC with 16 learners of Chinese divided into 8 dyads. Participants completed four collaborative tasks via Moodle, a course management system, and their dialogue was analyzed for the occurrence of LREs. Participants’ language learning was measured using an immediate posttest and a delayed posttest that were developed based on the identified LREs, resulting in 150 test items. The researchers found that learners assisted each other in attending to language forms, which subsequently improved their performance on the study’s measures of language learning.

Although interaction research in SCMC has empirically demonstrated that peer interaction on its own can be facilitative of L2 learning (Blake, 2000; Pellettieri, 2000; Smith, 2003; Yilmaz and Granena, 2010; Yilmaz, 2011; Zeng, 2017; Kim, 2017), the benefits of task-based peer interaction in SCMC when compared to interaction involving other types of interlocutors is less clear, at least from a cognitive-interactionist perspective. For instance, as discussed above, Bueno-Alastuey (2013) and Liu (2017) represent two of a limited number of
empirical studies addressing interaction in SCMC that have incorporated NS-NNS and NNS-NNS dyads in their research designs, but the researchers ultimately made contrasting conclusions regarding the effects of type of interlocutor on instances of negotiation, and they did not measure L2 development.

The comparison of expert (native speaker or teacher) and novice (learner) interlocutors, and more specifically, the incorporation of teacher interlocutors into research designs, is one that deserves attention in task-based interaction in SCMC for many reasons. As Avermaet, Colpin, Van Gorp, Bogaert and Van den Branden argue, “Still the expert (and acknowledged as such by the pupils), the teacher uses his knowledge in a very strategic way: he uses it not to give clear-cut answers, but to ask well-chosen questions that guide the learners’ active knowledge construction.” (2006, p. 189). Teachers can be the ‘more knowledgeable’ partner without explicitly asserting their role as expert in the interaction, and they may encourage instances of negotiation for form and modified output by using recasts, confirmation requests, and clarification requests. Furthermore, teacher interlocutors can guide learners to successful task completion through negative feedback when learners formulate incorrect hypotheses. They can also support learners on the affective level by encouraging them to stay motivated to complete the task (Avermaet et al., 2006). Other sociocultural researchers have offered support for Avermaet et al.’s arguments. For instance, Lantolf and Thorne (2006) have suggested that teacher mediation, defined as assistance that addresses learners’ language problems (Dao & Iwashita, 2018), can help learners develop better understanding when working towards the appropriation of language forms. Teacher mediation both assists learners in completing the task at hand and offers scaffolding in response to learners’ linguistic gaps (Poehner & van Compernolle, 2011).
On the other hand, socioculturists also argue that learner-led interaction can result in more creativity, self-regulation, and linguistic autonomy, characteristics which are necessary for the mastery of the L2 (Brooks, Donato, & McGlone, 1997; Donato, 1994; van Lier, 1996). Furthermore, as Sato and Ballinger (2016) highlight,

Despite the fact that native speakers are more likely to provide richer syntactical and lexical input, and learners may not always reach a target-like solution to language problems that arise during peer interaction, it seems that during peer interaction learners are able to shift their attention to formal aspects of the target language more than when they interact with native speaking partners or teachers. It is also clear that they engage in quantitatively richer interaction with each other in terms of output. (p. 4)

Although many early classroom-based studies addressed negotiation of meaning and comprehensible input in teacher-learner interaction (Chaudron, 1986; Long & Sato, 1983; Pica, 1987, 1991; Pica & Doughty, 1985, 1988; Pica & Long, 1986; Rulon & McCreary, 1986), only a handful of studies have empirically compared the effects of learner-learner versus learner-teacher dyadic interaction (e.g., Van den Branden, 1997), and all of them have been conducted in the FTF mode. In classroom-based studies that have examined learner-teacher and learner-learner interaction, it has been demonstrated that learners provide feedback and engage in negotiation more often with each other than when interacting with the teacher (Alcón, 2002; Toth, 2008). Some researchers within the sociocultural framework have posited that this is because peer interaction is less stressful than interaction with a teacher since it is not carefully monitored (Philp, Adams, & Iwashita, 2014). For instance, Alcón addressed the development of requests in learner-learner versus learner-teacher interactions with L1 Spanish L2 learners of English. Two groups of 12 learners were instructed in the use of requests. Participants were then given role-play contexts that prompted the use of requests. The findings indicated that learners in both groups improved their use of requests, but the participants in the learner-learner group engaged in more interaction.
In another empirical study that addressed the effects of type of interlocutor (teacher or learner), Zhao and Bitchener (2007) examined features of incidental focus on form (types of focus on form; types of feedback; linguistic forms focused on and types of immediate uptake) present in 10 two-way information tasks. The study produced a number of significant results. First, they found that teacher interlocutors were more active than learner interlocutors in responding to learners’ errors. In addition, learners were more likely to ask questions of each other than of their teacher. The researchers noted that, in both types of interactions, incidental form-focused episodes occurred frequently, facilitating opportunities for L2 learning. While these findings support the use of both types of interaction for L2 learning, it should be noted that learner-learner interaction was dyadic in nature, while the learner-teacher interaction was conducted via whole class discussion, which could have potentially skewed the results of the study.

Toth (2008) used both quantitative and qualitative measures to examine learner-learner vs. learner-teacher task-based interaction in the classroom setting with 78 L1 English beginning learners of L2 Spanish. The treatment involved five 50-minute lessons targeting the anticausative clitic se. Recordings of the treatment sessions were transcribed and participants were administered a grammaticality judgment task and guided production task in a pretest-posttest-delayed posttest design. The qualitative, transcribed data revealed that teachers had the potential to facilitate language learning by offering assistance when learners processed output and by directing their attention to target forms. The quantitative data further suggested that participants who interacted with a teacher rather than with other learners performed better on both tasks. However, similar to Zhao and Bitchener (2007), it should be noted that learners interacting with each other did so in small groups, while teacher-led instruction was conducted as a class. In Toth
(2008), the learners in small groups definitely showed evidence in the transcripts of focusing on form and providing scaffolded assistance during their interactions. However, the focus of their attention to L2 form was much wider than the teacher-led whole class group, which focused almost exclusively on the target structure. It’s not surprising then, that on assessments of target structure knowledge, the gains in the classes with small-group interactions were not as strong as the whole-class group for knowledge and use of the target structure.

Also from a sociocultural perspective, Dao and Iwashita (2018) addressed teachers’ language mediation and task-related assistance, and how these variables are affected by learners’ responsivity in classroom interaction. 16 L1 Chinese L2 learners of English completed meaning-focused tasks in the classroom setting over a period of four weeks. These sessions were recorded, transcribed, and analyzed for language mediation and task-related assistance. The results indicated that the teacher provided task-related assistance, such as task clarification, modeling, eliciting, and direction, and also varying degrees of task collaboration. All instances of highly collaborative language mediation were regarding lexical items. These instances of high collaboration resulted in greater opportunities for the teacher and learner to work together to co-construct knowledge of the lexical item, leading to more successful self-regulation. The researchers thus suggested that teacher-learner collaboration was key to the appropriation of L2 lexical forms. Overall, sociocultural research seems to indicate that teachers should not be overlooked as facilitators for L2 language development learning during interaction.

One of the only studies to compare learner-learner and teacher-learner dyadic interaction is Van den Branden (1997). He addressed the effects of different types of negotiation on the output of 48 child learners of Dutch. Some of the participants were NSs (slightly over half) and others were NNSs of Dutch, although the numbers were not explicitly quantified. The
participants completed an oral production task where they had to describe a series of images to an interlocutor. It was found that the learner-learner and teacher-learner dyads produced similar amounts of negotiation, with variables such as language proficiency and type of feedback received having greater effects on the amount of negotiation than the type of interlocutor. The least negotiation often occurred in learner-learner dyads where both participants possessed low proficiency, an observation that has been supported by more recent studies (e.g., Liu, 2017). Van den Branden noted the following regarding the results:

The fact that some of these pairs only negotiated minimally during their interaction underscores that no task in itself holds any guarantee of negotiation taking place. However, owing to the lockstep-type of education they usually received, the pupils were not used to depending on another pupil, rather than the teacher, for crucial information. (p. 613).

Van den Branden also stated that he (as the teacher interlocutor) engaged in six times as many content negotiations and twice as many meaning negotiations with low-proficiency learners compared to high-proficiency learners. However, these differences were not statistically significant.

While Van den Branden provides important insight into learner-teacher task-based interaction, his research is not without limitations. For instance, the study provided no empirical evidence for the effects of negotiation on subsequent language output. In addition, the fact that the learners knew each other but did not know the teacher (i.e., Van den Branden) may have altered the amount and types of negotiation observed (see Plough & Gass, 1993). Finally, although this is not a limitation of the study itself, given the age of the participants in this study, the findings are not necessarily applicable to interactions between adult L2 learners and teachers, especially since the power differential between these groups may be different. From the sociocultural perspective, there have been a few studies that have investigated collaboration
between expert-novice dyads (e.g., Storch, 2001; Watanabe & Swain, 2007). These studies found that the nature of the relationship within dyads had the potential to significantly affect learners’ language production, with expert and novice interlocutors negotiating breakdowns in communication and equally contributing to task completion.

In the past, SLA researchers have measured interlocutor power through amount of talk (output), disagreement, interruption patterns, turn unit, fillers, topic initiations, and backchannels among others (Beebe & Giles, 1984; Chiang, 2013; Gass & Varonis, 1985; Itakura, 2001; Thonus, 1999, 2002; Woken & Swales, 1989). For instance, Hardman and Hardman (2016) reported that, in comparison with student-teacher interaction, peer interaction is important given that it creates a more symmetrical environment in terms of status and power, which is more conducive to the co-construction of knowledge. However, in many cognitive-interactionist studies, power is often overlooked as an interfering variable. For example, Baralt and Gurzynski-Weiss (2011) used teacher-learner dyads to examine the effects of mode (CMC vs. FTF) on learners’ state anxiety. Twenty-five learners of Spanish enrolled in intermediate university-level courses completed two information-gap tasks with a teacher, one in each modality. The principal findings revealed that state anxiety was comparable across modality. However, citing Robinson and Gilabert (2007), the researchers pointed out that using a teacher as the interlocutor may have affected the results of the study, since the teacher’s position of power could have caused them greater anxiety. They concluded by suggesting that future studies might yield different outcomes by not using a teacher as an interlocutor, or by pairing students of similar or mixed proficiency. While these findings contribute to the understanding of modality effects on affective factors, L2 development of the participants was not measured.
In summary, although type of interlocutor has been widely addressed by sociocultural researchers with varying outcomes, cognitive-interactionist research on teacher-learner versus learner-learner dyadic interaction is particularly scant in the FTF mode, but nonexistent in the SCMC mode. Furthermore, although many cognitive-interactionist studies have addressed NS-NNS vs. NNS-NNS interaction in the FTF mode, an empirical comparison of these types of dyads has not been as widespread in the SCMC mode, and the studies that have addressed this variable have often produced conflicting results. However, given that aspiring language learners are increasingly turning to online teachers and tutors, who may or may not be native speakers, to achieve their linguistic goals, it is essential that researchers consider teachers’ and native speakers’ perceived linguistic authority and its potential impact on interactions with students and the subsequent L2 development of these students. The present study hopes to address whether type of interlocutor can differentially affect the occurrence of LREs, amount of talk on task, learners’ performance on receptive and written productive measures, and whether there exists a relationship between type of interlocutor and type of task and learners’ state anxiety in text-based SCMC.

**Type of Task**

In order to fully understand the potential effects of type of task on L2 learners’ output, amount of negotiation, and linguistic development, we must first detail the types of tasks that have been used in interaction studies. Pica, Kanagy, and Falodun (1993) defined five types of communicative tasks that are commonly employed in interaction-based research: jigsaw, information gap activity, problem-solving, decision-making, and opinion exchange. A jigsaw task, also known as a two-way information gap task, involves two interactants who hold two
different sets of information that together form one whole. Their information must be exchanged as they work towards a common task goal. Given that no one learner has all of the information needed to complete the task, interaction from both learners is required. The second type of task is a one-way information gap task. The one-way information gap task is similar to a jigsaw task; however, only one learner holds important, task-relevant information that is required by the other learner in order to complete the task. Therefore, the learner not holding the task-relevant information must request this information from the other learner. In this case, the information-requester has more opportunities for seeking input modification, while the information-giver has more opportunities to receive feedback on his or her production. A problem-solving task is where there is only one resolution to a problem, and the learners must work together to reach it. In a decision-making task, although learners are expected to work towards a single outcome, there are many alternative outcomes that are available. Finally, an opinion exchange task involves the discussion and exchange of ideas.

Pica et al (1993) have suggested that the jigsaw and information gap tasks are those most facilitative of second language learning, since they “generate opportunities for interactants to work toward comprehension, feedback, and interlanguage modification processes.”(p. 181). This is due to four reasons: 1) each learner holds a different portion of information that must be exchanged in order to successfully complete the task; 2) both learners must supply and request this information; 3) the learners have convergent goals; and 4) there is only one acceptable outcome for successful task completion. However, empirical research since their seminal chapter has suggested that it is not always the case that information gap tasks are the most facilitative of L2 learning. Furthermore, in addition to task design, factors such as required procedures, conceptual content, social relationships among interlocutors, and task familiarity
must be considered when evaluating L2 learning outcomes. Foster and Ohta (2005) have argued that “A developmentally appropriate topic and task, a desire to express oneself, a supportive listener, a friendly and non-face-threatening environment in which to monitor one’s own output—these are likely ingredients of a task that is good for promoting progress in the target language, whether the information exchange is one way, two way, required, optional, or any other task type label” (p. 426).

The types of tasks employed in interaction-based research have varied. For example, in Slimani-Rolls (2005), there were 20 participants from varying L1 backgrounds who also spoke L2 English but were taking a university-level course on French for business purposes. They completed a consensus task (a type of jigsaw task), a one-way information exchange task, and a two-way information exchange task. She found that the two-way information exchange led to the most negotiation for meaning; however, there was a great deal of individual variation, given that more than 50% of the instances of negotiation for meaning came from only three learners. Upon interviewing the participants, she learned that many were more focused on completing the task rather than resolving any breakdowns in communication, so an understanding of task goals affected frequency of negotiation. Some participants also mentioned that they chose not to negotiate for meaning out of politeness, meaning that participants’ sense of social identity and a concern with how they presented themselves also affected negotiation. Thus, she concluded that:

The inherent characteristics of two-way task design seem to focus learners’ attention on getting the missing information without engaging in much meaning negotiation. Where meaning negotiation happens, the use of interactional modifications is apparently ineffective, lexically based and quantitatively easily inflated (p. 208).

Her findings also led her to argue that, contrary to the suggestions of Pica et al (1993), one-way and decision-making tasks offer more opportunities for genuine communication and linguistic manipulation.
Ultimately, Slimani-Rolls (2005) suggested that individual differences must be considered when choosing tasks for interaction-based research. This suggestion is still very much relevant today, as Long (2015) has argued that the key feature of any task is its relevance to learners’ communicative needs, and the best tasks are locally produced, given that teachers and program designers typically have a greater understanding of their students’ needs when compared to an unknown researcher. However, while Slimani-Roll’s findings certainly contributed to interaction-based research, the identity of her participants as already experienced language learners calls into question the generalizability of her findings. In other words, perhaps because of their previous experience learning L2 English and thus their identity as experienced language learners, participants may have approached the task differently in their L3, focusing on completion rather than communication.

In another empirical investigation into the effects of task type, Gass et al. (2005) recruited 74 participants enrolled in third semester university-level Spanish courses to complete three communicative jigsaw tasks: a map task, a picture differences task, and a consensus task. The results revealed that significantly more negotiation took place on the map task and the picture differences task. They attributed their findings to a principle characteristic of the tasks: whether or not they required the exchange of information. For the consensus task, participants only had to give their opinion about information that both participants shared. For the map and picture differences tasks, the interlocutors possessed different information that they needed to exchange in order to successfully complete the tasks. The researchers also concluded that task type had a greater effect on amount of negotiation than the setting in which the experiment took place (classroom vs. laboratory). However, the complexity of the tasks and the proficiency of the participants are variables that should be considered but were not discussed.
In a design that addressed both type of task and L2 lexical development, Newton (2013) recruited two groups of four adult learners of English to complete two information gap (convergent) and two opinion gap tasks (divergent). He found that learners were less likely to negotiate for meaning on the information gap tasks than on the opinion gap tasks. He observed that the learners completed the information gap tasks very ‘mechanically’, while for the opinion gap tasks, the learners had to focus on the meaning of each other’s output in order to successfully complete the tasks. In terms of L2 development, Newton reported gain scores of approximately 4 words per every 30 minutes on task. This led him to argue that ‘communication tasks [are] a potentially valuable source of opportunities for incidental vocabulary learning’ (p. 165).

Many task-based interaction studies have considered the effects of collaborative tasks in terms of the occurrence of LREs, which can lead to L2 development because, it is assumed learners notice holes in their interlanguage, they notice certain forms in their interlocutor’s speech by repeating them, they self-repair, they engage in metalinguistic talk, and they check and question their hypotheses about the target language (Swain & Lapkin 1995, 1998, 2001; Swain 1995, 1998; Lyster & Ranta 1997; Lyster 1998; Alegría de la Colina & García-Mayo 2007). One of the first studies to empirically investigate the relationship between LREs and L2 development was LaPierre (1994), who examined the instances where specific language items were discussed during interaction (later referred to as LREs) while learner participants collaborated in dyads to complete a dictogloss task. A dyad-specific delayed posttest was administered to participants one week after the experiment, and she discovered that participants retained 80% of the lexical forms from correctly resolved LREs, and 70% of the lexical forms from LREs that were resolved incorrectly. In a similar vein, Swain and Lapkin (1998) analyzed the interactions of 12 dyads of
French immersion students while they completed a jigsaw task. The researchers found a positive relationship between the number of LREs produced and participants’ scores on the posttest.

In terms of the effects of task type on the production of LREs, García-Mayo (2002) used 5 different tasks with 7 peer dyads: a cloze task, multiple-choice task, text construction task, text editing task, and a dictogloss task. The researcher found that LREs only represented 14% of turns on the dictogloss task, which could potentially be contributed to the fact that all other tasks were written, while the dictogloss task contained an oral component. In addition, contrasting with previous empirical research in both the sociocultural and cognitive-interactionist frameworks, the researcher found many more grammatical LREs than lexical LREs, which may have been a result of the content of the tasks.

From a cognitive-interactionist perspective, Gilabert, Baron, and Llanes (2009) addressed task type and LREs using a narrative reconstruction task, an instruction-giving map task, and a decision-making task. Two different versions of each task, one simple and one complex, were completed by 60 learners of English. The researchers found no significant differences among the different task types regardless of the conditions under which they were performed. In contrast, Kim (2009) reported that low proficiency learners produced significantly more LREs on a simple picture narration task, whereas high proficiency learners produced more LREs on a complex picture narration task. The latter finding was in line with Robinson’s Cognition Hypothesis (2001), which says that increasing task complexity should result in more instances of negotiation and therefore opportunities for learning, while Kim attributed the former finding to the learners’ proficiency.

Thus, it would seem that not all jigsaw and information gap tasks are created equal in terms of their potential for promoting negotiation. Researchers must take a myriad of factors into
account when designing task-based interaction studies utilizing more than one task. However, task complexity is difficult to control for unless it is being addressed directly by the experimental research design because it is affected by learners’ familiarity with the topic and procedures involved. Furthermore, although the present study only addresses task type and not task complexity, it is difficult to say whether or not the information gap task employed here (or in any other study) is more complex than the ones that have been employed by other studies investigating task effects. Another problem from a social standpoint is that task-based researchers tend to employ methods that positively distort the effects of the tasks used. Seedhouse (1999) argues that “A quantitative, segmental methodology has been used which isolates and counts individual features which happen to be abundant in task-based interaction. It is then claimed that these individual features are particularly conducive to second language acquisition, from which it follows that task-based approaches are particularly conducive to second language acquisition” (p. 154). Ultimately, he argues, this type of methodology is “self-fulfilling”.

In computer-mediated communication (CMC) in particular, which task type elicits the most negotiation still remains to be seen (Ortega & Gonzalez-Lloret, 2016). A problematic pattern in the accumulated research on task-based interaction in SCMC is that the amounts of negotiation for meaning and negative feedback are highly variable, ranging from 3% to 30% of turns (Ortega, 2009). Ortega and Gonzalez-Lloret (2016) suggest that this may be the result of the type of tasks utilized in the experimental design. Numerous studies in SCMC have demonstrated that task design can significantly impact the production, output, instances of negotiation, and development of L2 learners (e.g., Baralt, 2013; Blake, 2000; Smith, 2003; Yilmaz, 2011; Yilmaz & Granena, 2010).
Despite the importance of task effects, only a small number of studies have explored the effect of task type on learner language in text-based SCMC contexts. With the exception of Keller-Lally’s (2006) dissertation, all other CMC studies in ISLA have reported that tasks can affect the amount of negotiation that occurs in dyadic interaction. For instance, in one of the first studies to investigate Pica et al.’s (1993) task predictions in text-based SCMC, Blake (2000) used jigsaw, one-way information gap, and decision-making tasks with 50 learners of intermediate Spanish divided into dyads. Contrasting with Pica et al. (1993) and the findings of many FTF interaction-based empirical studies, he found that the jigsaw task elicited the greatest number of instances of negotiation. However, instances of negotiation only represented between 0.3% to 3.8% of the overall conversational turns. Furthermore, although the jigsaw tasks produced significantly more negotiations than the other tasks, the effect sizes were extremely small: “.067 and .102 for the fall and spring, respectively” (p. 127). In addition, the large majority of negotiations (75% and 95%) resulted from lexical confusions, with miscommunications regarding phonology, morphology, and syntax ‘few and far between’. Thus, although the study was able to successfully demonstrate that negotiation does occur in CMC, whether these negotiations were significantly affected by the type of task, and not by other interfering variables such as the content of the task, is questionable.

Pellettieri (2000) also investigated the negotiation of form and meaning in SCMC with intermediate learners of Spanish as they completed jigsaw and guided conversation tasks. Her findings indicated that the interactional features present were similar to those found in FTF interaction, with learners offering corrective feedback and employing interactional moves such as clarification requests and confirmation checks in response to a lack of understanding or an inappropriate response. She also concluded that goal-oriented tasks that included items, concepts,
or vocabulary that the learners were unfamiliar with had the potential to increase negotiation between learners. Finally, she suggested that text-based SCMC may be more beneficial to learners than FTF interaction, given that learners are provided with a visual representation of their input, allowing them to attend to their language and complex grammar structures.

The relevance of the content of the task also comes into play upon closer examination of Smith (2003). He employed 14 learner-learner dyads to address the amount and types of negotiation that occurred during 2 jigsaw and 2 decision-making tasks that were “seeded” with 8 target lexical items. The results revealed that the amount of negotiation was once again mediated by type of task, with the decision-making task eliciting more instances of negotiation and focus on forms than the jigsaw task. However, as Smith highlights, “There was some evidence that jigsaw tasks may elicit more incidental negotiation, as predicted by Pica et al. (1993), but when target lexical items are infused into the task, decision-making tasks yield more negotiation sequences than jigsaw tasks.” (p. 52). Thus, perhaps it is not the tasks themselves, but rather the researchers’ conscious or unconscious manipulation of the content of those tasks that results in more negotiation. Furthermore, it is impossible to directly compare the results of many SCMC-based interaction studies, given that they have not all used the same tasks in the same conditions and with similar participants for their research design. For example, unless another study uses jigsaw, one-way information gap, and decision-making tasks with intermediate L2 learners of Spanish with similar proficiency and curricular backgrounds, there are too many interfering variables to allow Blake (2000) to be directly compared.

In more recent research, both Yilmaz and Granena (2010) and Yilmaz (2011) addressed peer negotiation in text-based SCMC through the use of two tasks: a jigsaw task and a dictogloss task. Participants in Yilmaz and Granena (2010) were 10 adult intermediate learners of L2
English who worked in dyads to complete both tasks. Interactions were coded for number and characteristics of LREs. The researchers reported that although the dictogloss task resulted in more LREs than the jigsaw task, the LREs were characterized by a range of incorrect, correct, and unresolved outcomes. The jigsaw task was ultimately superior since it resulted in more implicit LREs that did not lead to incorrectly solved outcomes. Yilmaz (2011) studied the task-based interactions of 54 L2 English learners using the same tasks. He again found that the dictogloss task resulted in a higher number of LREs than the jigsaw task. However, in this case the dictogloss task was deemed superior, since the LREs that occurred during this task were described as orthographic, negative feedback, and solved correctly, while the jigsaw task often resulted in unresolved LREs. These conflicting findings could potentially be attributed to the small sample size in Yilmaz and Granena (2010), which affects that study’s generalizability.

Producing similar results to Yilmaz (2011), Zeng (2017) recruited 32 L1 Chinese English learners who completed both a dictogloss and jigsaw task in either SCMC or FTF modes. Learners’ exchanges were analyzed for LREs, and the results revealed that the dictogloss task produced significantly more LREs than the jigsaw task in both SCMC and FTF. In terms of the effects of mode, Zeng reported that not only were LREs more frequent in SCMC than in the FTF mode, but the SCMC LREs were characterized as having ‘orthographical, correct, and self-correction outcomes’, whereas the FTF LREs had ‘incorrect and request for assistance outcomes’. Ultimately, a review of the studies that have addressed the relationship between type of task and amount of negotiation suggests that their findings have varied, and variables such as task content and context must be considered upon comparing these studies.

In terms of the relationship between type of task and learning outcomes in SCMC, Baralt (2013) employed a pre-posttest design to measure L2 learning of the Spanish past subjunctive
forms. 84 L1 English learners of L2 Spanish completed tasks varying in their degree of cognitive complexity while interacting one-on-one with a researcher who provided corrective recasts on the Spanish past subjunctive. Interestingly, she found an unexpected interaction of medium with task complexity. Participants in the FTF group who performed the cognitively complex task exhibited significantly more learning of the Spanish past subjunctive form than those who performed the simple task, while in the SCMC group the opposite effect was observed. While the SCMC group participants who had to perform the cognitively complex task did not learn much at all, the SCMC participants who completed the simpler task performed best out of all the groups. According to her results, the design and the complexity of the task played a crucial role in learning outcomes. Thus, research designs that include tasks carried out solely in SCMC must be chosen carefully and must be empirically and theoretically supported so as to avoid these sorts of complications.

In conclusion, while the findings of some SCMC-based studies have suggested that jigsaw tasks produce the greatest amounts of negotiation (e.g., Blake, 2000; Yilmaz & Granena, 2010), others have suggested that decision-making tasks (e.g., Smith, 2003) or dictogloss tasks (Yilmaz, 2011; Zeng, 2017) are superior for increased instances of negotiation. In terms of the effects of type of task on L2 development in SCMC, the dearth of studies, combined with the lack of consideration of the impact of other factors such as social relationships among interlocutors, means that the jury is still out.

**State Anxiety**

One potential explanation for L2 learners’ variable success stems from their individual differences, which are the affective and cognitive variables that make each learner unique.
Foreign language anxiety (FLA), defined as “the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening, and learning.” (MacIntyre & Gardner, 1994b, p. 284) is a constant presence in the L2 classroom that possesses the potential to obstruct the learning process (Arnold & Brown, 1999). Given that it has consistently correlated negatively with success in L2 learning (Horwitz, 2001), teachers, researchers, and learners alike consider it a variable of interest. FLA’s negative effects can be explained through its interference with both social and cognitive aspects of L2 learning. As FLA increases, learners are less able to process input, make connections between prior knowledge and new information, and access their linguistic system during L2 production (e.g., MacIntyre, 1999; MacIntyre & Gardner, 1991; Onwuegbuzie, Bailey, & Daley, 2000). In terms of the social consequences, learners with high FLA often participate less in class (Delaney, 2009; Robson 1994), produce less output (Cheng, Horwitz, & Schallert, 1999), and show avoidance behavior and less willingness to communicate (MacIntyre & Charos, 1996).

Although foreign language anxiety has been widely researched in ISLA, it is important to note that much of this research has measured trait anxiety and not state anxiety. For instance, in a recent meta-analysis, Teimouri, Goetze, and Plonsky (2018) analyzed the findings of 97 empirical studies that examined the effects of anxiety in language learning. They found that the effects of L2 anxiety varied depending on the type of language achievement measure, educational level, target language, and anxiety type. However, overall, there was a significant negative correlation between L2 anxiety and language achievement. In spite of these compelling findings, a large majority of the studies included in the sample had examined speaking anxiety, and only 2 out of the 97 total studies addressed learners’ state anxiety.
The negative effects of FLA have been empirically demonstrated in a number of studies that have addressed its relationship with L2 learners’ listening (Kim, 2000; Brunfaut & Revesz, 2015), reading (Sellers, 2000), writing (Cheng, 2002), and speaking skills (Price, 1991). One early perspective of the role of anxiety in SLA was Krashen’s (1982) Affective Filter Hypothesis:

Those whose attitudes are not optimal for second language acquisition will not only tend to seek less input, but they will also have a high or strong Affective Filter—even if they understand the message, the input will not reach the part of the brain responsible for language acquisition, or the language acquisition device. Those with attitudes more conducive to second language acquisition will not only seek and obtain more input, they will also have a lower or weaker filter. They will be more open to the input, and it will strike "deeper". (Krashen, 1982, p. 31)

Thus, Krashen believed that learners with less anxiety, whether classroom-based or personal, would have greater opportunities for second language acquisition. However, the relationship between anxiety and L2 learning has proven to be nonlinear. As MacIntyre (1995) notes, FLA can act as both the cause and the effect, both the chicken and the egg. Variables such as cognition, aptitude, and anxiety exhibit a cyclical pattern, with all three factors capable of influencing each other. The more learners experience failure, the more their anxiety may increase, and in turn, affect their learning and performance.

Early research on the relationship between FLA and language learning examined the effects of anxiety on FL achievements such as standard proficiency tests and course grades and found negative correlations (e.g., Gardner & MacIntyre, 1993; Horwitz, 1986; Phillips, 1992; Trylong, 1987). However, later studies began to address the relationship between anxiety and L2 development. One of the first studies to do so was MacIntyre and Gardner (1994). They compared learners’ performances on a vocabulary learning task completed on the computer, and while some of the participants were performing the task, a video camera watched them. Their
findings demonstrated that the learners’ exposure to the video camera caused higher state anxiety, which resulted in less vocabulary learning. Thus, the researchers were able to identify a direct correlation between L2 anxiety and language development.

However, regarding the empirical evidence that has found a negative correlation between anxiety and L2 achievement, Horwitz (2001) stated that “it is not possible to be sure of the direction of the correlation or to rule out the possibility that some uncontrolled variable is responsible for any relationship which has been observed between the two variables under study.” (p. 117). Therefore, despite extensive research on foreign language anxiety, opportunities to form concrete conclusions have often been limited or ill-founded. Nevertheless, empirical research has demonstrated that certain variables in the learning environment can be manipulated with the aim of reducing L2 learner anxiety. For instance, learners may feel less anxious when interacting with peers in small groups or pairs in comparison with whole-class discussions (Brown, 2001; Davis, 1997; Willis, 1996).

Trait anxiety refers to behavioral patterns that are stable over long periods of time and across situations (i.e., anxiety when using or learning the L2), while state language anxiety is experienced in a specific moment in time in reaction to a stimulus (MacIntyre, 2007). According to Spielberger’s (1983) trait-state theory, when individuals with trait anxiety encounter a stressful situation, the appraisal of the stimulus will provoke state anxiety. Thus, state anxiety is an interaction between trait anxiety and other external factors. Although learners with high trait anxiety can potentially experience greater state anxiety, this interaction with external factors can result in a variety of outcomes. Ultimately, since the measurement of state anxiety comes with methodological challenges (learners’ anxiety can fluctuate throughout the interaction), only a few studies have attempted to measure it (e.g., Baralt & Gurzynski-Weiss, 2011; Valmori; 2016).
Valmori’s (2016) dissertation employed a spot-the-difference task and a picture-story task with 21 L1 English learners of Italian. Participants were recorded during interaction, and they were then asked to watch the videos and comment on the fluctuation of their state anxiety during the interactions. In a pretest-posttest design, their learning outcomes were measured using accuracy and fluency gain scores, and their use of two target structures was analyzed. The researcher found that negotiation and interactional feedback contributed to all participants’ language learning, and that there was no significant correlation between participants’ gain scores and their state anxiety.

**Empirical Studies on Anxiety in Synchronous-Computer Medicated Communication**

Many researchers have suggested that one of the possible advantages of text-based SCMC in comparison with other modes of L2 interaction is its potential to lessen learners’ anxiety (Chun, 1994; Kelm, 1992; Kern, 1995; Warschauer, 1996; Abrams, 2003). Since many researchers have posited that the one of the greatest sources of language anxiety stems from oral production (e.g., Gregersen & Horwitz, 2002; Hauck & Hurd, 2005; Krashen, 2003), this argument seems intuitive. SCMC is ‘an anonymous, less pressured environment that tends to lower the affective filter’ (Beauvois, 1997, p.171), where learners are offered more equitable opportunities for turn-taking (Wang & Woo, 2007), and have more time to plan and process input, negotiate for meaning, and produce output (Beauvois, 1992). However there have been relatively few attempts to empirically address the variable of learner anxiety in SCMC (cf. Arnold, 2007; Baralt & Gurzynski-Weiss, 2011; Cote & Gaffney, 2018; Martin & Valdivia, 2017; Satar & Ozdener, 2008).
The studies that have compared the effects of mode of interaction on FLA have produced mixed results. Baralt and Gurzynski-Weiss (2011) and Arnold (2007) found no significant differences in learner anxiety between the CMC and FTF modes, while Satar and Ozdener (2008) and Cote and Gaffney (2018) found that learners who interacted via text-based SCMC experienced significantly lessened anxiety levels. However, these studies have often operationalized and measured anxiety in different ways and through different types of interaction (e.g., conversation vs. task-based), making it difficult to directly compare their findings. For example, Arnold (2007) addressed communication apprehension, a type of foreign language anxiety, in face-to-face (FTF) and synchronous and asynchronous discussions. The participants were 56 students enrolled in a third-semester university-level German course. They were assigned to one of three groups, either face-to-face (FTF), synchronous CMC, or asynchronous CMC, where they discussed open-ended personal topics in groups of three or four on six occasions. Portions of Horwitz, Horwitz, and Cope’s (1986) Foreign Language Classroom Anxiety Scale (FLCAS) were administered to participants at the beginning and the end of the semester. The results revealed no significant differences in anxiety reduction between the three groups.

While Arnold’s (2007) findings helped form the foundation for further research in CMC on FLA, there are key interactional variables that were not considered. For example, it was reported that the participants’ participation in discussions was not monitored, which may have affected the results because 1) teacher supervision could potentially increase anxiety, and 2) there is no way to confirm participants’ equitable participation in FTF discussion. If participants did not feel pressured to participate due to a lack of supervision, this could explain why some introverted participants did not experience heightened anxiety in the FTF group, where less
equitable participation is more probable than in CMC contexts. For that matter, if it cannot be confirmed that the more anxious learners were participating in FTF, their opportunities for interaction and subsequent L2 learning would potentially be greatly diminished. Nevertheless, at the very least it can be said that SCMC offers an environment that is similar or less anxiety-provoking than the FTF mode (Arnold, 2007; Baralt & Gurzynski-Weiss, 2011; Satar & Ozdener, 2008).

In another SCMC-based study, Baralt and Gurzynski-Weiss (2011) compared the effects of mode (SCMC vs. FTF) on learners’ state FLA. 25 L2 learners of intermediate Spanish completed two information gap tasks with their teacher, once in the FTF mode and once in the SCMC mode. Participants were instructed to complete a foreign language state anxiety questionnaire once halfway through each task and once after the completion of each task. Their findings revealed no significant differences across modality in learners’ reported state anxiety. Thus, although Baralt and Gurzynski-Weiss were unable to demonstrate the superiority of SCMC for reduction of anxiety, the results did indicate that SCMC offered participants with a similar learning environment to FTF interaction.

A recent study that addressed the role of text-based SCMC on learners’ foreign language anxiety is Cote and Gaffney (2018). 61 beginning learners of French were separated into two groups, one that interacted face-to-face and the other via SCMC. Participants learned new vocabulary and the present tense forms of two irregular verbs. They were then instructed to complete a picture description task that targeted the newly acquired grammatical and lexical forms. In a second experiment, participants switched modes and completed the experiment a second time. After completing each experiment, participants were asked to complete a foreign language anxiety questionnaire. The questionnaire results revealed that the participants
experienced significantly less anxiety while using SCMC than while in the FTF classroom setting. Participants also took more turns and their amount of talk was significantly higher in SCMC.

In regards to the effects of the task type on learner anxiety, Eysenck (1979) has argued that this may depend on the complexity of the task. For example, with a relatively simple task, the effects of anxiety may be limited or anxiety can sometimes even facilitate improved performance. However, as the task becomes more complex, anxiety can become a cognitive interference when taking into account the assumed greater cognitive load required to complete the task. Although there have been a few studies that have addressed the relationship between task complexity, L2 learner anxiety, and L2 development in the FTF mode (e.g., Kim & Tracy-Ventura, 2011; Robinson, 2007; Revesz, 2011), the research addressing the relationship between these variables in SCMC is non-existent. There have also been no studies that have addressed the relationship between task type (e.g., dictogloss, two-way information gap, etc.), L2 learner anxiety, and L2 development in the FTF or SCMC mode. This is a troubling gap in the literature, and demonstrates just how little research has been conducted into affective factors in L2 development. As Brown (1973) highlights, “while all the cognitive factors may be operating in the attempted solution of a given task, the learner can fail because of an affective block.” (p. 231-232). Thus, in order to fully understand the processes and products of second language learning, exploration of L2 linguistic development should be accompanied by an examination of affective variables, such as anxiety.

It can be argued that different types of tasks (e.g., dictogloss, jigsaw) have different levels of inherent complexity, given that task type often significantly affects the amounts of negotiation produced as well as learners’ complexity, accuracy, and fluency (Skehan, 2003). In light of the
gap in the literature on the relationship between type of task and learners’ L2 anxiety, the scant literature on the relationship between task complexity, anxiety, and L2 development will be briefly reviewed. For instance, Robinson (2007) utilized three increasingly complex interactive tasks and found that for learners with high output anxiety, the complexity of the task did not affect the complexity of their speech. For learners with low output anxiety, increasingly complex task demands lead to increasingly complex speech. It was also reported that input anxiety was not related to accuracy or complexity. In addition, Kim and Tracy-Ventura (2011) investigated the relationship between task complexity, anxiety, and the development of the simple past. Their findings suggested that low anxiety learners used simple past morphology significantly better, although the effect sizes for anxiety were small. In contrast, Revesz (2011) found no significant relationship between task complexity and L2 learner anxiety.

Corrective feedback is an interactional feature that has been widely studied in terms of the opportunities it presents for modified output and subsequent L2 learning. Given that no studies have empirically addressed state anxiety in relation to learner-teacher versus learner-learner dyadic task-based interaction, I will briefly review some of the findings in ISLA regarding the relationship between foreign language anxiety and corrective feedback (Renko, 2012; Rassaei, 2015; Sheen, 2008). These represent a few of only a handful of studies that have examined anxiety within an interactionist framework.

For example, Renko (2012) investigated the anxiety levels of 100 adolescent Finnish EFL learners in relation to their corrective feedback preferences. She found that learners with higher levels of anxiety preferred less corrective feedback from their teachers. In another study, Sheen (2008) recruited 45 ESL participants enrolled in intermediate-level courses with varying L1 backgrounds to investigate the effectiveness of recasts in relation with learners’ FLA. The
learners were administered a questionnaire on their trait language anxiety and subsequently divided into groups based on the results: high-anxiety learners who received recasts \( (N = 13) \), low-anxiety learners who received recasts \( (N = 11) \), high-anxiety learners who received no recasts \( (N = 11) \), and low-anxiety learners who received no recasts \( (N = 10) \). The findings revealed that recasts were most effective with low-anxiety learners who produced greater amounts of modified output. In terms of the high-anxiety learners, Sheen argued that anxiety “might have blocked their capacity to process the input provided by the recasts and thus limited their ability to benefit from them.” (p. 861).

Finally, Rassaei (2015) compared the effectiveness of different types of corrective feedback (recasts versus metalinguistic feedback) in relation to the trait anxiety levels of 101 EFL learners. Contrasting with Sheen (2008), Rassaei found that while low-anxiety learners benefited from both types of feedback, high-anxiety learners benefited most from recasts. Together with these inconclusive results, an arguable limitation of both of these studies was that they did not address how learners’ anxiety fluctuated in response to corrective feedback, and learners were separated into groups based on their anxiety levels prior to the experimental condition.

Martin and Alvarez Valdivia (2017) is the only published study that has addressed anxiety resulting from teacher versus peer feedback during a task in SCMC by administering two questionnaires on anxiety and corrective feedback to 50 undergraduate students in an online learning environment. Their findings indicated that the students with low anxiety valued teacher feedback significantly more than the high anxiety group. Feedback from peers was least valued by participants in both groups. However, it must be noted that these students were interacting via oral SCMC. In addition, there was also no analysis of participants’ performance on the task.
In summary, it would seem that SCMC offers an environment that is similar or less anxiety-provoking than the FTF mode, but very little research has investigated affective factors in SCMC. Given the scant research on anxiety in SCMC, an investigation examining learners’ state language anxiety levels based on type of task and type of interlocutor in relationship with learners’ output and subsequent L2 development would offer exciting opportunities to bridge a number of different research strands and provide answers to a number of remaining questions in ISLA research.
CHAPTER 4: RESEARCH DESIGN AND METHODOLOGY

This chapter outlines the pilot study’s research design, methodology, results, and modifications for the present study. Following this, the chapter provides the research questions guiding the main study, as well as its design, methodology, materials, procedure, and coding and analysis.

Pilot Study

The pilot study was conducted during the Spring 2018 semester. The pilot study was executed with a much smaller sample size; therefore, its purpose was not to draw any statistical conclusions based on the limited data. Instead, it served the test the validity and reliability of the materials and instructions, determine the desired proficiency level of participants, assess the use of Zoom as a medium for data collection, assess the average time spent on each task, and to examine the potential for obtaining significant results in the main study. In light of the numerous gaps in the literature identified in the previous two chapters, the following research questions motivated the pilot study:

1) To what extent does type of interlocutor (peer vs. professor) affect
   a) the number of LREs produced during interaction?
   b) the quantity of learner output during interaction?
   c) learners’ L2 lexical development? Does this effect last after one week?

2) To what extent does the type of task affect
   a) the number of LREs produced during interaction?
   b) the quantity of learner output during interaction?
   c) learners’ L2 lexical development? Does this effect last after one week?
3) To what extent does learners’ state anxiety affect
   a) the number of LREs produced during interaction?
   b) the quantity of learner output during interaction?
   c) learners’ L2 lexical development? Does this effect last after one week?

4) To what extent is learners’ state anxiety affected by…
   a) type of task?
   b) type of interlocutor?

Pilot Methodology

Participants

This study included two types of participants: learners, who also functioned as peer interlocutors, and a professor.

Learner Participants

Learner participants were 18 adults between the ages of 18 and 21 with native or near-native proficiency in English. One participant was a native speaker of Brazilian Portuguese and one was a heritage speaker of French. All were enrolled in either third or fourth-semester Spanish language courses at the university level. Mean scores at the pretest stage suggested that there were no significant differences based on whether 1) participants were enrolled in the third- or fourth-semester Spanish language courses and whether 2) participants were L2 or L3 learners of Spanish. 2 out of the initial 18 participants had to be excluded from all analyses, leaving a final sample size of 16 learner participants (5 males, 11 females). The first participant was excluded because he was caught looking back at the chat interaction while completing the
immediate posttests, therefore contaminating the data. The second participant was excluded because he admitted to looking up some of the targeted lexical items during the week between the second and third sessions. 10 of these participants interacted with a peer (i.e., another learner participant), making for a total of 5 learner-learner dyads. 6 of the participants interacted with a professor, with the same number of professor-learner dyads. Participants were acquainted with the type of tasks utilized in the pilot study since they had exposure to a task-based, communicative language teaching approach in the L2 classroom.

On the background questionnaire, 12 out of the 16 participants rated their reading and writing skills in Spanish as a 4/7 (“functional”) or 5/7 (“good”). Participants had been learning Spanish for anywhere from 3 months to 8 years. 12 out of 16 participants had never chatted online in Spanish, and the other 4 did so for 1-2 hours per week. 13 out of 16 participants reported that they enjoyed using a computer to communicate with others. Although all participants reported feeling fully comfortable typing in English (7 out of 7 on the Likert scale), their level of comfort with typing in Spanish varied greatly, with 13 out of 16 participants rating their comfort between a 3 and 5 out of 7. When participants were asked what they would like to improve about their Spanish language ability, almost all mentioned wanting to improve their speaking ability.

Although the pilot study focused on state anxiety, the participants were asked about other kinds of anxiety during the background questionnaire. 9 out of the 16 participants considered themselves to be anxious people, and 10 out of 16 reported that learning and/or using their Spanish made them anxious. The order of how anxious they felt using their Spanish in the following ways, from least anxious to most anxious, was: 1) reading, 2) writing, 3) listening, 4) speaking. In terms of who they felt most anxious interacting with, from least anxious to most
anxious, the order was: 1) a friend or classmate who is equally proficient in Spanish, 2) their Spanish professor, 3) a friend or classmate who is more proficient in Spanish, and 4) a native speaker of Spanish. Although it was particularly interesting that the participants felt more anxious interacting with a friend or classmate who is more proficient in Spanish than with their professor, it should be noted that the order of these two interlocutors came down to a difference of one participant. When asked if they could choose whether they could interact in Spanish with their Spanish professor via video chat or text chat, 7 participants said they preferred text chat, 6 participants said they preferred video chat, and 3 participants didn’t have a preference. When asked if they could choose whether to interact in Spanish with a classmate via video chat or text chat, 7 participants preferred text chat, 4 participants preferred video chat, and 5 participants didn’t have a preference.

Professor Participant

The interlocutor in professor-learner dyads was an instructor of Spanish and 4th year Ph.D. student at the same university. Although her official title at the university was not ‘professor’, participants in the professor interlocutor group were told that they would be interacting with a professor, and thus her role will be referred to as professor through the present study. She had three years of teaching experience and had taken courses on second language acquisition and foreign language teaching methodology. She was a native speaker of English with near-native proficiency in Spanish. Given that she was also the researcher, she was very familiar with the aims of the present study, so no training was required.
Materials

Materials used in the pilot study included two experimental tasks (decision-making and information gap), two assessment tasks (open-ended and fill-in-the-blank), a learner state anxiety questionnaire, and a language background questionnaire. Each is explained in further detail below.

Experimental Tasks

Two experimental tasks were employed in the pilot study: a decision-making task and an information gap task. These types of tasks were chosen for their potential to elicit equal participation from both members of a dyad (Pica et al., 1993). The lexical target items incorporated into both tasks were lesser-known household items that were not an integral part of the university’s mandated curriculum for the first four semesters of the Spanish language program. For the information gap task, this meant vocabulary such as los imanes (the magnets) and la licuadora (the blender), and for the decision-making task, targeted lexical items included el sartén (the frying pan) and la linterna (the flashlight). Participant performance on the pretest further confirmed a lack of familiarity with the target items. Both tasks were untimed and completed with a peer or professor interlocutor via iMessage, an online platform for instant messaging.

For the decision-making task (Appendix A), participants were told to imagine that they were on a sinking ship and that they must decide which six out of twelve total items to bring with them onto a nearby deserted island. Each participant in a dyad was given pictures of six items labeled in Spanish, and the six items possessed by each member of the dyad were different. This resulted in a total of 12 targeted lexical items. Due to the design of the task, no distractor items
were included. Each participant was instructed to describe their six items to their partner until they understood the target item being described. The professor interlocutor feigned ignorance of the six targeted lexical items not present in her picture in order to give the participants interacting with her opportunities for interaction that were similar to the opportunities provided to those interacting with a peer. Both members of the dyad had to agree on which of the six items to take on the island. Participants were again instructed to submit their responses via Blackboard before completing the anxiety questionnaire. The decision-making task for the pilot study may be found in the Appendix.

For the information gap task (see Appendix B), there were two versions of a picture of a kitchen containing various household items that were each labeled with their word in Spanish. Each version contained six labeled items that were not present in the other version of the picture. The six items unique to each version of the picture provided a total of 12 targeted lexical items. The picture also contained 12 labeled items that were present in both versions. These 12 items functioned as distractors and were more commonly known household items that had mostly been incorporated into the university’s lower-level Spanish language curriculum. After receiving one of the two versions of the kitchen scene, participants were instructed to compile a list of the twelve total items that they did not have in common by messaging with their partner via iMessage. To avoid participants simply sending a list of items to their partners and, in this fashion, quickly completing the task without fulfilling its experimental purpose, participants were also instructed to describe in detail any items that were missing from their partner’s picture until their partner understood what the item was. Again, the professor interlocutor feigned ignorance of the six items that the learner participant possessed. At the end of the task, in theory the participants should have understood the meaning behind all 12 lexical items on their list of
differences. They were instructed to submit their responses through Blackboard and then to complete the anxiety questionnaire. The information gap task for the pilot study may be found in the Appendix.

Assessment Tasks

Participants completed the same two production assessment tasks for the pretest, immediate posttest, and delayed posttest: one open-ended and one fill-in-the-blank. Both assessment tasks incorporated all of the target items (24) and distractor items (12) from the two experimental tasks. On the open-ended assessment task (Appendix C), the participants were provided with a kitchen scene that was similar to the one used in the information gap task, except the target items from the decision-making task were also present in the picture. Participants were told that their college friend was graduating and needed help packing up her campus apartment. In order to help the friend organize what she needed to pack, participants were instructed to make a list in Spanish of as many of the items that appeared in the picture as they could. If they did not know the word in Spanish for an item, they were also allowed to guess or describe the item in Spanish using other words. They were instructed to answer entirely in Spanish and their responses were submitted via Blackboard.

For the fill-in-the-blank assessment task (Appendix D), participants were given a story in English about Ronald’s day and had to provide the Spanish equivalents for the underlined words in the accompanying blanks. If they did not know the Spanish equivalent of a word, they could choose to guess or leave that item blank. These answers were also submitted via Blackboard.
The state anxiety questionnaire administered to the learner participants in the pilot study (see Appendix E) was adapted from the questionnaire used by Baralt and Gurzynski-Weiss (2011). The aim of the questionnaire was to measure learners’ state anxiety immediately after the completion of each experimental task, and if this anxiety was affected by whom they were interacting with and the modality in which the interaction took place (text-based SCMC). Furthermore, the questionnaire was administered immediately following the completion of each experimental task (i.e., twice) to see if learners’ state anxiety was affected by the task being completed. The version of the questionnaire created by Baralt and Gurzynski-Weiss contained 15 items, while the version adapted for the pilot study contained 22 items. Additional items were added in order to tailor the questionnaire to the needs of the pilot study, and these items included both regular order and reverse order statements, following the design of Baralt and Gurzynski-Weiss. These added items included statements related to type of interlocutor and mode. For example, for the participants in the peer interlocutor groups, added items included ‘I feel more anxious interacting with my peers in class than I did while completing this task with a peer.’ and ‘I would have felt more relaxed and comfortable completing this task face-to-face with my partner.’ Thus, two versions of the questionnaire were created: one for those participants who interacted with a peer and another for those who interacted with a professor. Learners were instructed to rate how much they agreed or disagreed with each of the 22 statements on a five-point Likert scale. The choices offered ranged from SD (strongly disagree) to SA (strongly agree). Responses were submitted via Google Forms.
Language Background Questionnaire

https://forms.gle/SY9gtEy6uVh6DiSQ8

During the first session of the pilot study, participants completed a language background questionnaire via Google Forms. The purpose of this questionnaire was for the researcher to learn more about participants’ past language learning experience, their use of technology for language learning and communication, their anxiety as it related to interactions in Spanish with different types of interlocutors, and their anxiety regarding their Spanish language skills in general. For example, participants were asked to rate their anxiety on a 7-point Likert scale when interacting, whether verbally or through writing, with different types of interlocutors (e.g., a native speaker, a classmate who is equally proficient, etc.). Another question asked whether participants would prefer to interact with their Spanish professor or a classmate via 1) video chat or 2) text chat. Participants were also asked to indicate the course their native language, languages studied and number of years for each, their birth place, where they grew up, how they utilize their Spanish outside of the classroom, and their motivations for learning Spanish. Key findings are reported below in the pilot results section.

Procedure

The first session was completed online via Zoom with the researcher and lasted approximately 20 to 30 minutes. Zoom is a platform similar to Skype that allows users to video chat and share their computer screens. At their pre-determined appointment time, participants received a request from the researcher to join a Zoom session. Participants had been previously instructed to find a quiet place to complete this session. Some participated from the library or from their dormitories, and others sat in empty classrooms. After joining the Zoom session,
participants were instructed to share their screen with the researcher, show the area surrounding their computer to the researcher (e.g., their desk), and place their cell phone out of reach. The participants were sent an electronic IRB-approved consent form, followed by a foreign language background questionnaire and two pretests that were completed via Blackboard: one fill-in-the-blank and one open-ended. The researcher monitored the participants’ screens and video feeds as they completed the pretests. Participants were also given the option to complete the consent form and foreign language background questionnaire on their own time in advance of the first session.

The second session was completed on campus at the language laboratory anywhere from a few hours to 4 or 5 days after the first session. This session lasted between 70 and 90 minutes. Participants were given verbal instructions for completing the tasks and state anxiety questionnaire, and were randomly assigned to interact either with a peer or a professor of Spanish from their university. The experimental tasks completed by the participants at the second session included one decision-making task and one information gap task. Participants completed these two tasks in a randomized order. They were not given a time limit to complete them since this had the potential to affect their self-reported state anxiety (see Baralt & Gurznyski-Weiss, 2011). While completing the tasks, participants were instructed to make their best attempt with spelling and to only worry about accents for the lexical targets and distractors, since accents were not a focus of the study. Spell-check was deliberately turned off in the iMessage settings. Participants were also instructed to use Spanish only while interacting with their partner.

After completing each task, they completed an anxiety questionnaire based on the one from Baralt & Gurzynski-Weiss (2011). Participants were given one of two versions of the adapted questionnaire depending on whom they interacted with (i.e., peer or professor). Upon
completion of the experimental tasks and the accompanying anxiety questionnaires, participants were instructed to minimize their chat windows and then completed the same assessment tasks that were completed during the first session: one open-ended and one fill-in-the-blank immediate posttest. These were also administered in a randomized order. One participant from the professor interlocutor group was caught looking back at the chat while completing the assessment tasks. This participant was promptly excluded from all analyses. Throughout the second session, the researcher monitored the participants to ensure that the experimental tasks were being completed as instructed and she made herself available for any questions.

The third session was also completed online via Zoom with the researcher and took place one week after session 2. This session lasted approximately 20 to 30 minutes. Participants were required to complete the same assessment tasks that were completed during the first and second sessions: one open-ended and one fill-in-the-blank delayed posttest. These were completed in a randomized order via Blackboard. The researcher monitored the participants’ screens and video feeds as they completed the delayed posttests. After completing all testing, participants were asked via Blackboard if they had looked up any of the vocabulary since the time of the second session. If they answered yes, they were promptly excluded from all analyses. Participant 9 from the professor interlocutor group was excluded for this reason. Participants received three extra credit points on one of their Spanish tests after completing the study.

Coding and Analysis

On the open-ended and fill-in-the-blank assessment tasks administered for the pretest, immediate posttest, and delayed posttest, one point was awarded for each correctly produced target item. No half credit was given. Therefore, the maximum score participants could receive
for the target items on each assessment was 24 points. Participants were explicitly instructed to include the gender article for each item, but many participants did not include them or used the incorrect articles. Given that gender was not a focus of the present study, points were not deducted for errors or total omission. In addition, points were not deducted for missing accents or minor spelling errors that did not impede the understanding of the word (e.g., *encendedor* instead of *encendedor; licudora* instead of *licuadora*). Since all tasks were completed on the computer, these could have been typos. Gain scores were also calculated in order to measure participants’ improvement over time more precisely. These included Time 2 gain scores (Time 1 subtracted from Time 2), Time 3 gain scores (Time 1 subtracted from Time 3), and Time 3 losses (Time 2 subtracted from Time 3). As will be seen in the Results section below, the term ‘Time 1’ is used interchangeably with the pretest, ‘Time 2’ with the immediate posttest, and ‘Time 3’ with the delayed posttest.

Participants’ task-based interactions with an interlocutor were coded for total word count, word count per participant, and number of LREs. Each of the three measures was calculated separately for each experimental task. While calculation of the word count per participant would suffice for the peer interlocutor group, I chose to also calculate the mean total word count for the entire interaction per task, since this would provide more insight into the interaction between the professor and learner participants. Although Swain and Lapkin’s (1998) definition of LREs includes both “languaging” and self- or other-correction, I only counted instances of “languaging” as LREs. LREs mostly pertained to the lexical target items, but occasionally the focus of an LRE was an unrelated lexical item. For example, the following exchange was counted as two separate LREs, one for *el cargador* and one for *proteger*:

Participant 2: Tienes el cargador?

*Do you have the charger?*
Professor: no

No

Professor: que es el cargador

What’s the charger

Participant 2: Es una cosa necesario para un teléfono celular

It’s a necessary thing for a cell phone

Professor: para protegerlo?

To protect it?

Participant 2: No se

I don’t know

Participant 2: Que es proteger?

What is protect?

Professor: proteger es como...

To protect is like...

Professor: para que no se dañe el teléfono

So that the phone doesn’t get damaged

Professor: para que no se rompa

So that it doesn’t break

Participant 2: No

No

Participant 2: El cargador no protegerlo el teléfono celular

The charger doesn’t protect the cell phone

Although not a focus of the pilot study, it should be noted that there were no grammatical LREs observed in any of the peer-peer interactions or learner-professor interactions.

For the state anxiety questionnaire employed in the pilot study, following Baralt and Gurzynski-Weiss (2011), each of the 22 items was coded on a scale of 1 through 5. For regular statements such as ‘I felt rushed during the task’, ‘strongly agree’ was assigned an anxiety score of 5 and ‘strongly disagree’ was assigned an anxiety score of 1, and so on. With reverse statements like ‘I felt like I had enough time to complete the task’, ‘strongly agree’ was assigned 1 point and ‘strongly disagree’ was assigned 5 points.

Data from the state anxiety questionnaire was used in a few distinct ways. For the purposes of addressing the third research question, participants’ state anxiety was categorized as
‘high’ or ‘low’ based on their responses to the 15 questionnaire items originating from Baralt and Gurznyski-Weiss’s (2011) study. Since the mean state anxiety score for all participants was 33.44 (SD = 8.74), participants with a score of 33 or below were labeled as low anxiety, and participants with a score of 34 or above were labeled high anxiety. The highest possible score was 75 and the lowest possible score was 15.

For the fourth and final research question, learners’ state anxiety was separated into multiple categories: 1) overall anxiety, which was based on the 15 questionnaire items from Baralt and Gurzynski-Weiss (2011), 2) interlocutor anxiety, which was based on 3 questionnaire items that were added for the present pilot study that focused specifically on type of interlocutor, and 3) mode anxiety, which was based on 4 added questionnaire items that focused on the modality of the interaction (text-based SCMC). Interlocutor and mode anxiety were calculated by averaging participant ratings on the relevant questionnaire items. This meant that the mean scores for interlocutor and mode anxiety ranged between 1 and 5. The highest possible score for the overall anxiety category was 75 and the lowest possible score was 15.

**Pilot Results**

**Research Question 1**

The first research question aimed to address the relationship between type of interlocutor and number of LREs, quantity of learner output, and L2 development. In order to answer the first part of the first research question, the number of LREs that occurred during each experimental task was calculated. These LREs mostly pertained to the lexical target items involved in each task, but occasionally they occurred when participants were attempting to describe a targeted lexical item. Take the following example:

Participant 5: Y el ultimo es las tijeras, no se como los describer
And the last one is the scissors, I don’t know how to describe them

Participant 5: Son metalico
   *They’re metallic*

Professor: para que las usas?
   *What do you use them for?*

Participant 5: no se la palabra, pero para dividir una hoja por ejemplo, hacerlo en dos
   *I don’t know the word, but to divide a sheet for example, to make it into two*

Professor: oooo, ok
   *Oooo, ok*

Professor: si, entiendo
   *Yes, I understand*

Professor: para cortar
   *To cut*

Participant 5: si! Es la palabra!
   *Yes! That’s the word!*

As seen in Table 1 and Figure 1, interactions involving a professor interlocutor seemingly resulted in significantly more LREs than interactions with a peer. The mean number of LREs for the professor interlocutor group was 11.75 \((SD = 2.34)\), while the mean number was 9.11 \((SD = 3.86)\) for the peer interlocutor group. Given these especially large standard deviations, it is difficult to make any sort of generalizations.

Table 1. Pilot study’s number of language-related episodes by type of interlocutor.

<table>
<thead>
<tr>
<th>Interlocutor</th>
<th>LREs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peer</td>
</tr>
<tr>
<td>Mean</td>
<td>9.11</td>
</tr>
<tr>
<td>SD</td>
<td>3.86</td>
</tr>
</tbody>
</table>
To address the effects of type of interlocutor on quantity of learner output, I calculated the total word count and the word count per participant for each experimental task completed via iMessage. While calculation of the word count per participant would suffice for the peer interlocutor group, I chose to also calculate the mean total word count for the entire interaction, since this would provide more insight into the interaction of professor-participant dyads. The mean scores found in Table 2 represent an average of the total word count on both tasks. However, it should be noted that, as will be discussed later in this section, participants in both groups produced more language on the information gap task than on the decision-making task, potentially skewing the means.

**Table 2. Pilot study’s word count by type of interlocutor.**

<table>
<thead>
<tr>
<th></th>
<th>Word count per participant</th>
<th>Total word count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peer</td>
<td>Professor</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>164.3</td>
<td>244.8</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>71.32</td>
<td>56.84</td>
</tr>
</tbody>
</table>

The word count per participant was a good deal higher for participants in the professor interlocutor group, with a mean word count of 244.8 \((SD = 56.84)\) for this group and a mean word count of 164.3 \((SD = 164.3)\) for the peer interlocutor group. Figure 2 offers a visual interpretation of the data. This indicates that participants produced more language while

---

**Figure 1.** Number of language-related episodes by type of interlocutor.
interacting with a professor than while interacting with a peer. This could have been because the professor interlocutor produced a lot more language during interaction than the peer interlocutors. Her mean contribution was 433 words on the information gap task and 324 words on the decision-making task, resulting in an average contribution of 378.5 words overall. Thus, the more the professor spoke, the more the participants interacting with her were prompted to respond and negotiate. This is further evidenced by the total word count by type of interlocutor. The mean total word count for the professor interlocutor group (623.3, $SD = 120.4$) was almost double the mean total word count for the peer interlocutor group (328.6, $SD = 137.6$). A visual representation is offered by Figure 3. These findings are not surprising given that the professor interlocutor group also produced more LREs. These results, supported by the raw data found in Table 4, indicate that the professor was always the dominant member of the dyad in the learner-professor pairs. In addition, review of the raw data in Table 3 suggests that one member of each peer-peer dyad always dominated the interaction. However, these means and observations are far from generalizable, given the extremely small sample sizes (peer interlocutor group: $n = 5$; professor interlocutor group: $n = 6$) and the large standard deviations.

Figure 2. Word count per participant.  
Figure 3. Total word count.
In terms of the relationship between type of interlocutor and L2 development, as noted in Table 5 and Table 6, participants who interacted with a peer and those who interacted with a professor performed similarly on both assessment tasks at Time 1 (i.e., the pretest), producing few to none of the target forms. At Time 2 (i.e., the immediate posttest), the peer interlocutor group \((n = 10)\) appeared to improve significantly from the time of the pretest, with a mean score of 6.00 \((SD = 3.06)\) on the open-ended assessment task and a mean score of 6.50 \((SD = 3.34)\) on the fill-in-the-blank assessment. The professor interlocutor group \((n = 6)\) also appeared to improve significantly from the pretest to the immediate posttest, with a mean score of 7.50 \((SD = 2.67)\) on the open-ended assessment and 9.67 \((SD = 3.98)\) on the fill-in-the-blank assessment. Therefore, at the time of the immediate posttest, the average participant was able to correctly produce at least 6 out of the 24 (25%) total lexical target items versus correctly producing none or one prior to the start of the experiment.
Table 5. Pilot study’s L2 lexical development – Peer interlocutor.

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Target items (Open-ended)</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Target items (Fill-in-the-blank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.40</td>
<td>6.00</td>
<td>3.70</td>
<td>0.90</td>
<td>6.50</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.97</td>
<td>3.06</td>
<td>2.06</td>
<td>1.52</td>
<td>3.34</td>
<td>2.58</td>
<td></td>
</tr>
</tbody>
</table>
Total: 24

Table 6. Pilot study’s L2 lexical development – Professor interlocutor.

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Target items (Open-ended)</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Target items (Fill-in-the-blank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.67</td>
<td>7.50</td>
<td>6.50</td>
<td>0.83</td>
<td>9.67</td>
<td>7.33</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.82</td>
<td>2.67</td>
<td>1.64</td>
<td>0.98</td>
<td>3.98</td>
<td>2.50</td>
<td></td>
</tr>
</tbody>
</table>
Total: 24

Although both groups seemed to improve significantly from Time 1 to Time 2, the mean scores indicate that the professor interlocutor group seems to have outperformed the peer interlocutor group on both assessment tasks, particularly on the fill-in-the-blank assessment, where the mean score for those who interacted with a professor was 3.17 points greater than the mean score for the peer interlocutor group. In other words, participants who collaborated with a professor to complete the experimental tasks correctly produced an average of three more target items on the immediate posttest when compared with the peer interlocutor group. Statistical analyses to determine the significance of overall improvement and differences between groups were not performed due to the small sample sizes. However, it should be noted that the large standard deviations for both groups on both assessment tasks indicate that there was a great deal of variability in participant performance.

In terms of retention of the lexical target items after one week (i.e., at Time 3), the mean scores indicate that participants in both groups were able to correctly produce more target items than at Time 1, but less target items than at Time 2. In other words, participants retained some but not all of their knowledge of the lexical items. On the open-ended assessment, mean scores were 3.70 ($SD = 2.06$) for the peer interlocutor group and 6.50 ($SD = 1.64$) for the professor.
interlocutor group. All participants performed slightly better on the fill-in-the-blank assessment, with a mean score of 4.00 (SD = 2.58) for the peer interlocutor group and 7.33 (SD = 2.50) for the professor interlocutor group. Figures 4 and 6 offer visual representations of the performance of the peer interlocutor group on each assessment task at Time 1, 2, and 3, while the performance of the professor interlocutor group is seen in Figures 5 and 7.

Figure 4. Peer interlocutor group.  
Figure 5. Professor interlocutor group. 

Upon comparison of the two groups, the findings at Time 3 indicate that the professor interlocutor group was also superior in participants’ retention of the target forms. For example, on the open-ended assessment, the mean score of participants in the peer interlocutor group decreased 2.30 points from the immediate to the delayed posttest, while the mean score of
participants in the professor interlocutor group only decreased 1 point. On the fill-in-the-blank assessment task, mean scores of the participants in both groups decreased a similar amount, with a 2.50 point decrease for the peer interlocutor group and a 2.34 point decrease for the professor interlocutor group. However, percentage-wise, the findings indicate that peer interlocutor participants retained approximately 61% of the target forms from Time 2, while professor interlocutor participants retained 76%. This difference in percentage of retention appears to be large, although no further statistical analyses were conducted due to sample sizes. What was particularly interesting when comparing the performance of these two groups over time was that the participants who interacted with a professor performed better on the delayed posttest than the peer interlocutor group performed on the immediate posttest. This offers a strong indication that type of interlocutor indeed might have played a role in participant performance.

Although not a focus of this pilot study, participant performance over time on the distractor items warrants further mention. As can be seen in Tables 7 and 8, participants from both interlocutor groups performed similarly on the distractor items at the time of the pretest. What is particularly noteworthy, however, is the peer interlocutor group’s performance at Time 3 on these items when compared with their performance on the target items at Time 3. At Time 3, participants in the peer interlocutor group correctly produced an average of 3.40 ($SD = 1.90$) distractor lexical items on the open-ended assessment task and an average of 3.80 ($SD = 2.86$) on the fill-in-the-blank task. Although the peer interlocutor group performed slightly better on the distractor items than on the target items at Time 1, it was not a large difference. It seems significant that this group performed similarly on both the target and distractor items at Time 3 given that the distractors were not the focus of the experimental tasks. It would seem that the
learners who interacted with a peer during the experiment walked away with a similar amount of knowledge of the distractor and target items.

Table 7. Pilot study’s L2 performance on distractor items - Peer interlocutor.

<table>
<thead>
<tr>
<th></th>
<th>Open-ended assessment</th>
<th>Fill-in-the-blank assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>Mean</td>
<td>0.60</td>
<td>4.20</td>
</tr>
<tr>
<td>SD</td>
<td>0.84</td>
<td>2.25</td>
</tr>
</tbody>
</table>

Table 8. Pilot study’s L2 performance on distractor items - Professor interlocutor.

<table>
<thead>
<tr>
<th></th>
<th>Open-ended assessment</th>
<th>Fill-in-the-blank assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>Mean</td>
<td>0.83</td>
<td>4.83</td>
</tr>
<tr>
<td>SD</td>
<td>0.75</td>
<td>1.47</td>
</tr>
</tbody>
</table>

Upon comparison of the mean scores at Time 2 and Time 3 for the professor interlocutor group, it was revealed that, on average, these participants retained a similar percentage of distractors and targets from Time 2 to Time 3 on both assessment tasks. For example, as seen in Table 8, on the fill-in-the-blank assessment, participants’ mean scores for the distractor items decreased by one point, or 18%, from Time 2 to Time 3, with a mean score of 4.83 ($SD = 2.56$) at Time 3. The professor interlocutor group’s mean scores for the fill-in-the-blank assessment deceased 24% from Time 2 to Time 3, with a mean score of 7.33 ($SD = 2.50$) at Time 3. Given that these distractor items were not the focus of the information gap task and that the professor, unlike the peer interlocutors, was aware of this, it was anticipated that the participants in this group would not retain knowledge of any of the distractor items over time. However, this was not the case.

In conclusion, the findings of this preliminary study with regards to research question one suggest the following:

1) Type of interlocutor appears to affect the number of LREs produced during interaction. Participants in the professor interlocutor group produced more LREs than the peer interlocutor group.
2) The quantity of learner output appears to be affected by type of interlocutor, with participants producing more language (and receiving more language input) when interacting with a professor than with a peer.

3) Type of interlocutor does appear to affect L2 lexical development and retention. Those who completed the experimental tasks with a professor outperformed the participants who interacted with a peer on both the immediate and the delayed posttests.

Research Question 2

The second research question pertained to type of task. The first part of this question aimed to address the effects of type of task on the number of language-related episodes that took place during participants’ dyadic task-based interaction. As seen in Table 9, the information gap task seemed to result in significantly more LREs than the decision-making task, with a mean of 12.31 (SD = 2.27) for the information gap task and 7.71 (SD = 3.15) for the decision-making task. Although this seems like a significant argument in favor of the information gap task for promoting interaction and thus L2 learning, there may have been other factors at play that altered the results. The findings regarding the effects of type of task on L2 lexical development (which are discussed below in this section) also support this conclusion.

Table 9. Pilot study’s number of language-related episodes by type of task.

<table>
<thead>
<tr>
<th>LREs</th>
<th>Decision-making</th>
<th>Information gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Mean</td>
<td>7.71</td>
<td>12.31</td>
</tr>
<tr>
<td>SD</td>
<td>3.15</td>
<td>2.27</td>
</tr>
</tbody>
</table>

As seen in Table 10, the mean word count per learner participant for the information gap task (228.9, SD = 228.9) is much higher than the word count for the decision-making task (159.4, 208.3).
$SD = 89.29)$. This is not particularly surprising, since a higher number of LREs usually means a greater amount of output. As Figure 8 demonstrates, there was clearly a good deal of variation in the length of the decision-making task interaction (with one participant in the peer interlocutor group only producing 58 words!), while the range for the information gap task was somewhat smaller.

**Table 10.** Pilot study’s word count by type of task.

<table>
<thead>
<tr>
<th>Word count</th>
<th>Decision-making</th>
<th>Information gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td>Mean</td>
<td>159.4</td>
<td>228.9</td>
</tr>
<tr>
<td>SD</td>
<td>89.29</td>
<td>43.86</td>
</tr>
</tbody>
</table>

**Figure 8.** Word count by type of task.

Tables 11 and 12 address L2 development based on type of task, which was the third part of the second research question. With regard to the target items from the information gap task, participants performed identically on the open-ended assessment ($0.06, SD = 0.25$) and fill-in-the-blank assessment tasks ($0.06, SD = 0.25$) at Time 1. For the target items from the decision-making task, Table 10 shows that participants also performed similarly on both assessment tasks at Time 1. At the time of both the immediate posttest (i.e., Time 2) and the delayed posttest (i.e., Time 3), participants again performed similarly on the open-ended and fill-in-the-blank
assessment tasks for the information gap target items. However, for the decision-making target items, there were some differences in performance on the assessment tasks. For the open-ended assessment task, mean scores for the immediate posttest were 3.19 (SD = 1.94) and mean scores for the delayed posttest were 2.63 (SD = 1.67). A comparison of mean scores at Time 2 and 3 suggest that participants retained most of their knowledge of the decision-making target forms. On the other hand, for the fill-in-the-blank assessment task at Time 2, participants seemed to perform a good deal better than on the open-ended assessment, with a mean score of 4.13 (SD = 2.53). Nevertheless, mean scores at Time 3 for the fill-in-the-blank assessment (3.13, SD = 1.93) decreased twice as much as the mean scores for the open-ended assessment, meaning that both assessments ultimately demonstrated similar retention of the decision-making target items.

Table 11. Pilot study’s L2 lexical development – Information gap task.

<table>
<thead>
<tr>
<th></th>
<th>Open-ended assessment</th>
<th>Fill-in-the-blank assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>Mean</td>
<td>0.06</td>
<td>3.38</td>
</tr>
<tr>
<td>SD</td>
<td>0.25</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Table 12. Pilot study’s L2 lexical development – Decision-making task.

<table>
<thead>
<tr>
<th></th>
<th>Open-ended assessment</th>
<th>Fill-in-the-blank assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>Mean</td>
<td>0.44</td>
<td>3.19</td>
</tr>
<tr>
<td>SD</td>
<td>0.81</td>
<td>1.94</td>
</tr>
</tbody>
</table>

When comparing participant performance based on type of experimental task, the participants seemed to improve similarly on the immediate posttest (i.e., Time 2) in their accurate production of the target forms from both experimental tasks. See Figures 9 and 11 for visual representations of participant performance on the information gap task, and Figures 10 and 12 for performance on the decision-making task. From Time 2 to Time 3, mean scores on the information gap target items decreased by 1.25 points on the open-ended assessment and by 1.43 points on the fill-in-the-blank assessment. With regard to the decision-making task, participants
correctly produced an average of 0.50 less target items on the open-ended assessment, and 1 target item less on the fill-in-the-blank assessment. Therefore, it would seem at first glance that the participants retained slightly more of the decision-making target items, but it should be noted that there were larger standard deviations across the board for performance on the decision-making task.

![Figure 9. Information gap task.](image)

![Figure 10. Decision-making task.](image)

![Figure 11. Information gap task.](image)

![Figure 12. Decision-making task.](image)

Furthermore, when participant performance at Time 1 is considered, it appears that participants ultimately retained knowledge of a similar number of lexical items from the information gap and decision-making tasks. These observations are further supported by Table 13 and Table 14, which demonstrate the mean gain scores at Time 2 and Time 3. With a larger sample size, these small differences between task type may have the potential to become
statistically significant, but the differences between tasks in the present pilot study appear insignificant. Overall, the mean scores over time for both the information gap and decision-making tasks suggest similar improvement in terms of the production and retention of the targeted lexical items.

Table 13. Pilot study’s gain scores at Time 2 (T2 – T1) by type of task.

<table>
<thead>
<tr>
<th></th>
<th>Decision-making</th>
<th>Information gap</th>
<th>Decision-making</th>
<th>Information gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.75</td>
<td>3.31</td>
<td>3.31</td>
<td>3.50</td>
</tr>
<tr>
<td>SD</td>
<td>1.92</td>
<td>1.74</td>
<td>2.02</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Table 14. Pilot study’s gain scores at Time 3 (T3 – T1) by type of task.

<table>
<thead>
<tr>
<th></th>
<th>Decision-making</th>
<th>Information gap</th>
<th>Decision-making</th>
<th>Information gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.19</td>
<td>2.06</td>
<td>2.31</td>
<td>2.06</td>
</tr>
<tr>
<td>SD</td>
<td>1.42</td>
<td>1.18</td>
<td>1.58</td>
<td>1.48</td>
</tr>
</tbody>
</table>

In summary, the findings suggest the following with regard to type of task:

1) The occurrence of LREs and participant word count during task-based interaction may be related to type of task; however, it is important to control for any interfering variables before reaching any concrete conclusions. If it were already a perfectly designed experiment, the findings of the present pilot study would indicate that the information gap task was superior for promoting LREs and increased output.

2) Type of task does not seem to affect learners’ L2 development and retention of the targeted lexical forms. Participants ultimately retained knowledge of a similar number of lexical items from the information gap and decision-making tasks.
Research Question 3

The third research question aimed to address the effects of learners’ state foreign language anxiety. Participants’ state anxiety was categorized as ‘high’ or ‘low’ based on their responses to the 15 questionnaire items originating from Baralt and Gurznyski-Weiss’s (2011) study. Since the mean state anxiety score for all participants was 33.44 ($SD = 8.74$), participants with a score of 33 or below were labeled as low anxiety, and participants with a score of 34 or above were labeled high anxiety. As a result, only 4 participants were classified as high anxiety and the other 12 were classified as low anxiety, so the findings for this research question should be viewed with caution. However, without coding the anxiety scores, it would have been difficult to offer any interpretation of this data.

The first part of this research question addressed the effects of learners’ state anxiety on the occurrence of LREs during interaction. As seen in Table 15, the numbers suggest that, on average, participants who experienced high state anxiety while completing the experimental tasks were involved in two more language-related episodes (11.36, $SD = 3.44$) during each task than low state anxiety participants (9.47, $SD = 3.50$). This difference between groups appears to be a bit large; however, the extremely large standard deviations must be considered.

Table 15. Pilot study’s number of language-related episodes in relation to state anxiety.

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.36</td>
<td>9.47</td>
</tr>
<tr>
<td>SD</td>
<td>3.44</td>
<td>3.50</td>
</tr>
</tbody>
</table>

With regard to the second part of the third research question, a calculation of the mean word count based on participants’ state anxiety level rendered no differences between groups (see Table 16). Low state anxiety participants produced an average of 197.9 words per task ($SD = 70.60$) and the high state anxiety participants produced an average of 194.0 words per task ($SD = 70.60$).
In other words, the amount of anxiety participants experienced while completing the experimental tasks did not appear to have affected the amount of output they produced while interacting with their partner.

**Table 16. Pilot study’s word count in relation to state anxiety.**

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>Word count per participant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Mean</td>
<td>194.0</td>
<td>197.9</td>
</tr>
<tr>
<td>SD</td>
<td>88.78</td>
<td>70.60</td>
</tr>
</tbody>
</table>

In order to address the effects of anxiety on L2 lexical development and retention, participants’ state anxiety was compared to their target item gains on the assessment tasks at Time 2 (T2 – T1) and Time 3 (T3 – T1), as well as their losses at Time 3 (T3-T2). Table 17 provides the mean gains at Time 2 for high and low anxiety participants. Figures 13 and 14 provide a visual representation of the differences between groups. The mean gains for the low anxiety group on both the open-ended (6.75, $SD = 2.14$) and fill-in-the-blank assessment tasks (7.67, $SD = 3.00$) suggest that participants with low anxiety improved significantly more than the high anxiety participants in their accurate production of the target items from the pretest to the immediate posttest. However, the large standard deviations for both groups as well as the small sample sizes mean that this interpretation should be regarded with caution.

**Table 17. Pilot study’s gains at Time 2 (T2 – T1) in relation to state anxiety.**

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>Time 2 gains (Open-ended)</th>
<th></th>
<th>Time 2 gains (Fill-in-the-blank)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Mean</td>
<td>4.00</td>
<td>6.75</td>
<td>4.25</td>
<td>7.67</td>
</tr>
<tr>
<td>SD</td>
<td>3.16</td>
<td>2.14</td>
<td>3.40</td>
<td>3.00</td>
</tr>
</tbody>
</table>
Table 18 offers the average target gains at the time of the delayed posttest (Time 3) compared to the mean pretest (Time 1) target scores for both groups. The mean gain scores for both groups seem to indicate that participants who had low anxiety while completing the experimental tasks gained more knowledge of the target items over time, although the differences between groups are not as striking as at Time 2. However, when the two anxiety groups’ losses from Time 2 to Time 3 are compared (see Table 19), high anxiety learners’ mean target item scores decreased less from the immediate to the delayed posttest, with a mean decrease of 1.25 points on both the open-ended ($SD = 2.06$) and fill-in-the-blank ($SD = 1.26$) assessment tasks. This is likely because the low state anxiety group had higher mean gain scores at Time 2, and therefore had more to lose. Overall, the means in Table 18 still indicate that, on average, the low state anxiety participants ultimately gained knowledge of a greater number of target items than the high state anxiety participants.

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>Time 3 gains (Open-ended)</th>
<th>Time 3 gains (Fill-in-the-blank)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Mean</td>
<td>2.75</td>
<td>4.75</td>
</tr>
<tr>
<td>SD</td>
<td>1.71</td>
<td>1.96</td>
</tr>
</tbody>
</table>
Table 19. Pilot study’s losses at Time 3 (T3 – T2).

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>Time 3 losses (Open-ended)</th>
<th>Time 3 losses (Fill-in-the-blank)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-1.25</td>
<td>-2.00</td>
</tr>
<tr>
<td></td>
<td>-1.25</td>
<td>-2.83</td>
</tr>
<tr>
<td>SD</td>
<td>2.06</td>
<td>2.13</td>
</tr>
<tr>
<td></td>
<td>1.26</td>
<td>1.90</td>
</tr>
</tbody>
</table>

In summary, for research question three, the preliminary findings suggest the following regarding the effects of participants’ state anxiety:

1) The task-based interactions of participants with high state anxiety were characterized by a higher number of LREs.

2) State anxiety levels did not seem to be related to the amount of language participants produced on task.

3) Low levels of state anxiety seemed to result in greater initial gains in L2 performance on the targeted lexical items, but the effects of state anxiety were less clear in terms of the retention of L2 lexical items.

Research Question 4

The final research question was intended to address the effects of type of interlocutor and type of task on learners’ state anxiety. For this purpose, learners’ state anxiety was separated into
multiple categories: 1) overall anxiety, which was based on the 15 questionnaire items from Baralt and Gurzynski-Weiss (2011), 2) interlocutor anxiety, which was based on 3 questionnaire items that were added for the present pilot study that focused specifically on type of interlocutor, and 3) mode anxiety, which was based on 4 added questionnaire items that focused on the modality of the interaction (text-based SCMC). Interlocutor and mode anxiety were calculated by averaging participant ratings on the relevant questionnaire items. All ratings were between 1 and 5 on the Likert scale, with 5 representing the highest level of anxiety and 1 representing the lowest anxiety level. The highest possible score for the overall anxiety category was 75.

As can be seen in Table 20, it would appear that participants who interacted with a professor reported experiencing less overall state anxiety during the experimental tasks (31.83, $SD = 4.95$). However, based on the standard deviation for the peer interlocutor group ($SD = 10.38$), it would seem that participants who interacted with a peer varied quite a bit in the state anxiety they experienced. Figure 17 offers a visual interpretation of this variation. In terms of interlocutor anxiety and mode anxiety, Table 20 shows that participants from both interlocutor groups responded very similarly on the associated questionnaire items. Therefore, it seems that neither of the two interlocutor groups felt more anxious than the other about the interlocutor they interacted with or about interacting via text chat.

<table>
<thead>
<tr>
<th></th>
<th>Overall anxiety</th>
<th>Interlocutor anxiety</th>
<th>Mode anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peer</td>
<td>Professor</td>
<td>Peer</td>
</tr>
<tr>
<td>Mean</td>
<td>34.40</td>
<td>31.83</td>
<td>1.88</td>
</tr>
<tr>
<td>SD</td>
<td>10.38</td>
<td>4.95</td>
<td>0.68</td>
</tr>
</tbody>
</table>
With regard to the relationship between state anxiety and type of task, Table 21 offers a look into the descriptive statistics. Participants’ mean overall anxiety score was 35.31 ($SD = 8.97$) after completing the information gap task and 31.56 ($SD = 8.53$) after completing the decision-making task. Although further statistical analyses could not be conducted, these mean anxiety scores suggest that the participants experienced more anxiety while completing the information gap task than while completing the decision-making task. However, the visual representation of the differences between state anxiety on tasks (Figure 18) suggests that these differences may not be significant. In terms of interlocutor anxiety and mode anxiety, the mean anxiety scores from Table 21 indicate that participants from both interlocutor groups responded very similarly on the associated questionnaire items. Therefore, it appears that type of task did not appear to have affected how anxious participants felt about whom they were interacting with or the modality in which the interaction took place (text-based SCMC).

**Table 21.** Pilot study’s state anxiety in relation to type of task.

<table>
<thead>
<tr>
<th></th>
<th>Overall anxiety</th>
<th>Interlocutor anxiety</th>
<th>Mode anxiety</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DM</td>
<td>IG</td>
<td>DM</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>31.56</td>
<td>35.31</td>
<td>2.06</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>8.35</td>
<td>8.97</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Figure 18. Overall state anxiety in relation to type of task.

In summary, there was somewhat of a difference between interlocutor groups in terms of overall state anxiety. The participants who interacted with a professor reported slightly lower state anxiety levels, but there were also extremely large standard deviations. Participants also experienced less anxiety after completing the decision-making task than after the information gap task, but the very large standard deviations also suggest a lack of reliability for these findings.

Pilot Modifications

The pilot study served to test the validity and reliability of the materials and instructions, determine any prior knowledge of the targeted lexical items, determine the desired proficiency level of participants, assess the use of Zoom as a medium for data collection, assess the average time spent on each task, and to examine the potential for obtaining significant results in the main study. While the pilot study was being carried out, numerous methodological issues arose that provided insight into how the study’s design could be improved. As a result, a number of modifications were made to the main study’s methodology and materials.
The pilot study included learner participants that were enrolled in either third-semester (Intermediate I) or fourth-semester (Intermediate II) university-level Spanish courses. For the main study, I decided to recruit participants only from Intermediate II. Although the mean scores at the time of the pretest indicated that level did not affect participants’ prior knowledge of the targeted lexical items, I observed while reviewing the participants’ chat transcripts that participants enrolled in fourth-semester Spanish courses were more effective communicators who had less difficulty expressing themselves throughout the interaction. This also improved the experience of their partners, who were able to respond to the input that they were exposed to with less confusion. This allowed them to focus more on the targeted lexical items while completing the tasks.

One of the biggest changes that I made to the main study was to the interlocutor participant. In the pilot study, participants either interacted with a peer or with a professor interlocutor. However, on the language background questionnaire, a majority of the participants noted that they would have felt the most anxious if they had been instructed to complete the experimental tasks with a native speaker. Given that the present study investigates the relationship between anxiety and type of interlocutor, the logical next step was to include a native speaker interlocutor to allow for further comparison of the effects of interlocutor type. Furthermore, countless empirical studies in the interaction strand of ISLA research have investigated the characteristics and effects of NS-NNS interaction, but have often not taken into account that the interlocutors used are often both educators and native speakers. One of the goals of the present study was to determine if this distinction truly matters when conducting interaction-based research.
Regarding the experimental tasks, and more specifically the target and distractor items used for them, one of the potential reasons that the participants performed so well at Time 2 and Time 3 during the pilot study was the activation of prior knowledge. Although participants were not able to produce almost any target items at Time 1, a number of them indicated while completing the tasks with me that they understood a couple of them. As discussed in the pilot study results section above, the decision-making task resulted in less LREs than the information gap task. Although other potential explanations may exist, this was likely in part because the participants recognized a few of the target items, so they did not need to discuss the meaning of them with their partner (which would have resulted in an LRE). This may also explain why participants performed similarly on target items from both assessment tasks, despite a lower number of LREs and lower word count for the decision-making task.

This led to a change that needed to be made to the study’s assessment measures. Both assessment measures utilized in the pilot measured participants’ written production of the targeted forms; however, there was no assessment of their recognition of the targeted forms. Based on the findings of the pilot, a recognition assessment task was added, and the written production assessment task was changed entirely. I replaced both the open-ended task and the fill-in-the-blank task at the suggestion of my dissertation committee. Both of these assessment tasks served very similar purposes: To encourage participants’ written production of the targeted lexical forms. However, since English was not used during the completion of the experimental tasks, these assessment tasks were not an appropriate measure of the experiment’s effects on participants’ written production.

Given participants’ familiarity with a few of the targeted lexical items on the pilot study, I also replaced a number of the targeted items on the experimental tasks for the main study with
lower-frequency lexical items. Before incorporating these new targeted items into the experimental tasks, I confirmed that these were not vocabulary words found in the University’s textbooks for the first, second, third, or fourth-semester Spanish courses. The targeted forms replaced on the decision-making task were as follows: *la linterna* (flashlight), *las pastillas* (pills), *el cepillo de dientes* (toothbrush), *las tijeras* (scissors), *el paraguas* (umbrella), *el mòvil* (cell phone), and *el reloj* (clock). The targeted forms replaced on the information gap task were as follows: *la jarra* (pitcher), *el cargador* (charger), *el guante de cocina* (oven mitt), *la licuadora* (blender), and *la plancha* (iron). My doctoral committee also advised me that 6 distractor items on the information gap task would suffice, so I removed the following distractor items: *la microondas* (microwave), *la tetera* (kettle), *el cajón* (drawer), *el tazón* (bowl), *el lavaplatos* (dishwasher), *las fotos* (photos), and *la taza* (mug). In fact, a few of the distractor items (i.e., *la tetera* and *el tazón*) were used as targeted lexical items for the main study based on the pilot study’s participants’ performance with these forms. A list of all targeted lexical items and distractors is provided in Appendix J.

With regard to the language background questionnaire, I decided to change a few of the questions based on the pilot study’s questionnaire responses. For the question that asks participants whether they would prefer to interact with “your Spanish professor” via text chat or video chat, I changed it to “a Spanish professor you don’t know”. Given that the participants were already familiar with their Spanish professors, I found that they provided a different answer when the question pertained to an unidentified professor, as outlined in the questionnaire results of the main study. Changing the language in the questionnaire aligned better with the design of the present study, since participants interacted with a Spanish professor that they did not know, and not with the Spanish professor they saw every day in class. I also added questions regarding
1) their use of instantaneous chatting applications (e.g., Whatsapp) and 2) their task preferences, following the example of Gurzynski-Weiss (2011).

In terms of the learner state anxiety questionnaire, I had to ensure that every regular statement has a reverse statement for the main study. For the pilot study, I added 7 questions to Baralt and Gurzynski-Weiss’ (2011) questionnaire: 3 regarding type of interlocutor and 4 about mode. Since these statements should have come in pairs, I needed to include one more question on type of interlocutor. However, since an additional interlocutor group was added to the study’s design, a total of 3 more questions regarding type of interlocutor was added.

For the main study’s experimental design, it was decided that the recognition and written production assessment tasks for Time 1 (i.e., the pretests) would be administered to participants 2 weeks prior to the experiment (Time 2). Since participants’ completion of the recognition assessment task at Time 1 offered exposure to the targeted items and could have affected their performance on the experimental and assessment tasks at Time 2, there needed to be a long period of time between the pretest and the experiment in order to limit the effects of this exposure. During the pilot study, there were only a few hours or days between the pretest and experimental stage, but because there was no recognition assessment task, the limited amount of time between sessions did not impact the findings of the pilot.

With regard to procedure during the experiment, a number of small changes were required to improve the design of the study. For instance, one of the pilot study participants had to be excluded because he was caught looking back at the chat while completing the immediate posttest. Instead of simply minimizing the chat while participants complete the assessment tasks, I decided to close out the chat window on the participants’ computer screens after the two tasks had been completed. For the pilot study, the chat was being minimized instead of closed because
I was having trouble with my iMessage and I needed access to the chat transcript. However, the chat transcripts for the main study were copied immediately after the participants completed the task so that the chat window could be promptly closed. Furthermore, because of the issues that I encountered while navigating iMessage, I chose to use Zoom as the messaging platform for the experimental stage of the main study.

In addition to giving participants verbal instructions, I also improved and elaborated on the written instructions for the main study’s assessment tasks. When I was giving verbal instructions for the pilot study, I found that I occasionally forgot to relay a piece of information to a participant (e.g., ‘Pay attention to spelling’). I also found that a few of the pilot study participants were not initially completing the tasks as directed until I intervened. I had trouble wording the directions in a way that would avoid participants falling into a pattern of ‘Do you have this? Yes. Do you have this? No.’ Thus, for the main study, I offered written examples in the assessment task instructions to help guide their interactions.

Main Study

The present study aims to contribute to the gap in research on task-based interaction by employing two types of tasks (decision-making and two-way information gap) in the SCMC mode along with three types of interlocutors (peer, professor or native speaker) to investigate the relationship of these variables with the occurrence of language-related episodes, learners’ interactional output, learners’ L2 development, and learners’ self-reported state anxiety. The research questions motivating this study are as follows:

1) To what extent does Type of interlocutor (peer vs. professor vs. native speaker) affect

   a) the number of LREs produced during interaction?
b) the amount of talk on task?

c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?

d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

2) To what extent does Type of task (decision-making vs. information gap) affect

   a) the number of LREs produced during interaction?

   b) the amount of talk on task?

   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?

   d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

3) What is the relationship between learners’ state anxiety and

   a) the number of LREs produced during interaction?

   b) the amount of talk on task?

   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?

   d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

   e) Type of interlocutor?

   f) Type of task?
Methodology

Participants

This study included two types of participants: Learners, who also functioned as peer interlocutors, and the researcher, who served as both the professor and native speaker interlocutor.

Learner Participants

Learner participants were 82 adults between the ages of 18 and 24 with native \((n = 77)\) or near-native \((n = 5)\) proficiency in English. The L1s of the participants with near-native English proficiency were Polish, Punjabi, Urdu, Vietnamese, and Portuguese. 35 participants were male and 47 participants were female. All were enrolled in fourth-semester Spanish language courses at the university level. Mean scores on the recognition and production assessment measures at Time 1 suggested that there were no significant differences based on whether participants were L2 or L3 learners of Spanish.

Out of an initial sample size of 86 participants, 4 participants had to be excluded from all analyses. Participant #16 was excluded because of her performance on the recognition assessment at Time 1. She scored more than 3 standard deviations (>50.50%) above the overall mean score \((15.71, SD = 11.58)\). Participant #34 was excluded for not following instructions on the written production assessment at Time 2, while participants #51 and #70 were excluded for not completing one or both of the assessments at Time 2. Finally, 2 participants, #55 and #56, had to be excluded from any analyses related to participants’ state anxiety on the decision-making task, since they did not complete the anxiety questionnaire for this task.
30 participants interacted with a peer (i.e., another learner participant), making for a total of 15 learner-learner dyads. 27 participants interacted with a professor interlocutor and 25 participants interacted with a “native speaker” interlocutor. All participants were acquainted with the type of tasks utilized in this study since they have exposure to a task-based, communicative language teaching approach in the L2 classroom.

*Professor and Native Speaker Interlocutor*

The same interlocutor participated in both the professor-learner and native speaker-learner dyads. She was an instructor of Spanish and 5th year Ph.D. candidate at the same university. She was a native speaker of English with near-native proficiency in Spanish. She had four years of teaching experience and had taken courses on second language acquisition and foreign language teaching methodology. Given that she was also the researcher, she was very familiar with the aims of the present study, so no training was required.

Although the interlocutor’s official title at the university was not professor and she was not a native speaker, participants in the professor and native speaker interlocutor groups were told that they would be interacting with either a professor they did not know or a native speaker. To explain further, in many empirical studies that have examined teacher-learner interaction (e.g., Van den Branden, 1997), the teacher participant was also a native speaker, but their status as native speaker and its effects on participants’ performance and the features of the interactions were not considered when interpreting the findings. Therefore, the present study has manipulated this independent variable (i.e., type of interlocutor) in order to 1) provide further insight into its effects on interactional features, participant performance, and state anxiety, 2) control for
potential differences in performance between the two interlocutor types and 3) offer potential support for the findings of previous empirical research.

Materials

Materials used in the present study included two experimental tasks (decision-making and information gap), two assessment tasks (recognition and written production), a learner state anxiety questionnaire, and a language background questionnaire. Each is explained in further detail below.

Experimental Tasks

Two experimental tasks were employed in this study: a decision-making task and a two-way information gap task (i.e., a jigsaw task). These types of tasks were chosen for their potential to elicit equal participation from both members of a dyad (Pica et al., 1993). In addition, although SCMC-based interaction research as a whole has produced mixed findings, some empirical studies such as Blake (2000) and Yilmaz and Granena (2010) have found that the jigsaw task was superior for promoting negotiation, while Smith (2003) found that the decision-making task resulted in more negotiation. Furthermore, Pelletieri’s (2000) findings led her to suggest that goal-oriented tasks that included items, concepts, or vocabulary that the learners were unfamiliar with had the potential to increase negotiation between learners. Thus, the lexical target items incorporated into both tasks were lesser-known household items that were not an integral part of the university’s mandated curriculum for the first four semesters of the Spanish language program. For the information gap task, this meant 12 vocabulary words such as los imanes (the magnets) and la licuadora (the blender), and for the decision-making task, the 12
targeted lexical items included *el sartén* (the frying pan) and *la linterna* (the flashlight). Participant performance on the pretests, particularly the production assessment, further confirmed a lack of familiarity with the target items. A list of the 24 total targeted lexical items, as well as the 6 distractor items found on the information gap task, are found in Appendix J. Both tasks were untimed and completed with a peer, professor, or native speaker interlocutor via Zoom, an online platform that allows for instant messaging.

For the decision-making task (Appendix F), participants were told to imagine that they were on a sinking ship and that they must decide which six out of twelve total items to bring with them onto a nearby deserted island. Via Google Forms, each participant in a dyad was randomly assigned to either Version A or Version B of the task, giving them access to pictures of six items labeled in Spanish, and the six items possessed by each member of the dyad were different. This resulted in a total of 12 targeted lexical items. Due to the design of the task, no distractor items were included. Each participant was instructed to describe their six items to their partner through Zoom until they understood the target item being described. While interacting via Zoom, all participants had their cameras turned off and were instructed to communicate via text chat. The professor and native speaker interlocutor feigned ignorance of the six targeted lexical items not present in her picture in order to give the learners interacting with her opportunities for interaction that were similar to the opportunities provided to those interacting with a peer. In addition, empirical research has shown that learning by teaching has the potential to improve participant learning gains (de Wit et al., 2018; Tanaka & Matsuzoe, 2012). Both members of the dyad had to agree on which of the six items to take on the island. Participants were instructed to submit their responses via Google Forms and then to complete the anxiety questionnaire.
For the information gap task (see Appendix G), there were two versions of a picture of a kitchen containing various household items that were each labeled with their word in Spanish. Each version contained six labeled items that were not present in the other version of the picture. This resulted in a total of 12 targeted lexical items between the two versions of the task. The picture also contained 6 labeled items that were present in both versions. These 6 items functioned as distractors and were more commonly known household items that had been incorporated into the university’s lower-level Spanish language curriculum. Participants were randomly assigned to one of the two versions of the kitchen scene on Google Forms and instructed to compile a list of the twelve total items that they did not have in common by messaging with their partner via Zoom. To avoid participants simply sending a list of items to their partners and, in this fashion, quickly completing the task without fulfilling its experimental purpose, participants were also instructed to describe in detail any items that were missing from their partner’s picture until their partner understood what the item was. Again, the professor and native speaker interlocutor feigned ignorance of the six items that the learner participant possessed. At the end of the task, the goal was for participants to understand the meaning behind all 12 lexical items on their list of differences. They were again instructed to submit their responses through Google Forms and to then complete the anxiety questionnaire.

Assessment Tasks

Participants completed one recognition assessment task and one production assessment task at Time 1 (pretest), Time 2 (immediate posttest), and Time 3 (delayed posttest). Both assessment tasks incorporated all of the target items (24) from the two experimental tasks and the distractor items (6) from the information gap task. The recognition assessment task (Appendix
H) was a 30-question multiple-choice assessment completed via Google Forms. For each question, the participants were provided with one image of one of the 30 total vocabulary items, with only one question appearing on each page of the task. Participants were instructed to choose the vocabulary word in Spanish that best matched the item found in the image. They were provided with 30 possible answers to choose from: the 24 target items and 6 distractor items from the experimental tasks. Question order and answer choice order were both randomized automatically for all participants by using the appropriate settings in Google Forms. The researcher verbally instructed participants to not guess, but rather to leave blank any questions that they were not sure about. Although multiple-choice assessments typically do not include 30 answer choices, this was also done to avoid participant guessing.

The production assessment task (Appendix I) was a fill-in-the-blank task completed via Google Forms. Participants were again provided with one image of one of the 30 total vocabulary items, with only one question appearing on each page of the task. A blank appeared underneath each image, and participants were instructed to fill the blank with the word in Spanish that best described the item in the image. Question order was randomized automatically for all participants by using the appropriate settings in Google Forms. If participants did not know the Spanish equivalent, they were verbally instructed to leave the question blank rather than guess.

*Learner State Anxiety Questionnaire*

The state anxiety questionnaire administered to the learner participants in this study (see Appendix K) was adapted from the questionnaire used by Baralt and Gurzynski-Weiss (2011). The aim of the questionnaire was to measure learners’ state anxiety immediately after the
completion of each experimental task, and if this anxiety was affected by whom they were interacting with and the modality in which the interaction took place (text-based SCMC). Furthermore, the questionnaire was administered immediately following the completion of each experimental task (i.e., twice) to see if learners’ state anxiety was affected by the task being completed. The version of the questionnaire created by Baralt and Gurzynski-Weiss contained 15 items, while the version adapted for the present study contained 25 items. Additional items were added in order to tailor the questionnaire to the needs of the present study, and these items included both regular order and reverse order statements, following the design of Baralt and Gurzynski-Weiss. These added items included statements related to type of interlocutor and mode. For example, for the participants in the peer interlocutor groups, added items included ‘I feel more anxious interacting with my peers in class than I did while completing this task with a peer.’ and ‘I would have felt more relaxed and comfortable completing this task face-to-face with my partner.’ Thus, three versions of the questionnaire were created: one for those participants who interacted with a peer, one for those who interacted with a professor, and a third version for those who interacted with a native speaker. Learners were instructed to rate how much they agreed or disagreed with each of the 25 statements on a five-point Likert scale. The choices offered ranged from SD (strongly disagree) to SA (strongly agree). Responses were submitted via Google Forms.

Language Background Questionnaire

https://drive.google.com/open?id=1J3yq5tLzPKeer_9fGYW8nvq9QTePrp7vb38TBhiw_90

During the first session of this study, participants completed a language background questionnaire via Google Forms. The purpose of this questionnaire was for the researcher to
learn more about participants’ past language learning experience, their use of technology for language learning and communication, their anxiety as it related to interactions in Spanish with different types of interlocutors, and their anxiety regarding their Spanish language skills in general. For example, participants were asked to rate their anxiety on a 7-point Likert scale when interacting, whether verbally or through writing, with different types of interlocutors (e.g., a native speaker, a classmate who was equally proficient, etc.). Another question asked whether participants would prefer to interact with their Spanish professor or a classmate via 1) video chat or 2) text chat. Participants were also asked to indicate the course they were enrolled in, their native language, languages studied and number of years for each, their birth place, where they grew up, how they utilize their Spanish outside of the classroom, and their motivations for learning Spanish.

On the background questionnaire, 80% of participants rated their reading skills in Spanish at a 4 out of 7 (“functional”) or 5 out of 7 (“good”), and 72% of participants rated their writing skills as “functional” or “good”. Participants had been learning Spanish for anywhere from 1 to 14 years. 95% of participants had never chatted online in Spanish, and the few that had did so for 1-2 hours per week. 85.2% of participants reported that they enjoyed using a computer to communicate with others. Although 92.6% of participants reported feeling fully comfortable typing in English (7 out of 7 on the Likert scale), their level of comfort with typing in Spanish varied greatly, with 83% of participants rating their comfort between a 3 and 5 out of 7. When participants were asked what they would like to improve about their Spanish language ability, almost all mentioned wanting to improve their speaking and/or listening skills.

Although the present study focuses on state anxiety, the participants were asked about other kinds of anxiety during the background questionnaire. 33.3% of participants considered
themselves to be anxious people all the time and 43.2% of participants considered themselves to be anxious sometimes. In addition, 37% of participants reported that learning and/or using their Spanish made them anxious all the time, and 44.4% only sometimes. The order of how anxious they felt using their Spanish in the following ways, from least anxious to most anxious, was: 1) reading, 2) writing, 3) listening, 4) speaking. In terms of who they felt most anxious interacting with, from least anxious to most anxious, the order was: 1) a friend or classmate who is equally proficient in Spanish, 2) a friend or classmate who is more proficient in Spanish, 3) their Spanish professor, 4) a native speaker of Spanish, and 5) a Spanish professor they don’t know. Although participants reported feeling most anxious interacting with a Spanish professor they don’t know, the differences in reported anxiety when interacting with a native speaker versus an unknown Spanish professor were minimal.

When asked if they could choose whether they could interact in Spanish with their Spanish professor via video chat or text chat, 60.5% of participants said they preferred text chat, 18.5% of participants said they preferred video chat, and 21% of participants did not have a preference. Also, when asked if they could choose whether they could interact in Spanish with a Spanish professor they don’t know via video chat or text chat, 71.6% of participants said they preferred text chat, 14.8% of participants said they preferred video chat, and 13.6% of participants did not have a preference. When asked if they could choose whether they could interact in Spanish with a native speaker via video chat or text chat, 63% of participants said they preferred text chat, 24.7% of participants said they preferred video chat, and 12.3% of participants did not have a preference. Finally, when asked if they could choose whether to interact in Spanish with a classmate via video chat or text chat, 49.4% of participants preferred
text chat, 27.2% of participants preferred video chat, and 23.5% of participants did not have a preference.

**Procedure**

The first session was completed online via Zoom with the researcher and lasted approximately 20 to 30 minutes. Zoom is a platform similar to Skype that allows users to video chat, text chat, and share their computer screens. At their pre-determined appointment time, participants received a request from the researcher to join a Zoom session. Participants had been previously instructed to find a quiet place to complete this session. Some participated from the library or from their dormitories, and others sat in empty classrooms. After joining the Zoom session, participants were instructed to share their screens with the researcher, show the area surrounding their computers to the researcher (e.g., their desk), and place their cell phones out of reach. The participants were sent an electronic IRB-approved consent form along with links to a foreign language background questionnaire and two pretests that were completed via Google Forms in the following order: 1) one production assessment and 2) one recognition assessment. The researcher monitored the participants’ screens and video feeds as they completed the pretests. Participants were also given the option to complete the consent form and foreign language background questionnaire on their own time in advance of the first session.

The second session was completed on campus at the language laboratory approximately two weeks after the first session. This session lasted between 70 and 90 minutes. Participants were given verbal instructions for completing the tasks and state anxiety questionnaire, and were randomly assigned to interact either with a peer, native speaker of Spanish, or a professor of Spanish from their university. The experimental tasks completed by the participants at the
second session included one decision-making task and one information gap task. Participants completed these two tasks in a randomized order. They were not given a time limit to complete them since this would have had the potential to affect their self-reported state anxiety (see Baralt & Gurznyski-Weiss, 2011), and the focus of the present study was the relationship between anxiety, task type, and interlocutor. While completing the tasks, participants were instructed to make their best attempt with spelling and to only worry about accents for the lexical targets and distractors, since accents were not a focus of the study. Spell-check was deliberately turned off in the Zoom settings, and participants’ cameras were turned off as well. Participants were also instructed to use Spanish only while interacting with their partner.

After completing each task, they completed an anxiety questionnaire based on the one from Baralt and Gurznyski-Weiss (2011). Participants were given one of three versions of the adapted questionnaire depending on whom they interacted with (i.e., peer, professor, or native speaker). Upon completion of the experimental tasks and the accompanying anxiety questionnaires, participants were instructed to minimize their chat windows and then complete the same assessment tasks that were completed during the first session and in the same order: First, a production assessment task, and second, a recognition assessment task. The researcher took care to ensure that participants did not refer back to the chat when completing the assessments. Throughout the second session, the researcher also monitored the participants to confirm that the experimental tasks were being completed as instructed and she made herself available for any questions.

The third session was also completed online via Zoom with the researcher and took place approximately one week after session 2. This session lasted anywhere from 10 to 20 minutes. Participants were required to complete the same assessment tasks that were completed during the
first and second sessions: One recognition and one production delayed posttest. The production assessment task was competed first, followed by the recognition assessment task. The links to these assessments in Google Forms were only sent to participants once the Zoom session began. The researcher monitored the participants’ screens and video feeds as they completed the delayed posttests. After completing all testing, participants were asked via Google Forms if they had looked up or studied any of the vocabulary at any point since the first session of the experiment. All 82 participants reported that they had not. Participants received three extra credit points on one of their Spanish tests upon completion of the study.

**Coding and Analysis**

The coding procedure for the recognition and written production assessment tasks for the main study was identical to the coding procedure employed in the pilot study. On the recognition and production assessment tasks administered for the pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3), one point was awarded for each correctly produced target item. Given that gender agreement was not a variable investigated in the present study, points were not deducted for errors or total omission. In addition, points were not deducted for missing accents or minor spelling errors that did not impede the understanding of the word (e.g., *encendedor* instead of *encendedor*; *licuadora* instead of *licuadora*). Since all tasks were completed on the computer, these could have been typos. Therefore, the maximum score participants could receive for the target items from each experimental task (decision-making and information gap) on each assessment task (recognition and written production) was 12 points. Thus, there were four scores, each with a maximum of 12 points, for each participant at Time 1, Time 2, and Time 3. For the
statistical analyses, these four scores for each participant were converted to percentages in order to offer a clearer picture of participant performance over time.

In order to conduct the correlation and regression analyses necessary to address RQ3, gain scores were calculated in order to measure participants’ improvement over time. These included Time 2 gain scores (Time 1 subtracted from Time 2), Time 3 gain scores (Time 1 subtracted from Time 3), and Time 3 losses (Time 2 subtracted from Time 3). As will be seen in the Results section below, the term ‘Time 1’ is used interchangeably with the pretest, ‘Time 2’ with the immediate posttest, and ‘Time 3’ with the delayed posttest.

Participants’ task-based interactions with an interlocutor were coded for total word count, word count per participant, and number of LREs. Each of the three measures was calculated separately for each experimental task. While calculation of the word count per participant would suffice for the peer interlocutor group, I chose to also calculate the mean total word count for the entire interaction per task, since this would provide more insight into the proportion of contributions from each interlocutor in the professor-learner and native speaker-learner dyads.

With regard to LREs, although Swain and Lapkin’s (1998) definition of LREs includes both “languaging” and self- or other-correction, I only counted instances of “languaging” as LREs. LREs mostly pertained to the lexical target items, but occasionally the focus of an LRE was an unrelated lexical item. For example, the following exchange was counted as two separate LREs, one for *el cargador* and one for *proteger*:

Participants’ task-based interactions with an interlocutor were coded for total word count, word count per participant, and number of LREs. Each of the three measures was calculated separately for each experimental task. While calculation of the word count per participant would suffice for the peer interlocutor group, I chose to also calculate the mean total word count for the entire interaction per task, since this would provide more insight into the proportion of contributions from each interlocutor in the professor-learner and native speaker-learner dyads.

With regard to LREs, although Swain and Lapkin’s (1998) definition of LREs includes both “languaging” and self- or other-correction, I only counted instances of “languaging” as LREs. LREs mostly pertained to the lexical target items, but occasionally the focus of an LRE was an unrelated lexical item. For example, the following exchange was counted as two separate LREs, one for *el cargador* and one for *proteger*:

```
Participant 2: Tienes el cargador?
             Do you have the charger?
Professor:   no
             No
Professor:   que es el cargador
             What’s the charger
Participant 2: Es una cosa necesario para un teléfono celular
```
It’s a necessary thing for a cell phone

Professor: para protegerlo?
To protect it?

Participant 2: No se
I don’t know

Participant 2: Que es proteger?
What is protect?

Professor: proteger es como...
To protect is like...

Professor: para que no se dañe el teléfono
So that the phone doesn’t get damaged

Professor: para que no se rompa
So that it doesn’t break

Participant 2: No
No

Participant 2: El cargador no protegerlo el teléfono celular
The charger doesn’t protect the cell phone

Although not a focus of the present study, it should be noted that there were no grammatical LREs observed in any of the peer-peer interactions, professor-learner, or native speaker-learner interactions.

For the state anxiety questionnaire employed in this study, following Baralt and Gurzynski-Weiss (2011), each of the 25 items was coded on a scale of 1 through 5. For regular statements such as ‘I felt rushed during the task’, ‘strongly agree’ was assigned an anxiety score of 5 and ‘strongly disagree’ was assigned an anxiety score of 1, and so on. With reverse statements like ‘I felt like I had enough time to complete the task’, ‘strongly agree’ was assigned 1 point and ‘strongly disagree’ was assigned 5 points.

For the third research question, learners’ state anxiety was separated into multiple categories: 1) overall anxiety, which was based on the 15 questionnaire items from Baralt and Gurzynski-Weiss (2011), 2) interlocutor anxiety, which was based on 6 questionnaire items that were added for the present pilot study that focused specifically on type of interlocutor, and 3)
mode anxiety, which was based on 4 added questionnaire items that focused on the modality of the interaction (text-based SCMC). Interlocutor and mode anxiety were calculated by averaging participant ratings on the relevant questionnaire items. This meant that the mean scores for interlocutor and mode anxiety ranged between 1 and 5. The highest possible score for the overall anxiety category was 75 and the lowest possible score was 15.
CHAPTER 5: RESULTS

Research Questions 1 and 2

1) To what extent does Type of interlocutor (peer vs. professor vs. native speaker) affect
   a) the number of LREs produced during interaction?
   b) the amount of talk on task?
   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?
   d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

2) To what extent does Type of task (decision-making vs. information gap) affect
   a) the number of LREs produced during interaction?
   b) the amount of talk on task?
   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?
   d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

RQ1a and RQ2a: Language-Related Episodes

In order to address research questions 1a and 2a, descriptive statistics were first calculated for the number of LREs produced during each experimental task by participants in each interlocutor group. Means and standard deviations for the three interlocutor groups on each experimental task were computed and can be found in Table 22. In order to confirm that the variances were not significantly different among interlocutor groups, Levene’s Test of Equality
of Error Variances was used. The sphericity test was not a valid measure given the number of groups. Levene’s Test revealed that homogeneity of variance was not violated, with $F(2,63) = 2.34, p = .11$ for the decision-making task and $F(2,63) = 1.61, p = .21$ for the information gap task. Therefore, a two-way ANOVA could be carried out. Significance level was set at $p \leq .05$.

Table 22. Number of language-related episodes, by interlocutor group and type of task.

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Decision-making task</th>
<th>Information gap task</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor ($n = 27$)</td>
<td>12.33 (1.07)</td>
<td>12.67 (1.21)</td>
<td>25.00 (1.47)</td>
</tr>
<tr>
<td>Native speaker ($n = 25$)</td>
<td>12.08 (1.38)</td>
<td>13.04 (1.51)</td>
<td>25.12 (2.39)</td>
</tr>
<tr>
<td>Peer ($n = 14$)</td>
<td>10.29 (2.13)</td>
<td>12.29 (0.83)</td>
<td>21.87 (3.58)</td>
</tr>
<tr>
<td>Total ($n = 66$)</td>
<td>11.80 (1.65)</td>
<td>12.73 (1.28)</td>
<td>24.34 (2.73)</td>
</tr>
</tbody>
</table>

A two-way $3 \times 2$ ANOVA was then conducted with the LRE data in which Type of interlocutor (professor vs. native speaker vs. peer) was entered as the between-subjects factor and Type of task (decision-making vs. information gap) was entered as the within-subjects factor. The dependent variable was mean number of LREs. The ANOVA results indicated that there were significant main effects for 1) Type of interlocutor ($F(2,63) = 8.03, p = .00, \eta^2_p = .20$), and 2) Type of task ($F(1,63) = 22.46, p = .00, \eta^2_p = .26$). A significant interaction between Type of task and Type of interlocutor group was also observed ($F(2,63) = 3.94, p = .02, \eta^2_p = .11$). These results indicated that it was a combination of both independent variables that determined the number of LREs that were produced by participants, that is, the significant main effects observed for both Type of interlocutor and Type of task need to be interpreted in light of this significant interaction.

In terms of Type of interlocutor, Tukey’s honestly significant difference (HSD) post hoc test showed that, on average, both the native speaker and professor interlocutor groups produced significantly more LREs than the peer interlocutor group ($p < .05$). When comparing only the native speaker and professor interlocutor groups, the native speaker interlocutor group produced
slightly more LREs; however, the difference was not significant ($p = .98$). All post hoc test results are found in Table 23.

**Table 23.** Tukey’s post hoc test for number of language-related episodes.

<table>
<thead>
<tr>
<th>(I) Interlocutor group</th>
<th>(J) Interlocutor group</th>
<th>Mean difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor</td>
<td>Native speaker</td>
<td>-.06</td>
<td>.286</td>
<td>.976</td>
<td>-.75 -.63</td>
</tr>
<tr>
<td>Professor</td>
<td>Peer</td>
<td>1.21*</td>
<td>.340</td>
<td>.002</td>
<td>.40 2.03</td>
</tr>
<tr>
<td>Native speaker</td>
<td>Professor</td>
<td>.06</td>
<td>.286</td>
<td>.976</td>
<td>-.63 .75</td>
</tr>
<tr>
<td>Native speaker</td>
<td>Peer</td>
<td>1.27*</td>
<td>.344</td>
<td>.001</td>
<td>.45 2.10</td>
</tr>
<tr>
<td>Peer</td>
<td>Professor</td>
<td>-1.21*</td>
<td>.340</td>
<td>.002</td>
<td>-2.03 -.40</td>
</tr>
<tr>
<td>Peer</td>
<td>Native speaker</td>
<td>-1.27*</td>
<td>.344</td>
<td>.001</td>
<td>-2.10 -.45</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level.

With regard to Type of task, participants produced more LREs on the information gap task regardless of which Type of interlocutor they interacted with. This finding is supported by the descriptive data presented in the “Total” row in Table 22. However, the significant interaction observed between the independent variables indicated that the highest number of LREs were produced by the native speaker interlocutor group on the information gap task, and the lowest number of LREs were produced by the peer group on the decision-making task. Ultimately, all three groups produced a similar number of LREs during collaboration on the information gap task, but the peer interlocutor group produced significantly less LREs than the other two groups on the decision-making task. However, the large standard deviation for the peer interlocutor group on the decision-making task should also be taken into consideration. These differences can be observed in Figure 19.
Figure 19. Estimated marginal means of number of language-related episodes, by interlocutor group.

In sum, a combination of Type of interlocutor (professor vs. native speaker vs. peer) and Type of task (decision-making vs. information gap) differentially affected the number of LREs produced during interaction. While participants from all interlocutor groups produced a similar number of LREs on the information gap task, participants from the peer interlocutor group produced significantly less LREs on the decision-making task. Ultimately, it would seem that participants’ prior knowledge of the L2 lexical items found on each experimental task may have played a role here, but this observation will be discussed in further detail in the following chapter.

RQ1b and RQ2b: Talk on Task

In order to address research questions 1b and 2b, descriptive statistics were calculated for the amount of talk on task, operationalized as participants’ word count, during each experimental
task by participants in each interlocutor group. Means and standard deviations for the three interlocutor groups on each experimental task were computed and can be found in Table 24. In order to confirm that the variances were not significantly different among interlocutor groups, Levene’s Test of Equality of Error Variances was used. The sphericity test was not a valid measure given the number of groups. Levene’s Test revealed that the variances were not significantly different among interlocutor groups, with $F(2,77) = .99, p = .38$ for the decision-making task and $F(2,77) = .57, p = .57$ for the information gap task. Therefore, a two-way ANOVA was subsequently conducted. Significance level was set at $p \leq .05$.

**Table 24.** Amount of talk on task, by interlocutor group and type of task.

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Decision-making task</th>
<th>Information gap task</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor ($n = 27$)</td>
<td>257.48 (64.34)</td>
<td>239.67 (67.30)</td>
<td>497.15 (121.06)</td>
</tr>
<tr>
<td>Native speaker ($n = 25$)</td>
<td>285.64 (77.85)</td>
<td>239.88 (50.42)</td>
<td>525.52 (120.41)</td>
</tr>
<tr>
<td>Peer ($n = 28$)</td>
<td>191.96 (64.47)</td>
<td>227.18 (58.00)</td>
<td>409.57 (110.98)</td>
</tr>
<tr>
<td>Total ($n = 80$)</td>
<td>243.35 (78.70)</td>
<td>235.36 (58.73)</td>
<td>473.76 (126.27)</td>
</tr>
</tbody>
</table>

* Defined as word count

A two-way 3 x 2 ANOVA was run with the data on amount of talk on task in which Type of interlocutor (professor vs. native speaker vs. peer) was entered as the between-subjects factor and Type of task (decision-making vs. information gap) was entered as the within-subjects factor. The dependent variable was mean amount of talk on task. The ANOVA results showed that there was a significant main effect for Type of interlocutor ($F(2,77) = 6.00, p = .00, \eta^2_p = .14$). No significant main effect was found for Type of task ($F(1,77) = 2.47, p = .12, \eta^2_p = .03$). However, similar to the LREs results, there was a significant interaction between Type of task and Type of interlocutor ($F(2,77) = 15.67, p = .00, \eta^2_p = .29$). These results indicated that it was a combination of both independent variables that determined the amount of talk on task by participants, that is, the significant main effect observed for Type of interlocutor needs to be interpreted in light of this significant interaction.
This significant interaction stems from the native speaker and professor interlocutor groups producing many more words on the decision-making task than the peer group, as can be seen in Figure 20. Tukey’s Test confirmed that learner participants who interacted with a professor or a native speaker produced significantly more talk on task than the learner participants who interacted with a peer ($p < .05$). In other words, the participants in both the professor and native speaker groups had significantly higher word counts. When comparing the findings of these analyses to the descriptive data found in Table 24, it is again clear that the difference between groups was especially prominent for the decision-making task, while no significant differences were found between the three groups on the information gap task. Post hoc test results for the decision-making task are found in Table 25.
Table 25. Tukey’s post hoc test for amount of talk on task.

<table>
<thead>
<tr>
<th>(J) Interlocutor group</th>
<th>(J) Interlocutor group</th>
<th>Mean difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lower bound</td>
<td>Upper bound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professor</td>
<td>Native speaker</td>
<td>-14.19</td>
<td>16.185</td>
<td>.657</td>
<td>-52.87</td>
</tr>
<tr>
<td></td>
<td>Peer</td>
<td>39.00</td>
<td>15.729</td>
<td>.040</td>
<td>1.41</td>
</tr>
<tr>
<td>Native speaker</td>
<td>Professor</td>
<td>14.19</td>
<td>16.185</td>
<td>.657</td>
<td>-24.50</td>
</tr>
<tr>
<td></td>
<td>Peer</td>
<td>53.19</td>
<td>16.046</td>
<td>.004</td>
<td>14.84</td>
</tr>
<tr>
<td>Peer</td>
<td>Professor</td>
<td>-39.00</td>
<td>15.729</td>
<td>.040</td>
<td>-76.59</td>
</tr>
<tr>
<td></td>
<td>Native speaker</td>
<td>-53.19</td>
<td>16.046</td>
<td>.004</td>
<td>-91.54</td>
</tr>
</tbody>
</table>

*. The mean difference is significant at the .05 level.

In summary, a combination of Type of interlocutor (professor vs. native speaker vs. peer) and Type of task (decision-making vs. information gap) differentially affected the amount of talk on the decision-making task. The native speaker and professor interlocutor groups produced many more words on the decision-making task than the peer group. In comparison, participants in all interlocutor groups produced similar amounts of language on the information gap task. Again, participants’ prior knowledge of the L2 lexical likely played a role here, a supposition that will be discussed in more detail in the following chapter.

RQ1c and RQ2c: L2 Lexical Recognition

To address research questions 1c and 2c, descriptive statistics were computed for participants’ performance on the recognition assessment tasks at Time 1 (pre-test), Time 2 (immediate posttest) and Time 3 (delayed posttest). These findings were separated by Type of interlocutor and by which experimental task the targeted lexical items originated from, either decision-making or information gap. Means and standard deviations for recognition data can be found in Table 26. For participant performance on the lexical recognition tasks, Levene’s Tests revealed that the variances among interlocutor groups were not significantly different on target.
items originating from either experimental task at Time 1, 2, or 3 ($p > .05$). As a result, a repeated-measures ANOVA was subsequently run. Significance level was set at $p \leq .05$.

**Table 26.** Descriptive data for recognition assessment task at Time 1, 2, and 3 (in percentages).

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Recognition assessment task, M (SD)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decision-making task*</td>
<td>Information gap task *</td>
<td></td>
</tr>
<tr>
<td>Professor ($n = 27$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>17.59 (10.16)</td>
<td>8.33 (10.84)</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>76.23 (18.01)</td>
<td>59.88 (24.46)</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>74.07 (19.93)</td>
<td>59.88 (28.22)</td>
<td></td>
</tr>
<tr>
<td>Native speaker ($n = 25$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>22.67 (12.39)</td>
<td>8.67 (10.89)</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>78.33 (19.25)</td>
<td>65.00 (20.97)</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>71.00 (20.00)</td>
<td>54.00 (22.32)</td>
<td></td>
</tr>
<tr>
<td>Peer ($n = 30$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>21.11 (14.64)</td>
<td>12.77 (12.71)</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>72.22 (18.09)</td>
<td>60.83 (22.33)</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>65.56 (19.91)</td>
<td>53.33 (22.60)</td>
<td></td>
</tr>
<tr>
<td>Total ($n = 82$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>20.43 (12.64)</td>
<td>10.06 (11.62)</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>75.41 (18.37)</td>
<td>61.79 (22.49)</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>70.02 (20.02)</td>
<td>55.69 (24.39)</td>
<td></td>
</tr>
</tbody>
</table>

*Percentages were calculated from a total of 12 targeted items originating from each experimental task.

A repeated-measures 3 x 2 x 3 ANOVA was conducted in which Type of interlocutor (professor vs. native speaker vs. peer) was entered as the between-subjects factor, and Type of experimental task (decision-making vs. information gap) and Time (pretest vs. immediate posttest vs. delayed posttest) were entered as the within-subjects factors. The ANOVA results indicated that there was no significant main effect for Type of interlocutor ($F(2,79) = .18, p = .83, \eta^2_p = .01$) and significant main effects for Type of task ($F(1,79) = 103.65, p = .00$), with a partial eta-squared effect size of $\eta^2_p = .57$, which means that 57% of variance in recognition scores could be attributed to Type of task, and Time ($F(2,79) = 619.96, p = .00$), with a partial eta-squared effect size of $\eta^2_p = .89$, which means that 89% of variance in recognition scores
could be attributed to Time. A significant interaction was also observed between Type of interlocutor and Time ($F(4,158) = 2.58, p = .04, \eta^2_p = .06$) as seen in Table 27.

Table 27. ANOVA summary table for within-subjects effects for recognition assessment data.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>Mean square</th>
<th>F</th>
<th>p</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>1</td>
<td>2.04</td>
<td>103.65</td>
<td>.00*</td>
<td>0.57</td>
</tr>
<tr>
<td>Task*Group</td>
<td>2</td>
<td>0.02</td>
<td>0.93</td>
<td>.40</td>
<td>0.02</td>
</tr>
<tr>
<td>Error (Task)</td>
<td>79</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>14.09</td>
<td>619.96</td>
<td>.00*</td>
<td>0.89</td>
</tr>
<tr>
<td>Time*Group</td>
<td>4</td>
<td>0.06</td>
<td>2.58</td>
<td>.04*</td>
<td>0.06</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>158</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task*Time</td>
<td>2</td>
<td>0.02</td>
<td>2.06</td>
<td>.13</td>
<td>0.03</td>
</tr>
<tr>
<td>Task<em>Time</em>Group</td>
<td>4</td>
<td>0.01</td>
<td>0.62</td>
<td>.65</td>
<td>0.02</td>
</tr>
<tr>
<td>Error (Task*Time)</td>
<td>158</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

In terms of the effects of Type of interlocutor, while the tests of between-subjects effects indicated that the Type of interlocutor that learner participants interacted with did not significantly affect recognition of the targeted L2 lexical items, a significant interaction was found between Type of interlocutor and Time. Given this significant interaction and in order to address whether the effect of Type of interlocutor on participants’ performance on L2 lexical recognition tasks lasts after one week, paired t-tests were conducted to compare the mean scores obtained on the pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3) by each experimental group. The results revealed that participants in all interlocutor groups statistically increased their receptive performance from the pretest to the immediate posttest (Professor: $p = .00$, Native speaker: $p = .00$, Peer: $p = .00$), but this performance statistically decreased from the immediate posttest to the delayed posttest for participants in the native speaker ($p = .01$) and peer interlocutor groups ($p = .00$). For participants in the professor interlocutor group, there was no significant decrease in performance from the immediate posttest to the delayed posttest ($p = .63$), which may explain the significant interaction between Type of interlocutor and Time. A review of the descriptive data found in Table 26 and the graphs in
Figure 21 shows that learner participants in the professor group retained, on average, almost all of their receptive knowledge of the targeted lexical items from both experimental tasks one week after the experiment was completed.

Decision-making Task

Information Gap Task

**Figure 21.** Participant performance on L2 lexical recognition assessment tasks over time.

Ultimately, the paired t-tests revealed that in all three groups, a statistically significant difference was maintained between the delayed posttest mean scores and the pretest mean scores.
(Professor: $p = .00$, Native speaker: $p = .00$, Peer: $p = .00$). This indicates that while participants in the native speaker and peer groups may have experienced a significant decrease in their receptive performance from Time 2 to Time 3, overall, interaction with any of the three interlocutor types still led to a lasting effect one week after the experiment. The paired t-test results for the recognition assessment data are found in Table 28.

Table 28. Paired t-test results for recognition assessment data, separated by type of interlocutor.

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Mean difference</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor ($n = 27$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>55.09</td>
<td>18.21</td>
<td>15.72</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-1.08</td>
<td>11.65</td>
<td>-0.48</td>
<td>.63</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>54.01</td>
<td>19.76</td>
<td>14.20</td>
<td>.00</td>
</tr>
<tr>
<td>Native speaker ($n = 25$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>56.00</td>
<td>16.14</td>
<td>17.35</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-9.17</td>
<td>11.47</td>
<td>-4.00</td>
<td>.01</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>46.83</td>
<td>14.60</td>
<td>16.04</td>
<td>.00</td>
</tr>
<tr>
<td>Peer ($n = 30$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>49.58</td>
<td>16.32</td>
<td>16.65</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-7.08</td>
<td>9.67</td>
<td>-4.01</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>42.50</td>
<td>15.10</td>
<td>15.42</td>
<td>.00</td>
</tr>
</tbody>
</table>

With regard to the effect of Type of task, the significant main effect indicates that participants were significantly more accurate in their recognition of the lexical target items from the decision-making task. However, the lack of significant interaction between Time and task combined with the significant main effect for Time suggest that, overall, participants improved significantly over time in their accurate recognition of the targeted lexical items, but this occurred regardless of whether the target items originated from the decision-making or information gap experimental task.

These conclusions are supported by the results of paired t-tests, which were used to address whether the effect of Type of task on participants’ performance on L2 lexical recognition tasks lasts after one week. The mean scores obtained on the pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3) for targeted lexical items originating from the decision-
making task and information gap task were compared. The results revealed that participants statistically increased their accurate recognition of targeted lexical items originating from both experimental tasks from the pretest to the immediate posttest (Decision-making: \( p = .00 \), Information gap: \( p = .00 \)), but this performance also statistically decreased from the immediate posttest to the delayed posttest for both tasks (Decision-making: \( p = .00 \), Information gap: \( p = .00 \)). Ultimately, for both experimental tasks, a statistically significant difference was maintained between the delayed posttest mean scores and the pretest mean scores (Decision-making: \( p = .00 \), Information gap: \( p = .00 \)). This indicates that both experimental tasks still led to a lasting effect one week after the experiment. The paired t-test results for the recognition assessment data are found in Table 29.

**Table 29.** Paired t-test results for recognition assessment data, separated by type of task.

<table>
<thead>
<tr>
<th>Experimental task</th>
<th>Mean difference</th>
<th>SD</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision-making ( (n = 82) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>54.98</td>
<td>18.18</td>
<td>27.38</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-5.39</td>
<td>12.79</td>
<td>-3.81</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>49.59</td>
<td>18.00</td>
<td>24.95</td>
<td>.00</td>
</tr>
<tr>
<td>Information gap ( (n = 82) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>51.73</td>
<td>21.35</td>
<td>21.94</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-6.10</td>
<td>13.92</td>
<td>-3.97</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>45.63</td>
<td>21.53</td>
<td>19.20</td>
<td>.00</td>
</tr>
</tbody>
</table>

In sum, participants in all interlocutor groups statistically increased their receptive performance on lexical items originating from both experimental tasks over time. Although the interlocutor with whom learner participants interacted did not significantly affect their recognition of the targeted lexical items on either task, participants who interacted with a professor exhibited significantly higher levels of retention of their newly acquired receptive knowledge of the targeted lexical items, regardless of experimental task. In addition, participants were significantly more accurate in their recognition of the targeted lexical items from the decision-making task, regardless of interlocutor group. As previously mentioned, this last finding
may be related to participants’ activation of prior knowledge, but this is explained further in the discussion chapter.

**RQ1d and RQ2d: L2 Lexical Production**

To address research questions 1d and 2d, descriptive statistics were computed for learner participants’ performance on the written production assessment tasks at Time 1 (pre-test), Time 2 (immediate posttest) and Time 3 (delayed posttest). These findings were separated by Type of interlocutor and by which experimental task the targeted lexical items originated from, either decision-making or information gap. Means and standard deviations for written production data can be found in Table 30 For participant performance on the lexical production tasks, Levene’s Test determined that the variances were significantly different among interlocutor groups at Time 1 for target items originating from the decision-making task ($F(2,79) = 15.65, p = .00$), and at Time 3 for performance on information gap task target items ($F(2,79) = 3.44, p = .04$). The other four Levene’s Tests conducted to assess homogeneity of variance for performance on the lexical production tasks did not reveal significant $p$ values ($p > .05$).

**Table 30.** Descriptive data for written production assessment task at Time 1, 2, and 3 (in percentages).

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Written production assessment task, M (SD)</th>
<th>Decision-making task *</th>
<th>Information gap task*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor ($n = 27$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>0.00 (0.00)</td>
<td>0.93 (2.67)</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>48.15 (26.08)</td>
<td>27.78 (19.88)</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>28.70 (18.39)</td>
<td>21.30 (17.35)</td>
<td></td>
</tr>
<tr>
<td>Native speaker ($n = 25$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time 1</td>
<td>1.67 (4.17)</td>
<td>1.00 (5.00)</td>
<td></td>
</tr>
<tr>
<td>Time 2</td>
<td>45.00 (26.24)</td>
<td>35.00 (21.11)</td>
<td></td>
</tr>
<tr>
<td>Time 3</td>
<td>31.67 (21.65)</td>
<td>24.00 (24.69)</td>
<td></td>
</tr>
</tbody>
</table>
Table 30 (cont.)

<table>
<thead>
<tr>
<th>Peer (n = 30)</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.67 (3.39)</td>
<td>0.83 (2.54)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40.28 (20.54)</td>
<td>33.33 (19.82)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25.28 (14.60)</td>
<td>19.72 (15.08)</td>
<td></td>
</tr>
<tr>
<td>Total (n = 82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.11 (3.14)</td>
<td>0.91 (3.47)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.31 (24.18)</td>
<td>32.01 (20.22)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>28.35 (18.18)</td>
<td>21.54 (19.02)</td>
<td></td>
</tr>
</tbody>
</table>

The two violations of homogeneity of variance among groups warranted further interpretation of the data before proceeding with the repeated-measures ANOVA. For the violation at Time 1 for decision-making target items, a review of the raw data determined that the distribution was very slightly leptokurtic, with a few scores pulling up the mean and creating a left-skewed distribution. 72 out of the 82 total study participants scored 0% on the written production pre-test for decision-making target items. Another 9 participants scored 8.3%, and the final participant scored 16%. For the violation at Time 3 for information gap target items, a review of the raw data once again determined that the distribution was slightly leptokurtic, with a few scores pulling up the mean and creating a left-skewed distribution. Computing the descriptive statistics for all of the participants, regardless of interlocutor group, further confirmed this, with a mean score of 21.03%, a median of 16.67%, a mode of 8.33%, a standard deviation of 18.86%, and a range of 83.3%. Overall, it was more common for participants to perform relatively poorly on the Time 3 written production assessment task than it was for them to perform extremely well – only 2 participants scored an 83.3% on this assessment, while 27 participants scored an 8.3%. Ultimately, due to the fact that ANOVA is considered to be fairly ‘robust’ to deviations from normality (Larson-Hall, 2010, p. 355), inferential analyses were carried out regardless of these violations, and a Greenhouse-Geisser correction was applied to the ANOVA results. Significance level was set at $p \leq .05$. 

137
A repeated-measures 3 x 2 x 3 ANOVA was conducted in which Type of interlocutor (professor vs. native speaker vs. peer) was entered as the between-subjects factor, and Type of experimental task (decision-making vs. information gap) and Time (pretest vs. immediate posttest vs. delayed posttest) were entered as the within-subjects factors. The ANOVA results revealed no significant main effect for Type of interlocutor ($F(2,79) = .38, p = .69, \eta^2_p = .01$) and significant main effects for Type of task ($F(1,79) = 48.10, p = .00$) with a partial eta-squared effect size of $\eta^2_p = .38$, which means that 38% of variance in written production scores could be attributed to Type of task, and Time ($F(1.73,79) = 199.16, p = .00$) with a partial eta-squared effect size of $\eta^2_p = .72$, which means that 72% of variance in written production scores could be attributed to Time. Significant interactions were also observed between 1) Type of task and Time ($F(1.94,79) = 24.51, p = .00, \eta^2_p = 0.24$), and 2) Type of interlocutor, Type of task, and Time ($F(3.88, 153.12) = 4.00, p = .01, \eta^2_p = 0.01$), as exhibited in the ANOVA summary in Table 31.

**Table 31.** ANOVA summary table for within-subject effects for written production assessment data.

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>Mean square</th>
<th>F</th>
<th>p</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>1</td>
<td>0.52</td>
<td>48.10</td>
<td>.00*</td>
<td>0.38</td>
</tr>
<tr>
<td>Task*Group</td>
<td>2</td>
<td>0.02</td>
<td>2.04</td>
<td>.14</td>
<td>0.05</td>
</tr>
<tr>
<td>Error (Task)</td>
<td>79</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>1.73</td>
<td>6.73</td>
<td>199.16</td>
<td>.00*</td>
<td>0.72</td>
</tr>
<tr>
<td>Time*Group</td>
<td>3.46</td>
<td>0.01</td>
<td>0.34</td>
<td>.83</td>
<td>0.01</td>
</tr>
<tr>
<td>Error (Time)</td>
<td>136.57</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task*Time</td>
<td>1.94</td>
<td>0.16</td>
<td>24.51</td>
<td>.00*</td>
<td>0.24</td>
</tr>
<tr>
<td>Task<em>Time</em>Group</td>
<td>3.88</td>
<td>0.03</td>
<td>4.00</td>
<td>.01*</td>
<td>0.01</td>
</tr>
<tr>
<td>Error (Task*Time)</td>
<td>153.12</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Since no significant main effect was found for Type of interlocutor, it would seem that the type of interlocutor with whom learner participants interacted did not impact their overall performance on productive measures. However, there was a significant triple interaction between Type of interlocutor, Type of task, and Time. This indicates that participants’ accurate written
production of the 24 targeted lexical items was statistically affected by a combination of the interlocutor they interacted with, the type of experimental task the target items originated from, and time. A review of the descriptive data found in Table 30 shows that the professor and native speaker groups outperformed the peer group on accurate written production of target items from the decision-making task; however, this difference was more marked at Time 2 (the immediate posttest). In terms of the accurate written production of information gap target items, the native speaker and peer interlocutor groups outperformed the professor group, but again, this difference was clearest at the time of the immediate posttest. This is represented in Figure 22, where it can be observed that these differences appear to be minor. Furthermore, the partial eta squared effect size observed ($\eta^2_p = 0.01$) means that only 1% of the variance in written production assessment scores could be attributed to the interaction between Type of interlocutor, Type of task, and Time. This indicates that the true significance of this interaction should be interpreted with caution.

Decision-making Task

![Decision-making Task Graph]

**Estimated marginal means**

- **Professor**
- **Native speaker**
- **Peer**

**Group**

**Time**

1, 2, 3
To address whether the effect of Type of interlocutor on participants’ performance on L2 lexical production tasks lasts after one week, paired t-tests were conducted to compare the mean scores obtained on the pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3) by each experimental group. The results revealed that participants in all interlocutor groups statistically increased their productive performance from the pretest to the immediate posttest (Professor: $p = .00$, Native speaker: $p = .00$, Peer: $p = .00$), but this performance statistically decreased from the immediate posttest to the delayed posttest for all interlocutor groups (Professor: $p = .00$, Native speaker: $p = .00$, Peer: $p = .00$). Finally, in all three groups, a statistically significant difference was maintained between the delayed posttest mean scores and the pretest mean scores (Professor: $p = .00$, Native speaker: $p = .00$, Peer: $p = .00$). This indicates that while participants may have experienced a significant decrease in their productive performance from Time 2 to Time 3, overall, interaction with any of the three interlocutor types still led to a lasting effect after one week. The paired t-test results for the written production assessment data are found in Table 32.
Table 32. Paired t-test results for written production assessment data, separated by type of interlocutor.

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Mean difference</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor (n = 27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>37.50</td>
<td>21.46</td>
<td>9.08</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-12.96</td>
<td>13.44</td>
<td>-5.01</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>24.54</td>
<td>15.65</td>
<td>8.15</td>
<td>.00</td>
</tr>
<tr>
<td>Native speaker (n = 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>38.83</td>
<td>20.86</td>
<td>9.31</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-12.17</td>
<td>13.50</td>
<td>-4.51</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>26.67</td>
<td>19.13</td>
<td>6.97</td>
<td>.00</td>
</tr>
<tr>
<td>Peer (n = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>35.56</td>
<td>18.37</td>
<td>10.60</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-14.31</td>
<td>17.22</td>
<td>-4.55</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>21.25</td>
<td>12.25</td>
<td>9.50</td>
<td>.00</td>
</tr>
</tbody>
</table>

In terms of Type of task, the significant main effect found indicates that participants were significantly more accurate in their written production of the lexical target items originating from the decision-making task, of which Figure 22 offers a clear visual. In addition, the significant main effect for Time indicates that all participants improved over time in their accurate written production of the targeted lexical items. However, the combination of the variables of Type of task and Time is key to interpreting the participants’ performances, which is evidenced by the significant interaction between Type of task and Time. This interaction indicates that not only did participants more accurately produce target items from the decision-making task when compared to the information gap task, but this difference held true over time: participants retained more productive knowledge of the decision-making task target items when their performance on the pretest and the delayed posttest are compared.

To statistically address whether the effect of Type of task on participants’ performance on L2 lexical production tasks lasts after one week, paired t-tests were conducted to compare the mean scores obtained on the pretest (Time 1), immediate posttest (Time 2), and delayed posttest (Time 3) for targeted lexical items originating from the decision-making task and information.
The results revealed that participants statistically increased their accurate written production of targeted lexical items originating from both experimental tasks from the pretest to the immediate posttest (Decision-making: \( p = .00 \), Information gap: \( p = .00 \)), but this performance also statistically decreased from the immediate posttest to the delayed posttest for both tasks (Decision-making: \( p = .00 \), Information gap: \( p = .00 \)). Ultimately, for both experimental tasks, a statistically significant difference was maintained between the delayed posttest mean scores and the pretest mean scores (Decision-making: \( p = .00 \), Information gap: \( p = .00 \)). This indicates that both experimental tasks still led to a lasting effect after one week. The paired t-test results for the written production assessment data are found in Table 33.

**Table 33.** Paired t-test results for production assessment data, separated by type of task.

<table>
<thead>
<tr>
<th>Experimental task</th>
<th>Mean difference</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision-making (( n = 82 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>43.19</td>
<td>23.54</td>
<td>16.61</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-15.96</td>
<td>18.66</td>
<td>-7.74</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>27.24</td>
<td>17.13</td>
<td>14.40</td>
<td>.00</td>
</tr>
<tr>
<td>Information gap (( n = 82 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2 vs. T1</td>
<td>31.10</td>
<td>19.29</td>
<td>14.60</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T2</td>
<td>-10.47</td>
<td>15.70</td>
<td>-6.04</td>
<td>.00</td>
</tr>
<tr>
<td>T3 vs. T1</td>
<td>20.63</td>
<td>17.68</td>
<td>10.57</td>
<td>.00</td>
</tr>
</tbody>
</table>

In sum, participants in all interlocutor groups statistically increased their productive performance on lexical items originating from both experimental tasks over time. Although Type of interlocutor did not play a large role in participants’ accurate written production of the targeted lexical forms, participants were able to accurately produce more targeted lexical forms originating from the decision-making task than the information gap task, and this difference in performance between experimental tasks held true over time.
Summary of Findings for Research Questions 1 and 2

A combination of Type of interlocutor (professor vs. native speaker vs. peer) and Type of task (decision-making vs. information gap) differentially affected 1) the number of LREs produced during interaction and 2) the amount of talk on task. On the decision-making task, participants from the native speaker and professor interlocutor groups produced more LREs and more talk on task than the peer interlocutor group. In comparison, on the information gap task, participants produced a similar number of LREs and amount of talk on task, regardless of interlocutor group. With regard to accurate recognition and written production of the targeted lexical forms, participants in all interlocutor groups statistically increased their receptive and productive performance on lexical items originating from both experimental tasks over time. However, participants who interacted with a professor retained more receptive knowledge over time of the targeted lexical items from both experimental tasks. Nonetheless, Type of interlocutor did not play a large role in participants’ accurate written production of these forms. Finally, in terms of the effects of Type of task, participants were significantly more accurate in their recognition and written production of the targeted lexical items originating from the decision-making task.

Research Question 3

3) What is the relationship between learners’ state anxiety and
   a) the number of LREs produced during interaction?
   b) the amount of talk on task?
   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?
d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

e) Type of interlocutor?

f) Type of task?

**RQ3a: Language-Related Episodes**

First, the means and standard deviations for participants’ overall state, interlocutor, and mode anxiety were computed. For overall state anxiety, the maximum score participants could receive was 75, with each question being worth 5 points. For interlocutor and mode anxiety, each participant’s score was calculated by averaging their responses on the questions connected to these subcategories, which resulted in a maximum score of 5. The descriptive statistics for all self-reported anxiety data collected in the present study, separated by Type of interlocutor and Type of task, can be seen in Table 34.

**Table 34. Participants’ self-reported state anxiety levels.**

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Anxiety levels, M (SD)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decision-making task</td>
<td>Information gap task</td>
<td></td>
</tr>
<tr>
<td>Professor (n = 27)</td>
<td>Overall anxiety*</td>
<td>33.04 (7.99)</td>
<td>34.41 (9.30)</td>
</tr>
<tr>
<td></td>
<td>Interlocutor anxiety**</td>
<td>2.53 (0.47)</td>
<td>2.51 (0.37)</td>
</tr>
<tr>
<td></td>
<td>Mode anxiety**</td>
<td>2.13 (0.52)</td>
<td>1.96 (0.57)</td>
</tr>
<tr>
<td>Native speaker (n = 25)</td>
<td>Overall anxiety</td>
<td>30.42 (6.98)</td>
<td>32.40 (10.26)</td>
</tr>
<tr>
<td></td>
<td>Interlocutor anxiety</td>
<td>2.40 (0.59)</td>
<td>2.37 (0.62)</td>
</tr>
<tr>
<td></td>
<td>Mode anxiety</td>
<td>1.81 (0.78)</td>
<td>1.92 (0.88)</td>
</tr>
<tr>
<td>Peer (n = 30)</td>
<td>Overall anxiety</td>
<td>31.38 (6.03)</td>
<td>33.43 (6.83)</td>
</tr>
<tr>
<td></td>
<td>Interlocutor anxiety</td>
<td>1.87 (0.47)</td>
<td>1.86 (0.47)</td>
</tr>
<tr>
<td></td>
<td>Mode anxiety</td>
<td>2.45 (0.84)</td>
<td>2.41 (0.80)</td>
</tr>
<tr>
<td>Total (n = 82)</td>
<td>Overall anxiety</td>
<td>31.65 (7.02)</td>
<td>33.44 (8.73)</td>
</tr>
<tr>
<td></td>
<td>Interlocutor anxiety</td>
<td>2.25 (0.58)</td>
<td>2.23 (0.57)</td>
</tr>
<tr>
<td></td>
<td>Mode anxiety</td>
<td>2.15 (0.76)</td>
<td>2.12 (0.78)</td>
</tr>
</tbody>
</table>
In order to examine the relationship between learners’ state anxiety and the number of LREs produced, six Pearson’s Product-Moment correlation analyses were conducted. The first statistical analysis, which addressed the relationship between learner participants’ overall state anxiety (based on the 15 questionnaire items from Baralt and Gurzynski-Weiss, 2011) and number of LREs produced on the decision-making task, found no correlation between these variables ($r = -.02, p = .85$). Regarding the relationship between learner participants’ overall state anxiety and the number of LREs produced on the information gap task, there was once again no correlation ($r = .02, p = .80$).

The other four correlation analyses examined the two subcategories of state anxiety used in the present study: Interlocutor anxiety, defined as the anxiety participants felt based on the person they were interacting with, and mode anxiety, defined as the anxiety participants felt based on the mode of interaction (i.e., text-based SCMC). In terms of interlocutor anxiety, no correlation was detected in relation to the number of LREs produced on the decision-making task ($r = .08, p = .40$) or the information gap task ($r = .10, p = .33$). In addition, there was no correlation between LREs produced on the information gap task and mode anxiety ($r = .10, p = .32$). However, there was a small negative correlation between LREs produced on the decision-making task and mode anxiety ($r = -.23, p = .02$). Therefore, the more anxious participants felt about interacting via text-based SCMC, the lower their number of LREs during the decision-making task.

**RQ3b: Talk on Task**

To address research question 3b, six Pearson’s Product-Moment correlation analyses were again conducted. For the decision-making task, there was no correlation between the
amount of talk on task and overall state anxiety \((r = -0.07, p = .35)\) or the amount of talk on task and mode anxiety \((r = -0.09, p = .27)\). However, a small positive correlation between amount of talk on task and interlocutor anxiety was found \((r = 0.18, p = .03)\). Thus, the more anxious participants were about interacting with the interlocutor with whom they were paired, the higher their word count during completion of the decision-making task. See Figure 23 for a visual representation.

![Figure 23. Correlation between interlocutor anxiety and amount of talk on the decision-making task.](image)

With regard to the information gap task, no significant correlations were found between participants’ amount of talk on task and their overall state anxiety \((r = -0.14, p = .07)\), their interlocutor anxiety \((r = 0.13, p = .09)\), or their mode anxiety \((r = 0.08, p = .30)\).

**RQ3c: L2 Lexical Recognition**

In order to address research question 3c, the following variables were considered: 1) participants’ overall state, interlocutor, and mode anxiety, and 2) participants’ lexical recognition gain scores. Participant gain scores were calculated by subtracting their Time 1 (pretest) scores
from their Time 2 (immediate posttest) scores (T2-T1), subtracting their Time 2 (immediate posttest) scores from their Time 3 (delayed posttest) scores (T3-T2), and subtracting their Time 1 (pretest) scores from their Time 3 (delayed posttest) scores (T3-T1). In this manner, it could be seen how participants improved in their lexical recognition immediately after completing the experiment, how much they retained one week later, and what they took away overall from participating in the experiment. The means and standard deviations for gain scores on the recognition assessment task, separated by Type of interlocutor and Type of task, are found in Table 35.

**Table 35.** Descriptive data for recognition assessment task gains (in percentages).

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Recognition gain scores, M (SD)</th>
<th>Decision-making task*</th>
<th>Information gap task *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor (n = 27)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>58.64 (18.41)</td>
<td>51.54 (24.35)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-2.16 (14.73)</td>
<td>0.00 (13.87)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>56.48 (18.10)</td>
<td>51.54 (25.53)</td>
<td></td>
</tr>
<tr>
<td>Native speaker (n = 25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>55.67 (18.43)</td>
<td>56.33 (17.72)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-7.33 (11.11)</td>
<td>-11.00 (14.77)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>48.33 (16.49)</td>
<td>45.33 (17.53)</td>
<td></td>
</tr>
<tr>
<td>Peer (n = 30)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>51.11 (17.61)</td>
<td>48.06 (21.18)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-6.66 (12.07)</td>
<td>-7.50 (11.44)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>44.44 (17.69)</td>
<td>40.56 (19.91)</td>
<td></td>
</tr>
<tr>
<td>Total (n = 82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>54.98 (18.18)</td>
<td>51.73 (21.34)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-5.39 (12.79)</td>
<td>-6.10 (13.92)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>49.59 (18.00)</td>
<td>45.63 (21.53)</td>
<td></td>
</tr>
</tbody>
</table>

*Percentages were calculated from a total of 12 targeted lexical items originating from each experimental task.

18 correlation analyses were conducted: 9 for the decision-making experimental task and 9 for the information gap experimental task, which were split into 3 gain scores (T2-T1, T3-T2, and T3-T1) and the 3 types of anxiety measured (overall, interlocutor, and state). Given that
numerous correlations were needed to address all aspects of these questions, Kendall’s tau-b correlation was used as a more conservative alternative to Pearson’s correlation.

For a review of the findings on the correlation between participants’ L2 lexical recognition gains and their self-reported state anxiety levels, see Table 36. While the majority of

<table>
<thead>
<tr>
<th>Anxiety type</th>
<th>Experimental task</th>
<th>Gains</th>
<th>n</th>
<th>Kendall's tau-b correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>Decision-making</td>
<td>T2-T1</td>
<td>80</td>
<td>-.11</td>
<td>.19</td>
</tr>
<tr>
<td>Overall</td>
<td>Decision-making</td>
<td>T3-T2</td>
<td>80</td>
<td>.10</td>
<td>.25</td>
</tr>
<tr>
<td>Overall</td>
<td>Decision-making</td>
<td>T3-T1</td>
<td>80</td>
<td>-.01</td>
<td>.89</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Decision-making</td>
<td>T2-T1</td>
<td>80</td>
<td>.16</td>
<td>.06</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Decision-making</td>
<td>T3-T2</td>
<td>80</td>
<td>-.06</td>
<td>.47</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Decision-making</td>
<td>T3-T1</td>
<td>80</td>
<td>.10</td>
<td>.22</td>
</tr>
<tr>
<td>Mode</td>
<td>Decision-making</td>
<td>T2-T1</td>
<td>80</td>
<td>-.10</td>
<td>.25</td>
</tr>
<tr>
<td>Mode</td>
<td>Decision-making</td>
<td>T3-T2</td>
<td>80</td>
<td>.03</td>
<td>.72</td>
</tr>
<tr>
<td>Mode</td>
<td>Decision-making</td>
<td>T3-T1</td>
<td>80</td>
<td>-.07</td>
<td>.44</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>T2-T1</td>
<td>82</td>
<td>-.19*</td>
<td>.02</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>T3-T2</td>
<td>82</td>
<td>.18*</td>
<td>.03</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>T3-T1</td>
<td>82</td>
<td>-.07</td>
<td>.41</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Information gap</td>
<td>T2-T1</td>
<td>82</td>
<td>.08</td>
<td>.31</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Information gap</td>
<td>T3-T2</td>
<td>82</td>
<td>.07</td>
<td>.43</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Information gap</td>
<td>T3-T1</td>
<td>82</td>
<td>.12</td>
<td>.14</td>
</tr>
<tr>
<td>Mode</td>
<td>Information gap</td>
<td>T2-T1</td>
<td>82</td>
<td>-.12</td>
<td>.14</td>
</tr>
<tr>
<td>Mode</td>
<td>Information gap</td>
<td>T3-T2</td>
<td>82</td>
<td>.08</td>
<td>.34</td>
</tr>
<tr>
<td>Mode</td>
<td>Information gap</td>
<td>T3-T1</td>
<td>82</td>
<td>-.07</td>
<td>.42</td>
</tr>
</tbody>
</table>

the correlation analyses conducted did not detect a correlation between these variables, there was a significant negative correlation ($\tau_b = -.19$, $p = .02$) between participants’ overall state anxiety and their lexical recognition performance gains from Time 1 to Time 2 on target items originating from the information gap task. In other words, the more anxious participants felt while completing the information gap task, the less they improved from pretest to immediate posttest in their recognition of the targeted lexical items found on the information gap task. See Figure 24 for a graph of the correlation.
Figure 24. Correlation between overall anxiety and gains from Time 1 to Time 2 in the recognition of targeted lexical items from the information gap task.

The correlations run with the recognition assessment data also detected a significant positive correlation (see Figure 25) between participants’ overall state anxiety and the difference in their performance from Time 2 to Time 3 ($\tau_b = .18, p = .03$). Therefore, the more anxious participants felt, the better they performed on the delayed posttest in their accurate recognition of the targeted lexical items when compared to their performance on the immediate posttest.

Figure 25. Correlation between overall anxiety and gains from Time 2 to Time 3 in the recognition of targeted lexical items from the information gap task.
RQ3d: L2 Lexical Production

In order to address research question 3d, the following variables were considered: 1) participants’ overall state, interlocutor, and mode anxiety, and 2) participants’ lexical production gain scores. The means and standard deviations for the gain scores on the written production assessment task, separated by Type of interlocutor and Type of task, are found in Table 37.

Table 37. Descriptive data for written production assessment task gains (in percentages).

<table>
<thead>
<tr>
<th>Interlocutor group</th>
<th>Written production gain scores, M (SD)</th>
<th>Decision-making task*</th>
<th>Information gap task*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor (n = 27)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>48.15 (26.08)</td>
<td>26.85 (18.82)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-19.44 (18.78)</td>
<td>-6.48 (11.63)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>28.70 (18.39)</td>
<td>20.37 (16.40)</td>
<td></td>
</tr>
<tr>
<td>Native speaker (n = 25)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>43.33 (24.65)</td>
<td>34.00 (19.68)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-13.33 (18.00)</td>
<td>-11.00 (15.91)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>30.00 (19.25)</td>
<td>23.00 (22.60)</td>
<td></td>
</tr>
<tr>
<td>Peer (n = 30)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>38.61 (19.75)</td>
<td>32.50 (19.37)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-15.00 (19.25)</td>
<td>-13.61 (18.24)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>23.61 (13.68)</td>
<td>18.88 (14.17)</td>
<td></td>
</tr>
<tr>
<td>Total (n = 82)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2-T1</td>
<td>43.19 (23.54)</td>
<td>31.10 (19.29)</td>
<td></td>
</tr>
<tr>
<td>T3-T2</td>
<td>-15.96 (18.67)</td>
<td>-10.47 (15.70)</td>
<td></td>
</tr>
<tr>
<td>T3-T1</td>
<td>27.24 (17.13)</td>
<td>20.63 (17.68)</td>
<td></td>
</tr>
</tbody>
</table>

*Percentages were calculated from a total of 12 targeted lexical items originating from each experimental task.

18 correlation analyses were conducted with the lexical production data: 9 for the decision-making experimental task and 9 for the information gap experimental task, which were split into 3 gain scores (T2-T1, T3-T2, and T3-T1) and the 3 types of anxiety measured (overall, interlocutor, and state). Given that numerous correlations were needed to address all aspects of these questions, Kendall’s tau-b correlation was used as a more conservative alternative to Pearson’s correlation.
For a review of the findings on the correlation between participants’ L2 lexical production gains and their self-reported state anxiety levels, see Table 38. While the majority of the analyses conducted did not find significant correlations between these variables, there was a very large negative correlation ($\tau_b = -.23, p = .004$) between participants’ overall state anxiety and their lexical production performance gains from Time 1 to Time 2 on target items originating from the information gap task. In other words, the more anxious participants felt while completing the information gap task, the less they improved from pretest to immediate posttest in their written production of the targeted lexical items found on the information gap task. See Figure 26 for a graph of the correlation.

### Table 38. Kendall’s tau-b correlations between participants’ self-reported anxiety and gain scores on the written production assessment task.

<table>
<thead>
<tr>
<th>Anxiety type</th>
<th>Experimental task</th>
<th>Gains</th>
<th>n</th>
<th>Kendall's tau-b correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>Decision-making</td>
<td>T2-T1</td>
<td>80</td>
<td>-.04</td>
<td>.62</td>
</tr>
<tr>
<td>Overall</td>
<td>Decision-making</td>
<td>T3-T2</td>
<td>80</td>
<td>.10</td>
<td>.25</td>
</tr>
<tr>
<td>Overall</td>
<td>Decision-making</td>
<td>T3-T1</td>
<td>80</td>
<td>.08</td>
<td>.34</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Decision-making</td>
<td>T2-T1</td>
<td>80</td>
<td>.13</td>
<td>.13</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Decision-making</td>
<td>T3-T2</td>
<td>80</td>
<td>-.09</td>
<td>.30</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Decision-making</td>
<td>T3-T1</td>
<td>80</td>
<td>.08</td>
<td>.35</td>
</tr>
<tr>
<td>Mode</td>
<td>Decision-making</td>
<td>T2-T1</td>
<td>80</td>
<td>-.05</td>
<td>.55</td>
</tr>
<tr>
<td>Mode</td>
<td>Decision-making</td>
<td>T3-T2</td>
<td>80</td>
<td>.005</td>
<td>.96</td>
</tr>
<tr>
<td>Mode</td>
<td>Decision-making</td>
<td>T3-T1</td>
<td>80</td>
<td>-.09</td>
<td>.27</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>T2-T1</td>
<td>82</td>
<td>-.23*</td>
<td>.00</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>T3-T2</td>
<td>82</td>
<td>.16*</td>
<td>.05</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>T3-T1</td>
<td>82</td>
<td>-.08</td>
<td>.31</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Information gap</td>
<td>T2-T1</td>
<td>82</td>
<td>-.04</td>
<td>.63</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Information gap</td>
<td>T3-T2</td>
<td>82</td>
<td>.007</td>
<td>.93</td>
</tr>
<tr>
<td>Interlocutor</td>
<td>Information gap</td>
<td>T3-T1</td>
<td>82</td>
<td>-.02</td>
<td>.77</td>
</tr>
<tr>
<td>Mode</td>
<td>Information gap</td>
<td>T2-T1</td>
<td>82</td>
<td>-.13</td>
<td>.13</td>
</tr>
<tr>
<td>Mode</td>
<td>Information gap</td>
<td>T3-T2</td>
<td>82</td>
<td>.12</td>
<td>.13</td>
</tr>
<tr>
<td>Mode</td>
<td>Information gap</td>
<td>T3-T1</td>
<td>82</td>
<td>.01</td>
<td>.90</td>
</tr>
</tbody>
</table>
Figure 26. Correlation between overall anxiety and gains from Time 1 to Time 2 in the production of targeted lexical items from the information gap task.

The correlations run with the written production assessment data also found a significant positive correlation (see Figure 27) between participants’ overall state anxiety and the difference in their performance from Time 2 to Time 3 ($τ_b = .16, p = .05$). Thus, the more anxious participants felt, the more they retained from the immediate posttest to the delayed posttest in terms of accurate written production.

Figure 27. Correlation between overall anxiety and gains from Time 2 to Time 3 in the written production of targeted lexical items from the information gap task.
Overall, for RQ3c and RQ3d, the more anxious participants felt while completing the information gap task, the less they improved from pretest to immediate posttest in their recognition and written production of the targeted lexical items, as evidenced by the significant negative correlations found. However, the more anxious participants also retained more receptive and productive knowledge of the information gap target items from posttest to delayed posttest, as evidenced by the significant positive correlations found. These findings are summarized in Table 39.

Table 39. Significant Kendall’s tau-b correlations between participants’ overall state anxiety and gain scores on the assessment tasks.

<table>
<thead>
<tr>
<th>Anxiety type</th>
<th>Experimental task</th>
<th>Assessment task</th>
<th>Gains</th>
<th>n</th>
<th>Kendall's tau-b correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>Recognition</td>
<td>T2-T1</td>
<td>82</td>
<td>-.19*</td>
<td>.02</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>Recognition</td>
<td>T3-T2</td>
<td>82</td>
<td>-.18*</td>
<td>.03</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>Production</td>
<td>T2-T1</td>
<td>82</td>
<td>-.23*</td>
<td>.00</td>
</tr>
<tr>
<td>Overall</td>
<td>Information gap</td>
<td>Production</td>
<td>T3-T2</td>
<td>82</td>
<td>.16*</td>
<td>.05</td>
</tr>
</tbody>
</table>

RQ3e and RQ3f: Type of Interlocutor and Type of Task

Anxiety data was collected separately for each experimental task and could not be combined for analysis since this would have influenced the findings regarding its relationship with Type of task and Type of interlocutor. As a result, research questions 3e and 3f will be addressed here together. In order to answer these research questions, four linear regressions were conducted. Participants’ overall state anxiety was entered as the dependent variable, and interlocutor group was input as the independent variable. Given that there were three interlocutor groups, a different interlocutor group was used as the baseline group for each of the two regressions conducted per experimental task to allow for a more detailed comparison of groups. In other words, two regressions were run with the anxiety data from the decision-making task, and the other two were run with the anxiety data from the information gap task. On the decision-
making task, no significant regression equation was found \((F(2,77) = .92, p = .40)\), with an \(R^2\) of .023. This indicates that interlocutor group did not predict participants’ overall state anxiety well. More specifically, only 2.3% of the total variation in participant state anxiety could be explained by their interlocutor group on the decision-making task. The coefficients further demonstrated that interlocutor group did not play a role in participants’ overall state anxiety for the decision-making task, with \(p = .19\) when comparing the native speaker and professor groups, \(p = .38\) when comparing the peer and professor groups, and \(p = .62\) when comparing the peer and native speaker groups.

On the information gap task, no significant regression equation was found \((F(2,79) = .34, p = .72)\), with an \(R^2\) of .008. This again indicates that interlocutor group did not predict participants’ overall state anxiety well. More specifically, only 0.8% of the total variation in participant state anxiety could be explained by their interlocutor group on the information gap task. The coefficients further demonstrated that interlocutor group did not play a role in participants’ overall state anxiety for the information gap task, with \(p = .41\) when comparing the native speaker and professor groups, \(p = .68\) when comparing the peer and professor groups, and \(p = .67\) when comparing the peer and native speaker groups. Therefore, neither task nor interlocutor seem to be predictors of the participants’ state anxiety.

**Summary of Findings for Research Question 3**

In summary, although for the most part there was no correlation between participants’ state anxiety and the number of LREs they produced on task, correlation analyses did show that the more anxious participants felt about interacting via text-based SCMC, the lower their number of LREs during the decision-making task. Regarding the relationship between state anxiety and
the amount of talk on task, the more anxious participants felt about interacting with the interlocutor with whom they were paired, the higher their word count during completion of the decision-making task. In terms of participant performance on receptive and productive measures, the more anxious participants felt while completing the information gap task, the less they improved from pretest to immediate posttest in their recognition and written production of the targeted lexical items. However, the more anxious participants also retained more receptive and productive knowledge of the information gap target items from posttest to delayed posttest. Finally, the regression analyses revealed that neither interlocutor nor task were predictors of participants’ overall state anxiety while completing the tasks.
CHAPTER 6: DISCUSSION, PEDAGOGICAL IMPLICATIONS, LIMITATIONS, FUTURE DIRECTIONS, AND CONCLUSION

This chapter discusses the results presented in the previous chapter and is structured similarly, with a discussion of the findings separated by research question. First, the findings on type of interlocutor and type of task (i.e., Research Questions 1 and 2) are discussed together in terms of their relationship with the dependent variables in the following order: 1) production of LREs, 2) talk on task, 3) learners’ performance on L2 recognition tasks over time, 4) learners’ performance on L2 production tasks over time. Then, this chapter discusses the relationship between learners’ state anxiety and 1) production of LREs, 2) talk on task, 3) learners’ performance on L2 lexical recognition tasks over time, and 4) learners’ performance on L2 lexical production tasks over time (i.e., Research Question 3). This is followed by a discussion of the pedagogical implications of these findings, as well as a reflection on the study’s limitations and potential avenues for future research addressing interaction in SCMC.

Discussion: Research Questions 1 and 2

1) To what extent does Type of interlocutor (peer vs. professor vs. native speaker) affect
   a) the number of LREs produced during interaction?
   b) the amount of talk on task?
   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?
   d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

2) To what extent does Type of task (decision-making vs. information gap) affect
   a) the number of LREs produced during interaction?
   b) the amount of talk on task?
c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?

d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

**RQ1a and RQ2a: Language-Related Episodes**

Results in this dissertation showed that a combination of Type of interlocutor (professor vs. native speaker vs. peer) and Type of task (decision-making vs. information gap) differentially affected the number of LREs produced during interaction. While participants from all interlocutor groups produced a similar number of LREs on the information gap task, participants from the peer interlocutor group produced significantly less LREs on the decision-making task. There was no significant difference between the native speaker and professor interlocutor groups in terms of LRE frequency.

First, it should be noted that since this is arguably the first study to compare professor and native speaker interlocutors, these findings provide further support for past interaction-based empirical studies where researchers have used either native speaker or teacher interlocutors but did not consider if these interlocutor characteristics may have impacted the study’s findings. An example of this is Van den Branden (1997), where the researcher presented himself to participants as the teacher, but the fact that he was also a native speaker of the language for the experiment was not taken into consideration when interpreting the results. “Teacher” and “native speaker” are not necessarily characteristics that go hand in hand – there are plenty of nonnative teachers found in L2 classrooms. At least in the current study’s mode (i.e., SCMC) and with the tasks employed, it would seem that these interlocutor characteristics did not differentially influence the interactional features present.
In terms of peer interaction, given that in the present study, negotiation occurred either just as often or less frequently when learners collaborated with a peer in comparison with interaction with the other two interlocutor types, these findings would seem to lend support for studies such as Van den Branden (1997), Zhao and Bitchener (2007), Slimani-Rolls (2005), Bueno Alastuey (2013), and Lee and Son (2019). Van den Branden (1997) found that the learner-learner and teacher-learner dyads produced similar amounts of negotiation, with variables such as language proficiency and type of feedback received having greater effects on the amount of negotiation than the type of interlocutor. Zhao and Bitchener’s (2007) study produced similar results, finding that incidental form-focused episodes occurred frequently during interaction between both teacher-learners and learner-learner.

On the other hand, these findings also contradict those of numerous empirical studies that have found that peer interaction results in the most negotiation for meaning when compared to interaction with other types of interlocutors, particularly native speakers (Alcón, 2002; Liu, 2017; Sato & Lyster, 2007; Shehadeh, 2001; Varonis & Gass, 1985). For a long time, researchers have argued that peer interaction is less stressful (Philp et al., 2014), affords a greater comfort level (Foster & Ohta, 2005), results in more opportunities for modified output (Shehadeh, 2001, 2003), and thus more learning opportunities (Sato & Lyster, 2007). In spite of these findings, very few empirical studies have compared different interlocutor types in SCMC (Bueno Alastuey, 2013; Liu, 2017). While Bueno Alastuey (2013) employed voice-based SCMC, Liu (2017) is the only other study to use text-based SCMC to investigate the effects of type of interlocutor on the features of interaction. She investigated the frequency of instances of negotiation, successfully resolved instances, and interactional strategy in NNS-NS and NNS-NNS dyads. NNSs were native speakers of Mandarin learning English, and the NSs hailed from
various parts of the English-speaking world. NNS-NNS dyads were divided into five H-H pairs and five L-L pairs (H and L representing high- and low-proficiency learners). Similarly, NNS-NS dyads were made up of five H-N pairs and five L-N pairs (N being native speakers).

Ultimately, Liu (2017) found that negotiation was more frequent in peer interaction, which contrasted with the findings of Bueno-Alastuey (2013).

In spite of Liu’s (2017) findings, researchers have long posited that SCMC can result in increased participation and improved learner attitudes (Chun, 1994; Kern, 1995), which could certainly explain why participants in the present study often produced a similar number of LREs regardless of their interlocutor. It could be that text-based SCMC has the potential to offer similar opportunities for negotiation and subsequent L2 learning regardless of the interlocutor, or it could be, as Lee and Son (2019) argue, that “there may be little influence of interlocutor – learner or NS, familiar or unfamiliar – on the frequency of LREs, which suggests that learners negotiate meaning in similar ways regardless of whom they interact with” (p. 163).

In terms of Type of task, an initial review of the findings suggests that the two-way information gap task resulted either in more or a similar number of LREs when compared to the decision-making task, which also required a two-way exchange of information but allowed for different solutions. This finding supports numerous previous empirical studies that have concluded that task type can influence the amount of negotiation occurring in interaction (e.g., Blake, 2000; Gilabert et al., 2009, Smith, 2003; Yilmaz, 2011; Yilmaz & Granena, 2010), and more specifically, those that have found that two-way information gap tasks, also known as jigsaw tasks, lead to more negotiation of meaning (e.g., Blake, 2000; Gilabert et al., 2009). On the other hand, these findings also contradict Smith (2003), who reported that his decision-making tasks resulted in more negotiation than jigsaw tasks. However, Smith also noted that it
may have been features of the task’s design (i.e., infusion of targeted lexical items) and not the task itself that resulted in more negotiation sequences on the decision-making task.

On the surface, it would thus seem that the findings of this dissertation with regard to LRE production can be neatly positioned within the framework of previous empirical research. However, upon closer examination of the data, it is very likely that the significant interaction found between Type of interlocutor and Type of task stems from participants’ prior knowledge of the targeted items on the decision-making task. Although the researcher made a concerted effort to choose unknown targeted lexical items for both experimental tasks by reviewing the pilot study data and the participants’ university-specific language curricula, it would seem that although many participants had not reached the stage in their knowledge where they could accurately produce some of these targeted lexical forms in writing, they could certainly recognize some of them. This observation is supported by a review of the descriptive data for participants’ lexical recognition and production at Time 1 (see Tables 26 and 30). Although participants in all interlocutor groups performed similarly at Time 1 in their production of both the decision-making and information gap task target items, there was a seemingly large difference between their recognition of the decision-making and information gap target items at Time 1. At the time of the pretest (Time 1), participants recognized more of the target items originating from the decision-making task than from the information gap task. If many participants already had receptive knowledge of some of the decision-making target items, it is logical that they would not need clarification regarding their meaning during completion of the decision-making experimental task, thus resulting in a lower number of LREs on that particular experimental task.
However, Table 26 also shows that the level of receptive knowledge of the decision-making target items was similar for all three interlocutor groups. Logically, the next question is, why did only the peer interlocutor group produce a lower number of LREs on the decision-making task? As previously discussed, the professor and native speaker interlocutors feigned ignorance of the targeted lexical items in order to allow the participants who interacted with them more opportunities for interaction and subsequent language learning. In addition, in some cases where a participant in one of these two interlocutor groups stated that they already knew the meaning of one of the professor or native speaker’s target items during completion of the experimental tasks, the professor or native speaker interlocutor performed comprehension checks and asked the participants to explain the target item, thus resulting in a language-related episode that didn’t occur during peer interaction under the same circumstances. During these comprehension checks, sometimes it was demonstrated that the participant did have knowledge of the targeted item, while other times the participant had misunderstood the targeted item. In the first example below from the decision-making task, the comprehension check revealed that the participant did not accurately recognize the targeted lexical item as she had claimed:

**Professor:** yo tengo: la licuadora, el sartén, la pala, el encendedor, la cubeta, y el tazón. Conoces alguno?
*I have: the blender, the frying pan, the shovel, the lighter, the bucket, and the bowl. Do you know any of them?*

**Part. 22:** Conozco una pala y un encendedor
*I know a shovel and a lighter.*

**Professor:** sí conoces la pala? que es?
*You do know the shovel? What is it?*

**Part. 22:** Es un contenedor, no?
*It’s a container, no?*

**Professor:** No, la pala es algo que usas para limpiar y mover la nieve de tu casa y del camino alrededor de tu casa en el invierno. todos los que tienen casas aquí en DC, también necesitan palas. a veces los niños llevan palas a la playa
para jugar también. Es muy larga, pero no es un contenedor. Creo que estas pensando en la cubeta.

No, the shovel is something you use to clean and move the snow from your house and the path around your house in the winter. Everyone who has a house here in DC, they also need shovels. Sometimes children take shovels to the beach to play too. It’s very long, but it’s not a container. I think you’re thinking of the bucket.

Part. 22: Una cubeta es algo que tu pones cosas en?

A bucket is something you put things in?

Professor: si, la cubeta es algo grande que puede ser usada para transportar agua, o también para limpiar.puedes poner jabon y agua dentro de la cubeta para limpiar el piso.

Yes, the bucket is a big thing that can be used to transport water, or also to clean. You can put soap and water in the bucket to clean the floor.

Part. 22: Si, yo entiendo ahora

Yes, I understand now

In this second example from the decision-making task, the comprehension check revealed that the participant did have knowledge of the targeted lexical item and an LRE was produced:

Nat. speaker: entonces entiendes la comida enlatada y la hielera bien?

So you understand the canned food and the cooler?

Part. 58: si

Yes

Nat. speaker: qué es la hielera?

What’s the cooler?

Part. 58: Mantiene la temperatura de una cosa.

It maintains the temperature of something.

Nat. speaker: Ok

Although interactions of these sorts were the most common, there were other cases where the participant would explain the targeted lexical form without being prompted to do so, or where LREs occurred because the participant conveyed a lack of understanding regarding a lexical form that the professor or native speaker interlocutor used to explain one of the target items, as described in the previous chapter.
Another factor that may have played a role in the significantly lower number of LREs produced in the peer interaction group on the decision-making task was the participants’ occasional inattention to the task’s directions. The nature of the information gap task required that participants exchange information on all twelve target items in order to successfully complete the task, while the decision-making task could have, in theory, been completed without discussing all twelve targeted lexical forms since participants only had to decide on six total items to bring with them to the deserted island. The researcher made a conscious effort to avoid this issue by providing the participants with very explicit instructions and by closely monitoring the peer interactions on a computer adjacent to hers, but there were occasions where participants did not follow these instructions closely and the researcher overlooked it. The excerpt below provides an example where participants decided to end the decision-making task even though they had not discussed all twelve target items:

Part. 19:  
tenemos seis, corecto?  
*We have six, right?*

Part. 76:  
si, tu piensas que nosotros necesitamos otras cosas? Me alegro con mis cosas  
*Yes, do you think we need other things? I’m happy with my things*

Part. 19:  
no pienso y tu?  
*I don’t think so and you?*

Part. 76:  
no pienso tambien. Gracias y adios  
*I don’t think so either. Thanks and goodbye*

In these cases, it was observed that participants were more concerned with completing the task as quickly as possible than with their comprehension of the targeted lexical items. However, it can be argued that the same situation is often found in the FTF or online classroom, where without close monitoring, learners can just as easily ignore their teacher’s instructions regarding discussing all targeted vocabulary for the sake of completing a task more quickly.
In summary, it would seem that a combination of Type of interlocutor and Type of task differentially affects the production of LREs, at least on the decision-making task and in the context of the current study. However, there is a strong likelihood that participants’ prior receptive knowledge of the targeted lexical items originating from the decision-making experimental task may have played a role. Therefore, I argue that in the SCMC medium, interaction with any of the three interlocutor types may provide similar opportunities for negotiation in terms of the production of LREs. However, it is difficult to make any concrete or generalizable arguments regarding the effects of Type of task on frequency of LREs. Ultimately, it may have been certain features of the tasks (i.e., target lexical items) and not the tasks themselves that determined the results of the study.

**RQ1b and RQ2b: Talk on Task**

Results in this dissertation indicated that a combination of Type of interlocutor (professor vs. native speaker vs. peer) and Type of task (decision-making vs. information gap) differentially affected the amount of talk on task. Participants in the native speaker and professor interlocutor groups produced many more words on the decision-making task than the peer group. In comparison, participants in all interlocutor groups produced similar amounts of language on the information gap task.

Although output is a key component of interaction according to Long’s (1996) Interaction Hypothesis, the quantification of participants’ output is not something that has been widely addressed by empirical research in ISLA. Past empirical research has often focused on specific features of the output, addressing these features using both qualitative and quantitative analysis. However, as Poupore (2016) argued, “while producing more language does not necessarily entail
a direct link with acquisition, it may increase its chances” (p. 731). When comparing this dissertation’s findings with previous empirical studies that have quantified output, the results of this study lend support for some while contradicting others. For instance, in terms of type of interlocutor, this dissertation reported that participants who interacted with a native speaker or professor produced similar amounts of output on both experimental tasks. However, when the peer interlocutor group is compared to the native speaker and professor interlocutor groups, participants in the peer interlocutor group produced many less words on the decision-making task but performed similarly to the other groups on the interaction gap task. Thus, this study’s findings are in line with Sato (2015), who compared dyadic interaction between L2 learners and native speakers using two information gap tasks and found that learners and native speakers produced similar amounts of output. In addition, they support the findings of Seedhouse (1999), who argued that during peer interaction, there is a tendency for learners to minimize their linguistic output and produce only the language that is necessary for completion of the task. However, it is important to consider that any effort by participants to minimize talk on task may have been exacerbated by the effort required to type out their responses in a language other than their L1.

These findings also contradict those of numerous researchers (e.g., Loewen & Sato, 2018; Philp, Adams, and Iwashita, 2014; Poupore, 2016; Sato & Ballinger, 2016) who have argued that a greater feeling of comfort will help learners produce more linguistic output when interacting with a peer. This argument will also be discussed below in relation to the findings on the relationship between anxiety and type of interlocutor. Most importantly, the fact that the present study contradicts the findings of many researchers who have investigated the features of peer interaction highlights the importance of task type when interpreting the results of any task-based
interaction study, such as in the present study, where the significant interactions observed indicate that the combined effects of these two variables cannot be separated.

The interpretation of these findings does not come without limitations. As with research questions 1a and 2a, it is likely that participants’ prior knowledge of the targeted lexical items on the decision-making task played a role. Before participating in the experiment, most participants could not accurately produce any of the targeted lexical forms in writing, but it would seem they could recognize some of them, as demonstrated in Tables 26 and 30. If many participants in the peer interlocutor group already had receptive knowledge of some of the decision-making target items, it makes sense that they would not describe these items in detail to a peer. A reduced need to describe target items from the decision-making task means that participants would have to talk less in order to successfully complete the task. On the other hand, the information gap task required an exchange of information regarding all 12 targeted lexical items, regardless of interlocutor group. Furthermore, as previously discussed, the professor and native speaker interlocutors feigned ignorance of the targeted lexical items in order to allow the participants who interacted with them more opportunities for interaction and subsequent language learning. In addition, in some cases where a participant in one of these two interlocutor groups stated that they already knew the meaning of one of the professor or native speaker’s target items during completion of the experimental tasks, the professor or native speaker interlocutor performed comprehension checks and asked the participants to explain the target item, thus producing additional linguistic output that was not produced during peer interaction under the same circumstances. This would explain why participants in the native speaker and professor interlocutor groups produced many more words than the peer group on the decision-making task, but a similar number of words on the information gap task.
Another factor that may have played a role in the significantly lower amount of talk on the decision-making task in the peer interaction group was the participants’ occasional inattention to the task’s directions. The nature of the information gap task required that participants exchange information on all twelve target items in order to successfully complete the task, while the decision-making task could have, in theory, been completed without discussing all twelve targeted lexical forms since participants only had to decide on six total items to bring with them to the deserted island. The researcher made a conscious effort to avoid this issue by providing the participants with very explicit instructions and by closely monitoring the peer interactions on a computer adjacent to hers, but there were occasions where participants did not follow these instructions closely and the researcher overlooked it.

In summary, it would seem that a combination of Type of interlocutor and Type of task differentially affects the amount of talk produced on task, at least in the context of the current study’s experimental design. However, there is a strong likelihood that participants’ receptive knowledge of the decision-making task items prior to the start of the experiment may have played a role. Therefore, I argue that in the SCMC medium, interaction with any of the three interlocutor types may provide similar opportunities for talking on task. However, it is difficult to make any concrete or generalizable arguments regarding the effects of Type of task on amount of talk on task. Ultimately, it may have been certain features of the tasks (i.e., target lexical items) and not the tasks themselves that determined the results of the study.

**RQ1c and RQ2c: L2 Lexical Recognition**

ANOVA results in the present study showed that in terms of the relationship between type of interlocutor and L2 lexical recognition, no main effect was found, which meant that Type
of interlocutor did not significantly affect participants’ recognition of the targeted lexical items. However, a significant interaction between Type of interlocutor and Time was found. This is because the lexical recognition performance of the professor interlocutor group did not significantly decrease from the immediate posttest to the delayed posttest. In fact, a review of the descriptive data determined that, on average, the professor group retained almost all of their receptive knowledge of the targeted lexical items from both experimental tasks one week after the experiment was completed.

The lack of main effect for Type of interlocutor on L2 lexical recognition can likely be attributed to the patterns of interaction observed during the experiment. A qualitative review of the transcript data determined that participants and interlocutors most often engaged in collaborative patterns of interaction regardless of interlocutor type, with all but two dyads in the peer interlocutor group engaging in collaborative patterns of interaction. This supports the empirical arguments of Storch (2002), whose analysis, conducted within a sociocultural framework, showed that the collaborative pattern was the most conducive to language development. According to Storch, collaborative patterns of interaction require both high mutuality and high equality, where mutuality refers to “the level of engagement with each other’s contribution” (p. 127) and high mutuality is described as being “rich in reciprocal feedback and a sharing of ideas” (p. 127). Storch also explained that equality is not just equal turn-taking or equal contributions from both interactants, but rather an “equal degree of control over the direction of a task” (p. 127). A collaborative pattern of interaction that was observed in a peer dyad can be seen in the excerpt below:

Part. 5: en similar, tengo las cerillas, los son poquitos objetos de madera que usar a encender cosas. como tiene un encendedor, no necesitamos llevarlas. ¿Piensas?
Similarly, I have matches, they’re small wooden objects that are used to light things. Since you have a lighter, we don’t need to bring them. Don’t you think?

Part. 11: estoy de acuerdo con tu. 
I agree with you.

Part. 5: esto es bueno, llevaremos el encendedor. 
This is good, we’ll bring the lighter.

Part. 11: yo tengo un sartén que es importante para cocinar comida en el fuego. 
I have a pan which is important to cook food on the fire.

Part. 5: ¿es necesario si tenemos un encendedor y podemos buscar a madera? No podemos traer mucho, solamente seis objetos. 
Is it necessary if we have a lighter and we can look for wood? We can’t bring a lot, only six objects.

Part. 11: ok. que otras cosas debemos traer? Que cosas tu tienes? Ahora, temenos el encendedor, la licuadora y las cerillas. Necesitamos 3 otras objetos. 
Ok. What other things should we bring? What things do you have? Now, we have the lighter, the blender and the matches. We need 3 other objects.

Part. 5: el bronceador, aplique esto para salvas de el sol. es necesario si estamos en una isla que no tenga personas? 
The sunscreen, apply this to save yourself from the sun. Is it necessary if we’re on an island that doesn’t have people?

Part.11: yo creo que el bronceador es importante para nosotros. Debemos llevarlo. 
I think the sunscreen is important for us. We should bring it.

These peers provided each other with feedback on the importance of their items (i.e., the matches, the lighter, the pan, and the sunscreen), and they had equal opportunities to describe their items, showing high mutuality. While Participant 5 described the matches, Participant 11 described the pan. They also both ultimately made decisions about which items to bring to the island – together they decided to bring the sunscreen and the lighter, but not the matches. This shows that they controlled the direction of the task equally, thus demonstrating high equality.

Beyond the collaborative patterns that were seen in peer interaction, it was also observed that some participants in this interlocutor group chose to incorporate humor, something that would also be expected in the face-to-face mode, and could have potentially lowered any task-related anxiety. On the state anxiety questionnaire, one participant noted that “my anxiety was
much lower than I anticipated. I expected to be really nervous like a test, but it was actually pretty fun.” The use of humor could also indicate that participants generally felt extremely comfortable with one another. In the excerpt below, one participant joked about the uses of one of the items, and the other participant joked about the experimental tasks:

Part. 4: Tu tienes más cosas?  
*Do you have more things?*

Part. 14: una, pero no pienso necesitamos la licuadora. porque la licuadora es por preparar las margaritas.  
*One, but I don't think we need. The blender. Because the blender is for preparing margaritas.*

Part. 4: Jaja, vale. Un momento.  
*Haha, ok. One moment.*

Part. 14: yo quiero una margarita, pero no es necesario  
*I want a margarita, but it's not necessary.*

Part. 4: Sí, después de esos ejercicios!  
*Yes, after those exercises!*

However, as previously mentioned, collaborative patterns of interaction were not limited to the peer interlocutor group; these patterns were observed in the professor interlocutor and native speaker interlocutor groups as well. Below is an example from one of the learner-professor dyads:

Professor: piensas que es importante la cubeta para nosotros?  
*Do you think the bucket is important for us?*

Part. 13: Es bueno pero tengo otro objetos que son más importante para nosotros como el encendedor.  
*It’s good but I have other objects that are more important for us like the lighter.*

Professor: bueno! y que es el encendedor?  
*Okay! And what is the lighter?*

Part. 13: Es un instrumento que crea fuego. Con madera de la isla y un encendedor, es posible para crea fuego y prepara comida y calor.  
*It’s an instrument that creates fire. With wood from the island and a lighter, it’s possible to create fire and prepare food and warmth.*
Ok I understand. I have a similar object: matches. Matches make fire but you can only use them once and then you throw them in the trash. Is the lighter the same or different from the matches?

Part. 13: Es diferente porque un encendedor usa el gasolina para crear fuego. Es disponible para muchos usos, y no solamente uno para cada cerilla.


Part. 13: Bueno, de acuerdo con tú.

This interaction between learner and professor, which is characteristic of most, if not all, of the interactions in the professor interlocutor group, represents both high mutuality and high equality. The interaction demonstrates high mutuality because both interactants had equal opportunity to describe their items. While the learner participant described the lighter, the professor participant described the matches. They also provided each other with feedback about the importance of the items – the learner participant indicated that he thought they did not need to bring the bucket on the island, and the professor participant expressed that she thought the lighter was more useful than the matches. In addition, high equality was maintained since ultimately both participants had a say in the items that they chose to bring to the island.

These patterns of interaction were also observed in the native speaker interlocutor group, as seen in the following excerpt:

Native: la cubeta es algo grande que usas para limpiar típicamente. puedes poner mucha agua adentro y también jabón. y usas el agua con jabón para limpiar el piso. esta cubeta tiene como 5 galones

Part. 70: si entiendo. podemos llevar la cubeta
Yes, I understand. We can bring the bucket.

Native: bien! estoy de acuerdo. la otra cosa que tengo es el tazón. el tazón es un Plato que puedes usar para comer cereal con leche

Good! I agree. The other thing I have is the bowl. The bowl is a plate that you can use to eat cereal with milk.

Part. 70: o si entiendo pero no se si necesitamos

Oh yes I understand but I don’t know if we need.

Native: no creo que sea importante. entonces tenemos 6 cosas?

I don’t think it’s important. So we have 6 things?

Part. 70: no tenemos 5 cosas. necesitamos uno mas

No, we have 5 things. We need one more.

This excerpt comes from the very end of the decision-making task, when the nonnative speaker participant had already described all of his items, and the native speaker was finishing describing hers. The nonnative speaker participant thought that they should bring the bucket on the island, and the native speaker agreed. Then, the nonnative speaker expressed uncertainty regarding the utility of the bowl, and the native speaker said she did not think it was important. Therefore, together they decided to bring the bucket, but not the bowl. Again, this interaction is demonstrative of both high mutuality and high equality, making it collaborative in nature.

Although most dyads demonstrated collaborative patterns of interaction on this task, there were instances where participants engaged in other patterns of interaction as well. The following is an example of a dyad that formed an expert/novice relationship:

Part. 36: Lo llevamos?

Should we bring it?

Part. 60: quieres describir todos, y después quieres decidir?

Do you want to describe them all, and then do you want to decide?

Part. 36: Pienso que solo necesitamos describir seis cosas.

I think we only need to describe six things.

Part. 60: “YOUR PARTNER MUST UNDERSTAND THE MEANING OF ALL 6 OF YOUR ITEMS” pienso que necesitamos describir todos los objetos “YOUR PARTNER MUST UNDERSTAND THE MEANING OF ALL 6 OF YOUR ITEMS” I think we need to describe all the objects.

Part. 36: Bueno. También, you tengo el encendedor. Uso eso para empezar un fuego. es rojo, y es muy caliente.

172
Fine. Also, I have the lighter. I use this to start a fire. It's red, and it's very hot.

[Later on]

Part. 60: pienso que las cosas más importantes son: el sartén, encendedor, el tazón, la cubeta, la comida enlatada, y las cerillas

I think the most important things are: the pan, lighter, the bowl, the bucket, the canned food, and the matches

Part. 36: está bien conmigo.

That’s fine with me.

Part. 60: bueno

Fine.

Part. 36: Fin?

The end?

Part. 60: sí

Yes

Storch (2002) defines an expert/novice relationship as having high mutuality but low equality. In this excerpt, the participants argue over the direction of the task and how to best complete it. It is clear that Participant 60 (i.e., the expert) takes control of the task and determines its direction, which translates to low equality. At the same time, both participants share information, engage with each other’s comments, and offer each other feedback, making it a relationship of high mutuality. However, it should be reiterated that this type of interactional pattern was not typical even among the peer interlocutor group, with only two dyads forming an expert/novice relationship. Nevertheless, this does indicate a small amount of variation in the interactional patterns observed during the decision-making task.

The reader may have noticed that the above excerpts all come from the decision-making task. It could be argued that the collaborative patterns of interaction observed time and time again in these excerpts could be a result of the decision-making task design, which requires that both participants in each dyad explain their items and then decide together which items to take on the island. On the other hand, it would seem that the design of the information gap task did
not allow for much variation in the interactional patterns between participants of any dyad due to the more controlled nature of the task, which only required equal sharing of information from both parties. Therefore, participants always completed the information gap task in a collaborative manner. Given that a collaborative relationship is purported to be most conducive to language development, it would be expected that participants would accurately recognize more of the information gap target items on the assessment tasks over time. Nevertheless, this was not the case since participants performed similarly on target items from both experimental tasks. This potentially offers further evidence for the role of learner participants’ prior knowledge of the decision-making task target items.

In summary, if participants created a collaborative relationship regardless of the interlocutor, then according to Storch’s argument, it is logical that all interlocutor groups would benefit linguistically from the interactions in similar ways. In the present study, participants in all interlocutor groups improved significantly over time in their recognition of the targeted lexical items. Thus, perhaps the argument can be made that Type of interlocutor is not as instrumental to L2 lexical development as the patterns of interaction are. Rejecting Type of interlocutor as a key variable in interaction would support the arguments of Lee & Son (2019), who also claimed that Type of interlocutor did not play a significant role in L2 development.

In a more general sense, this finding supports a large body of research that has concluded that interaction facilitates L2 learning (see Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010; Ziegler, 2016 for reviews), and more particularly, it supports previous empirical studies such as Ellis et al. (1994), Kim (2017), Newton (2013), and Tocaimaza-Hatch (2016), which all found a positive effect for interaction on vocabulary learning. Furthermore, given that participants in the peer interlocutor group did significantly
improve their lexical recognition of the target items over time, the findings of the present study support a large body of research that has found that peer interaction, when examined on its own or in comparison with NNS-NS interaction, can facilitate L2 development in both via FTF interaction (Lyster, Saito, & Sato, 2013; McDonough, 2004; Philp & Iwashita, 2013; Sato & Lyster, 2012; Toth, 2008) and SCMC (Blake, 2000; Pellettiere, 2000; Smith, 2003; Yilmaz and Granena, 2010; Yilmaz, 2011; Zeng, 2017; Kim, 2017). At the same time, it also contradicts the findings of Sippel (2019), who found that peer interaction on its own was not effective in supporting lexical development – participants also needed to be provided with instruction on the targeted lexical items. Ultimately, the findings of this dissertation with regard to native speaker interlocutors and peer interlocutors supports extensive previous research, but the addition of the findings for professor interlocutors extends this research.

Nevertheless, it still must be addressed that, unlike the other interlocutor groups, the lexical recognition performance of the professor interlocutor group did not significantly decrease from the immediate posttest to the delayed posttest. This finding is particularly interesting in light of the fact that the native speaker interlocutor and professor interlocutor were the same person (i.e., the researcher), although participants were unaware of this. This indicates that it was learners’ perception of their interlocutor, and not necessarily the patterns of interaction or the interlocutor herself, that affected their retention of the targeted lexical items. As will be discussed in the context of the third research question, it is possible that participants’ anxiety may have played a role. For instance, on the language background questionnaire, the majority of learner participants noted that in terms of who they felt most anxious interacting with, from least anxious to most anxious, the order was: 1) a friend or classmate who is equally proficient in Spanish, 2) a friend or classmate who is more proficient in Spanish, 3) their Spanish professor, 4)
a native speaker of Spanish, and 5) a Spanish professor they don’t know. However, it should be noted that the effects of participants’ state anxiety, or any other unknown interfering variable, were not significant enough to impact participants’ written lexical production, unlike their lexical recognition. As previously explained, knowledge works on a continuum, so although these learners may not have been able to accurately produce the same number of lexical items, they had enough knowledge to be able to recognize a significant number of them.

Another variable that may have potentially had an impact on the lexical retention of learner participants in the professor interlocutor group is their depth of processing (DOP). Hulstijn (1992, 2001) has argued that greater depth of processing (also referred to as “elaborate processing”), through inferring and hypothesis-testing, results in improved lexical retention. As learners attend to more features of a lexical item (e.g., orthographic, prosodic, semantic), depth of processing increases. Unfortunately, it cannot be ascertained definitively if DOP influenced the professor interlocutor group’s performance on the lexical recognition task since no concurrent measures (e.g., think aloud protocols) were employed in this dissertation. However, it is possible that these participants, upon being informed that they would be working with a professor to complete the experimental tasks, experienced higher DOP because they felt motivated to be on their “best behavior” with their interlocutor, who would likely judge them on their performance, even if said performance was not going to be graded. After all, these students were recruited directly from their language classes and the experiment took place in a computer laboratory on a college campus. Perhaps if the students had been recruited under different circumstances, or if the experiment had been carried out in a different environment, the participants may have behaved differently. Researchers can attempt to control for as many of the potential interfering variables as possible, but in the end, sometimes they can only offer possible
explanations and not definitive answers. On the state anxiety questionnaire, one participant noted the following:

The fact that I was interacting with a Spanish professor was stressful in the sense that they knew the language well and I didn't, but it was also comforting because I could rely on the fact they could tell what I was trying to say, even if I didn't say it perfectly. It was also nice that it was over chat so I had time to think, but I did still feel rushed because I didn't want them to have to wait a long time for me to respond.

In brief, it is possible that participants’ perception of the professor interlocutor, and therefore either their DOP or anxiety played a role in this particular interlocutor group’s superior recognition gains from Time 2 to Time 3. The researcher treated all participants the same regardless of whether she was acting as the professor or native speaker interlocutor. Thus, these are the most logical explanations, given that a qualitative comparison of interactions in the professor and native speaker interlocutor groups also found no significant differences between the interactional patterns in the two groups.

With regard to the effects of Type of task, the repeated measures ANOVA conducted to address Research Questions 1c and 2c found a significant main effect, indicating that participants were significantly more accurate in their recognition of the lexical target items from the decision-making task. However, the lack of significant interaction between Time and Type of task, paired with the significant main effect for Time, along with the paired t-test results for Type of task indicate that, overall, participants improved significantly over time in their accurate recognition of the targeted lexical items, but this occurred regardless of whether the target items originated from the decision-making or information gap experimental task.

As previously mentioned in the results chapter and discussed in light of the findings on number of LREs and amount of talk on task, participants’ significantly more accurate recognition of the decision-making task target items over time may be related to participants’ activation of
prior knowledge. Although the researcher made a concerted effort to choose unknown targeted lexical items for both experimental tasks by reviewing the pilot study data and the participants’ university-specific language curricula, it would seem that although many participants had not reached the stage in their knowledge continuum where they could accurately produce some of these targeted lexical forms in writing, they could certainly recognize some of them. This observation is supported by a review of the descriptive data for participants’ lexical recognition and production at Time 1 (see Tables 26 and 30). Although participants in all interlocutor groups performed similarly at Time 1 in their production of both the decision-making and information gap task target items, there was a seemingly large difference between the recognition of the decision-making and information gap target items at Time 1. At the time of the pretest (Time 1), participants recognized more of the target items originating from the decision-making task than from the information gap task.

If many participants already had receptive knowledge of some of the decision-making target items, this likely would have freed some attentional resources that could have been directed towards a lower number of unknown target items on the decision-making task. Since most participants had little to no receptive or productive knowledge of any of the lexical targets on the information gap task, there would have been a greater cognitive load, with attentional resources needing to be divided among a larger number of unknown target items. Given that diverting attentional resources to a larger number of target items would have required more cognitive effort, in order for participants to receptively retain these items, a higher level of DOP would have been needed. At the same time, since participants’ attentional resources were, in most cases, directed towards a smaller number of lexical items on the decision-making task, a lower level of DOP may have been sufficient for participants to recognize these items at Time 2.
and Time 3. Alternatively, it is also possible that participants’ DOP may have been higher while completing the decision-making experimental task simply due to the task’s design. However, any potential relationship between DOP and Type of task unfortunately cannot be determined for certain since no concurrent data was collected in this dissertation. Nevertheless, this argument is supported by the findings of empirical research such as Calderón (2013), who employed eye-tracking and think aloud protocols to address the depth of processing involved with different types of linguistic items (i.e., grammatical versus lexical). With regard to depth of processing with lexical items, she reported that participants’ accurate recognition of the target items increased as their DOP increased.

Another observation worth highlighting is that a number of participants stated on the state anxiety questionnaire that they felt that the information gap task was “harder”. One participant from the professor interlocutor group said, “it was harder for me to describe the objects well. I also had a harder time understanding.” Another participant from the peer interlocutor group commented on the number of lexical items included in the task: “I was anxious [because] the words were a bit harder than the last activity and there were more of them.” Although no distractor items could be used on the decision-making task due to the nature of the task’s design, there were 6 distractor items included in the information gap task. Therefore, each participant was presented with an image that included 12 labeled items, rather than the 6 they were presented with on the decision-making task. Although the distractor items were carefully chosen based on the findings of the pilot study and, in theory, should not have significantly diverted participants’ attentional resources from the targeted lexical items, participants may have felt visually overloaded by the information gap task, particularly if it was the second experimental task they completed. Furthermore, it is possible that the distractor items served their purpose and
effectively distracted participants from what was important (i.e., the targeted lexical items). In this sense, it can be argued that the decision-making task may have been superior for promoting recognition of novel L2 vocabulary over time.

Ultimately, participants improved significantly over time in their accurate recognition of the targeted lexical items, but this occurred regardless of whether the target items originated from the decision-making or information gap experimental task. Thus, although the decision-making task may be superior for promoting accurate lexical recognition, both experimental tasks could serve as important tools for L2 vocabulary learning via task-based interaction. This finding supports empirical research that has concluded that interaction facilitates L2 learning (see Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010; Ziegler, 2016 for reviews), and more particularly, it supports previous empirical studies such as Ellis et al. (1994), Kim (2017), Newton (2013), and Tocaimaza-Hatch (2016), which all found a positive effect for task-based interaction on vocabulary learning.

**RQ1d and RQ2d: L2 Lexical Production**

ANOVA results in the present study showed that in terms of the relationship between Type of interlocutor and L2 lexical production, no main effect was found, which meant that Type of interlocutor did not significantly affect participants’ accurate written production of the targeted lexical items. However, results from the paired t-tests demonstrated that from the time of the pretest to the time of the delayed posttest, participants from all interlocutor groups significantly improved in their accurate written production of the targeted lexical forms.

Given the previously discussed findings on the variables of number of LRES, amount of talk on task, and L2 lexical recognition, it is unsurprising that Type of interlocutor did not
significantly affect participants’ accurate written production over time. Just as with participants’ L2 lexical recognition, the lack of main effect for Type of interlocutor on L2 lexical production can likely be attributed to the patterns of interaction observed during the experiment. As discussed above in detail in relation to the findings for RQ1c, participants and interlocutors most often engaged in collaborative patterns of interaction regardless of interlocutor type. This finding supports the empirical arguments of Storch (2002), whose analysis showed that the collaborative pattern was the most conducive to language development. If participants in all three interlocutor groups tended to form collaborative relationships with their interlocutors while completing the experimental tasks, then it is logical, according to Storch’s arguments, that participants in all three interlocutor groups would perform similarly on measures assessing their receptive and productive knowledge of the lexical forms targeted on those experimental tasks. As a result, the argument can be made that perhaps Type of interlocutor does not have as much of an impact on L2 lexical development as the patterns of interaction do. Rejecting Type of interlocutor as a key variable in interaction would support the arguments of Lee & Son (2019), who also claimed that Type of interlocutor did not play a significant role in L2 development, as well as provide support for interaction-based empirical studies where researchers may have not considered the impact of interlocutor characteristics on their studies’ findings.

At the same time, a key component of the present study’s design that should be considered when interpreting these findings is the use of the SCMC medium. While the argument that the collaborative patterns of interaction observed in all interlocutor groups lead to similar receptive and productive gains is solid, it is important to consider if the medium of interaction was ultimately what caused interactional patterns in all interlocutor groups to be so similar. Researchers have posited that since text-based SCMC provides learners with a written
record of the interaction, it can provide added opportunities to attend closely to form and meaning (Ziegler & Mackey, 2017) and facilitate accuracy (Salaberry, 2000). Given that intonation, facial expressions, and gestures that often guide visual interaction are absent in text-based SCMC, learners are encouraged to use language (e.g., pragma-linguistic forms) to express themselves (Sykes, 2005) as they encounter and attempt to resolve gaps in their interlanguage (Swain, 2005). Furthermore, the added time to process linguistic input and plan output can help foster a ‘safe space’ for anxious learners (Beauvois, 1992), who often fear judgment based on linguistic insecurities such as imperfect pronunciation (Kern, 1995). In addition, SCMC can result in increased participation and improved learner attitudes (Chun, 1994; Kern, 1995).

In the present study, it seems that a combination of all of these factors may have helped to create a similar environment for interaction regardless of the interlocutor type. This is more than just conjecture, since most learner participants mentioned these factors on the state anxiety questionnaire. One participant in the peer interlocutor group stated the following: “I wasn't worried that my partner would judge my ability and I had more time to think about my responses than in face-to-face interactions.” Also, one participant in the professor interlocutor group said, “because I couldn't see the person I was talking to, I felt more comfortable taking time to write my answer. I felt more comfortable making silly mistakes because I couldn't see her face.” Another in the same group commented the following about the SCMC medium and her interlocutor: “It made me a lot more comfortable than speaking to someone face-to-face. I feel like I got comfortable more quickly than I would with a professor in person.” Finally, a number of participants felt that completing the task via text-based SCMC improved their accuracy: “I really think the reading/writing part was what made everything so much easier. It gives you time to think and completely relax rather than having to anxiously stare into someone's face. I felt like
I was better able to demonstrate my Spanish skill because I could think about my answers.”

Thus, clear patterns emerged in terms of participants’ opinions on the SCMC medium. Many mentioned that they felt more comfortable interacting with their partner than they would have in person, and they appreciated the added time to process input and produce output. In addition, many participants found the SCMC medium to be more comfortable because they felt more confident in their reading and writing skills in Spanish than in their listening and speaking skills. Ultimately, it would seem that the lack of main effect for Type of interlocutor may be related to the collaborative patterns of interaction found in all three interlocutor groups, and these collaborative patterns may have emerged in similar ways in all groups at least partly due to the SCMC medium.

In the context of prior empirical research on interaction, these findings offer support for the overwhelming majority of studies that have investigated the effects of interaction on L2 learning and found facilitative effects (see Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010; Ziegler, 2016 for reviews). In terms of research that has addressed the effects of interaction on L2 lexical development, the present study can be added to those that have found that interaction has a positive impact (e.g., Ellis et al., 1994; Kim, 2017; Newton, 2013; Tocaimaza-Hatch, 2016). In addition, in line with De la Fuente (2003) and many other empirical studies, this dissertation found that all groups performed better on receptive measures than on written productive measures over time. This is no surprise, given that language learners can often recognize new words before they can produce them in writing. Furthermore, given that participants in the peer interlocutor group did significantly improve their lexical production of the target items over time, these findings support a large body of research that has found that peer interaction can facilitate L2 development in the SCMC medium (Blake, 2000; Pellettieri,
However, many of these studies measured L2 development in terms of number of LREs, or using other qualitative measures. Therefore, the present study contributes to previous research by offering both a qualitative and quantitative perspective on the data.

In terms of the relationship between Type of task and participants’ L2 production, the repeated measures ANOVA revealed a significant main effect for Type of task. This indicates that participants were significantly more accurate in their written production of the lexical target items originating from the decision-making task. In addition, the significant main effect for Time indicates that all participants improved over time in their accurate written production of the targeted lexical items. This was further confirmed by the paired t-test results. Finally, a significant interaction between Type of task and Time was also observed. This interaction indicates that not only did participants more accurately produce target items from the decision-making task when compared to the information gap task, but this difference held true over time.

Given that most participants in all three interlocutor groups had some receptive knowledge of the decision-making tasks at the time of the pretest (i.e., Time 1), this finding is unsurprising. Although participants in all interlocutor groups came into the experiment with little to no productive knowledge of the lexical target items originating from either experimental task, their receptive knowledge of some decision-making target items prior to the experiment may have improved their performance on the written productive assessment measures over time. According to Leow’s (2015) Model of the Learning Process in Instructed SLA, repeated exposure to any lexical targets that participants could recognize prior to the experiment would have likely resulted in an activation of prior knowledge, thus facilitating the incorporation of these items into the learner’s systemized knowledge when combined with a certain level of
awareness. Under the right circumstances, this would lead learners to successful written production (i.e., the output stage).

In addition, as discussed in relation to the findings on participants’ L2 lexical recognition, since many participants already had receptive knowledge of some of the decision-making target items prior to participating in the study, it is possible that DOP played a role in the development of both their receptive and productive knowledge of the targeted lexical items. However, the existence of any relationship in this dissertation between DOP and Type of task cannot be determined with any certainty since no concurrent data was collected. In the case of most participants, not only were there fewer target items on the decision-making task that they needed to divert their attentional resources to, but there were also an additional six distractor items included on the information gap task. Therefore, it could be argued that the information gap task would have required more cognitive effort in order for participants to improve their recognition and written production assessment scores over time. In this sense, it can be argued that the decision-making task may have been superior for promoting accurate production of novel L2 vocabulary over time.

Although there was a significant main effect for Type of task, it should be remembered that participants improved significantly over time in their accurate production of the targeted lexical items, but this occurred regardless of whether the target items originated from the decision-making or information gap experimental task. Therefore, although the decision-making task may be superior for promoting accurate lexical production, both experimental tasks encourage L2 vocabulary learning, despite that lexical retention in the present study was relatively limited. This finding offers support for empirical studies that have concluded that interaction facilitates L2 learning (see Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006;
Spada & Tomita, 2010; Ziegler, 2016 for reviews), and more particularly, it supports previous empirical research such as Ellis et al. (1994), Kim (2017), Newton (2013), and Tocaimaza-Hatch (2016), which all found a positive effect for task-based interaction on vocabulary learning.

In summary, in terms of participants’ lexical recognition and production performance over time, Type of interlocutor mostly did not play a role. This can likely be attributed to the collaborative patterns of interaction observed in all three interlocutor groups, which in turn may have been a result of the medium of interaction: text-based SCMC. With regard to Type of task, the results indicate that participants were significantly more accurate in their recognition and production of the lexical target items from the decision-making task, a finding that is likely related to participants’ activation of prior knowledge. The relationship between these variables and participants’ state anxiety (i.e., RQ3) will be addressed in the section that follows.

**Discussion: Research Question 3**

3) What is the relationship between learners’ state anxiety and

   a) the number of LREs produced during interaction?

   b) the amount of talk on task?

   c) learners’ performance on L2 lexical recognition tasks? Does this effect last after one week?

   d) learners’ performance on L2 lexical production tasks? Does this effect last after one week?

   e) Type of interlocutor?

   f) Type of task?
RQ3a: Language-Related Episodes

There were no significant correlations found between 1) overall state anxiety or 2) interlocutor anxiety and the production of LREs on either experimental task. For mode anxiety, which was defined as the anxiety participants felt based on the mode of interaction (i.e., text-based SCMC), no correlation was detected with the production of LREs on the information gap task; however, there was a small negative correlation with the production of LREs on the decision-making task. This meant that the more anxious participants felt about interacting via text-based SCMC, the lower their number of LREs during the decision-making task.

A review of the mean scores for participants’ self-reported state anxiety levels (see Table 34), reveals that participants in the peer interlocutor group generally seemed to have the highest levels of mode anxiety. Also, it is important to recall that participants in the peer interlocutor group also produced significantly less LREs on the decision-making task than the other two interlocutor groups. These two findings, when viewed together as two pieces of one puzzle rather than two separate, unrelated puzzles, may explain this small negative correlation. On the decision-making task, learner participants tended to produce less LREs potentially due to the activation of prior knowledge, but they also tended to feel more anxious about interacting via SCMC. Given that LRE production was higher and mode anxiety was lower in the other interlocutor groups, these differences may have been just enough to create a small negative correlation between these two variables. Thus, the correlation, which is small, may be coincidental in nature, with activation of prior knowledge potentially serving as a confounding variable.

This finding should be interpreted with caution for a few different reasons. First, the correlation between LREs produced on the decision-making task and mode anxiety was
relatively small ($r = -0.23, p = 0.02$), meaning that the association between these variables was not particularly strong. Furthermore, the mode anxiety category was based on only 4 questionnaire items: 2 regular order statements and 2 reverse order statements.

Nevertheless, the lack of significant correlation between overall state anxiety and the production of LREs on either experimental task seems to contradict a large body of research that has established a negative relationship between FLA and L2 learning. However, this contradiction may have to do with how both learning and foreign language anxiety have been operationalized in previous studies. On the other hand, the fact that a negative relationship was not detected in the present study may be due to the medium of interaction: text-based SCMC.

For decades, researchers have argued that one of the possible advantages of text-based SCMC in comparison with other modes of L2 interaction is its potential to lessen learners’ anxiety (Chun, 1994; Kelm, 1992; Kern, 1995; Warschauer, 1996; Abrams, 2003). Since many researchers have posited that the one of the greatest sources of language anxiety stems from oral production (e.g., Gregersen & Horwitz, 2002; Hauck & Hurd, 2005; Krashen, 2003), this argument seems intuitive. However, there have been relatively few attempts to empirically address the variable of learner anxiety in SCMC (cf. Arnold, 2007; Baralt & Gurzynski-Weiss, 2011; Cote & Gaffney, 2018; Martin & Valdivia, 2017; Satar & Ozdener, 2008). The findings of this study lend support for Satar and Ozdener (2008) and Cote and Gaffney (2018) who found that learners who interacted via text-based SCMC experienced significantly lessened anxiety levels when compared to the FTF medium.

On the state anxiety questionnaire employed in this dissertation, participants noted time and time again, regardless of their interlocutor group, that they felt relaxed or less stressed because they were interacting with their partner through chat. One participant in the native
speaker interlocutor group stated, “my anxiety level was mostly very low throughout the whole thing and Spanish usually gives me lots of anxiety!” A review of the descriptive data in Table 34 shows that this was likely the case with most participants, given the relatively low mean scores for overall anxiety in all three interlocutor groups. The standard deviations also seem small considering that overall state anxiety was calculated out of a total of 75 possible points. Therefore, it is possible that the limited variability in participants’ self-reported overall state anxiety levels may have been the reason that no correlation was detected between overall state anxiety and LRE production.

RQ3b: Talk on Task

For the decision-making task, the correlational analyses on amount of talk on task and participants’ self-reported state anxiety detected no correlation between talk on task and 1) overall state anxiety or 2) mode anxiety. However, a small positive correlation between amount of talk on task and interlocutor anxiety was found. In other words, the more anxious participants felt about interacting with their interlocutor, the higher their word count during completion of the decision-making task. With regard to the information gap task, no significant correlations were found between participants’ amount of talk on task and their 1) overall state anxiety, 2) their interlocutor anxiety, or 3) their mode anxiety.

As was the case with LRE production, it is likely that the lack of significant correlation between overall state anxiety and participants’ amount of talk on either experimental task was due to the medium of interaction (i.e., text-based SCMC) and the low variability in overall state anxiety scores. Before interpreting the small positive correlation between amount of talk on task and interlocutor anxiety, it should be reiterated that participants completed these state anxiety
questionnaires after finishing each of the experimental tasks. Thus, some of the anxiety that participants felt about their interlocutor during the experimental tasks was likely a result of the behavior of their interlocutor during interaction, and not necessarily a result of participants’ preconceptions about their interlocutor. After all, the purpose of the state anxiety questionnaire was to measure participants’ anxiety in the moment of the experiment rather than to measure their trait anxiety.

Also key to interpreting this correlation is a review of the results and discussion regarding amount of talk on task (i.e., RQ1b and RQ2b). In Table 24, it can be seen that while all three interlocutor groups produced a similar amount of output on the information gap task, participants in the peer interlocutor group seemed to produce significantly less output on the decision-making task than the other two interlocutor groups. This observation is supported by the ANOVA results, which found a significant interaction between Type of task and Type of interlocutor. In addition, Tukey’s test confirmed that learner participants who interacted with a professor or a native speaker produced significantly more talk on task than the learner participants who interacted with a peer. As discussed previously, this is partly because in cases where a participant in the native speaker or professor interlocutor groups stated that they already knew the meaning of one of the target items during completion of the experimental tasks, the professor or native speaker interlocutor performed comprehension checks and asked the participants to explain the target item, thus producing additional linguistic output that was not produced during peer interaction under the same circumstances.

In light of the findings from RQ1b and RQ2b, the small positive correlation detected is not surprising. Since participants in the native speaker and professor interlocutor groups were often pushed to produce more output than those in the peer interlocutor group, this might have
caused them to feel anxiety or discomfort about the person they were interacting with. A review of the descriptive data on state anxiety levels in Table 34 also indicates that the native speaker and professor interlocutor groups generally reported higher levels of interlocutor anxiety than the peer interlocutor group. This increase in anxiety, however small, may have been a bit more noticeable on the decision-making task since many participants came into this task with a certain level of receptive knowledge of the target items. This receptive knowledge was questioned by the native speaker and professor interlocutor through the use of comprehension checks, and through insistence that participants in these groups describe the target items they claimed to already know. Therefore, the more participants felt they had to explain themselves to their interlocutor – in turn producing more output – the more anxiety they felt about their interlocutor.

However, as with the relationship between LREs and state anxiety, this finding should be interpreted with caution for a few different reasons. First, the correlation between amount of talk produced on the decision-making task and interlocutor anxiety was relatively small ($r = .18, p = .03$), meaning that the association between these variables was not particularly strong. Furthermore, the interlocutor anxiety category was based on only 6 questionnaire items: 3 regular order statements and 3 reverse order statements.

**RQ3c: L2 Lexical Recognition**

For the most part, no correlation was detected between participants’ L2 lexical recognition gains on the decision-making or information gap tasks and participants’ self-reported state anxiety levels (overall, mode, or interlocutor). Out of the 18 correlation analyses conducted, only 2 produced significant correlations. First, there was a significant negative correlation between participants’ overall state anxiety and their lexical recognition performance gains from
Time 1 to Time 2 on information gap target items. This meant that the more anxious participants felt while completing the information gap task, the less their recognition of the target items improved from the pretest to immediate posttest. Second, there was a significant positive correlation between participants’ overall state anxiety and their lexical recognition performance gains from Time 2 to Time 3 on information gap target items. Therefore, the more anxious participants felt while completing the information gap task, the more their accurate recognition of the target items was retained from the immediate to delayed posttest.

For the information gap task, the significant negative correlation between participants’ overall state anxiety and their lexical recognition performance gains from Time 1 to Time 2 will be addressed first. Previous empirical research has demonstrated that as FLA increases, learners are less able to process input, make connections between prior knowledge and new information, and access their linguistic system during L2 production (e.g. MacIntyre, 1999; MacIntyre & Gardner, 1991; Onwuegbuzie, Bailey, & Daley, 2000). In addition, increased anxiety can lead to less L2 vocabulary learning (MacIntyre & Gardner, 1994) and less L2 grammatical development (Kim & Tracy-Ventura, 2011). This is a plausible explanation for the gains from Time 1 to Time 2 in relation to participants’ overall state anxiety on the information gap task. Given that all of the lexical target items presented on the information gap task were unknown to participants, it is possible that some participants felt increased anxiety as a result. They may have been concerned about their ability to process a larger amount of information (when compared with the decision-making task) in a short period of time. Those participants who felt increased anxiety may not have been able to perform as well on the immediate posttest. As MacIntyre (1995) highlights, FLA can act as both the cause and the effect, both the chicken and the egg. The more learners
experience failure, the more their anxiety may increase, and in turn, affect their learning and performance.

Another variable that could have contributed to both of the significant correlations is how deeply participants were processing the information they were given. It is possible that if participants were attempting to process the lexical target items on the information gap task at a deeper level, they could have felt more anxious while completing the task. While their increased state anxiety might have initially hindered their performance on the lexical recognition assessment tasks at Time 2, when the participants completed the delayed recognition assessment a week later, any anxiety felt while completing the experimental task would have dissipated. Therefore, their high level of DOP during the task could have helped them retain more of their receptive knowledge of the targeted lexical items. Ultimately, this explanation would support numerous previous studies that have addressed the relationship between DOP and L2 learning. For instance, Hulstijn (1992, 2001) and Rott (2005) have argued that greater depth of processing leads to better retention of vocabulary, while Hsieh, Moreno, and Leow (2016) have posited that it also leads to potentially better retention of grammatical information.

As has come to be a common theme throughout this chapter, the participants’ potential prior knowledge of some of the decision-making target items combined with the use of text-based SCMC as the medium of interaction is likely at least part of the reason why no correlations were detected between participants’ state anxiety and their lexical recognition scores for the decision-making task.

In summary, while increased state anxiety during the information gap task resulted in less receptive gains from Time 1 to Time 2, more anxious participants’ high level of processing likely aided them in retaining these receptive gains over time. However, it should be noted that this is
simply conjecture – given that no concurrent data measures were employed in this dissertation, it cannot be empirically determined if participants’ DOP played a role in these findings.

**RQ3d: L2 Lexical Production**

Identical to the lexical recognition production data, only two significant correlations were detected out of the 18 conducted to address the relationship between participants’ L2 lexical production gains on the decision-making or information gap tasks and participants’ self-reported state anxiety levels (overall, mode, or interlocutor). A significant negative correlation was again detected between participants’ overall state anxiety and their lexical production performance gains from Time 1 to Time 2 on information gap target items. This meant that the more anxious participants felt while completing the information gap task, the less their production of the target items improved from the pretest to immediate posttest. In addition, there was again a significant positive correlation between participants’ overall state anxiety and their lexical production performance gains from Time 2 to Time 3 on information gap target items. In other words, the more anxious participants felt while completing the information gap task, the more their accurate production of the target items was retained from the immediate to delayed posttest.

As was the case with the L2 recognition data, it is possible that some participants felt increased anxiety on the information gap task because they were not familiar with any of the lexical targets. They may have been concerned about their ability to process a larger amount of information (when compared with the decision-making task) in a short period of time. Furthermore, it is possible that if participants were attempting to process the lexical target items on the information gap task at a deeper level, they could have felt more anxious while completing the task. While their increased state anxiety may have initially hindered their
performance on the written lexical production assessment tasks at Time 2, when the participants completed the delayed production assessment a week later, any anxiety felt while completing the experimental task would likely have dissipated. Therefore, their high level of DOP during the task could have helped them retain more of their productive knowledge of the targeted lexical items. This explanation would support numerous previous studies that have addressed the relationship between DOP and L2 learning and retention (e.g., Bistline-Bonilla et al., 2019; Hsieh et al., 2016; Hulstijn, 1992, 2001; Rott, 2005).

As was the case with the recognition data, participants’ potential activation of prior knowledge combined with the use of text-based SCMC as the medium of interaction is likely at least part of the reason why no correlations were detected between participants’ state anxiety and their lexical production scores on the decision-making task. However, when looking beyond the potential role of prior knowledge, this finding would seem to lend support for the findings of Valmori (2016). In her dissertation, she employed a spot-the-difference task and a picture-story task with 21 L1 English learners of Italian. Participants were recorded during interaction, and they were then asked to watch the videos and comment on the fluctuation of their state anxiety during the interactions. In a pretest-posttest design, their learning outcomes were measured using accuracy and fluency gain scores, and their use of two target structures was analyzed. She ultimately reported no significant correlation between participants’ gain scores and their state anxiety.

In summary, while increased state anxiety during the information gap task resulted in less written productive gains from Time 1 to Time 2, more anxious participants’ high level of processing likely aided them in retaining these written productive gains over time. However, it should be noted, as with lexical recognition, that this is pure supposition. No concurrent data was
collected in the present study, so no concrete conclusions can be made about participants’ levels of DOP.

**RQ3e and RQ3f: Type of Interlocutor and Type of Task**

The linear regression analyses revealed that neither Type of interlocutor nor Type of task were predictors of participants’ overall state anxiety while completing the tasks. The finding regarding Type of interlocutor is unsurprising. As previously discussed, participants engaged in collaborative interactional patterns regardless of who they interacted with. If participants in all interlocutor groups had a similar experience completing the experimental tasks, where they worked collaboratively with their partners, their opinions were valued and they were able to offer and receive feedback, it makes sense that Type of interlocutor would not be a source of their anxiety. On the anxiety questionnaire, one participant in the native speaker interlocutor group said, “My partner was very supportive and would complement my explanations, and that helped me feel more relaxed.” A participant in the professor interlocutor group expressed similar sentiments: “I felt mostly relaxed because I understood the explanations and was also able to learn new vocabulary. The professor was also very supportive when I gave my explanations.” Participants in the peer group also felt the same way. One participant said, “I think we were both relaxed enough to come to an agreement and I do think that texting and the fact that it was a peer helped.” Another peer interlocutor participant said that “working with a peer helped reduce my anxiety level because I knew my peer was also struggling when I was.” Ultimately, it seemed that most, if not all, participants felt comfortable with their interlocutor while completing the tasks. Not one participant made any negative comments about their interlocutor.
The same argument could also explain the lack of relationship between state anxiety and type of task. The leveling off of anxiety between the two tasks may have been due to the potential role of prior knowledge of some lexical items. While the decision-making task may have been less anxiety-producing than the information gap task because most participants already had receptive knowledge of a few of the target items, the information gap task could have made participants more anxious because the vocabulary words were perceived as “harder” and there were more of them, especially when the distractor items are considered. On the other hand, the decision-making task could have been more stressful than the information gap task because not only did participants have to exchange information with each other, but they also had to express their opinions and work together to reach a common decision. On the other hand, the information gap task may have been less stressful than the decision-making task because it was more “mechanic” in nature, with participants only needing to ask each other, “do you have ___?” and then describe their items to their partner if they did not have them. Not as much analytical thinking would have been involved since they did not need to share their opinions on the items or reach a consensus about their usefulness. It seems that this may be common when comparing decision-making and information gap tasks. For example, Newton (2013) observed the same task differences with his participants. Two groups of four adult learners of English with low-intermediate proficiency completed two information gap tasks (convergent) and two opinion gap tasks (divergent). He reported recognition gain scores of approximately 4 words per every 30 minutes on task. Slimani-Rolls (2005) also supported these conclusions. Her participants were 20 learners from varying L1 backgrounds who also spoke L2 English but were taking a university-level course on French for business purposes. They completed a consensus task (a type of jigsaw task), a one-way information exchange task, and a two-way information exchange.
task. She learned that many were more focused on completing the task rather than resolving any breakdowns in communication. She also argued that decision-making tasks offer more opportunities for genuine communication and linguistic manipulation. In summary, it is possible that the tasks balanced each other out, offering participants a similar emotional experience regardless of task.

There are also many arguments that can apply to both of these variables and why they may not have predicted participants’ state anxiety. First, there was not a lot of variability in participants’ scores on the state anxiety questionnaire, with most participants scoring in the low 30s. Considering that the maximum score on the questionnaire was 75, this demonstrates that overall, participants’ state anxiety was generally low, and the descriptive statistics in Table 34 indicate that it remained mostly stable regardless of interlocutor or task. There was some limited variability in mean scores from task to task or interlocutor to interlocutor, but these differences were not significant.

The strongest argument for why there was little variability in participants’ state anxiety levels may be the medium of interaction: text-based SCMC. Many researchers have posited that the one of the greatest sources of language anxiety stems from oral production (e.g. Gregersen & Horwitz, 2002; Hauck & Hurd, 2005; Krashen, 2003). Therefore, perhaps interacting via SCMC is a way of leveling the playing field for learners who might feel anxious for any number of reasons. Time and time again, participants noted on the state anxiety questionnaire that they felt relaxed because the task involved “typing” or “reading and writing” rather than “speaking and listening”. Out of 82 participants, only one mentioned that they would have preferred speaking in person to their partner rather than typing. From a social standpoint, the age group of the participants should be considered. All learner participants were between the ages of 18 and 24. In
this day and age, most young adults are accustomed to texting or chatting online with friends and family with frequency. Thus, the chatting aspect of this dissertation’s design would have likely contributed to the comfort level of most participants. In fact, on the language background questionnaire, 85.2% of participants reported that they enjoyed using a computer to communicate with others. On the state anxiety questionnaire, one participant in the peer interlocutor group noted that, “I felt relaxed because I didn't feel pressure to come up with an on-the-spot answer. I could think a little bit before I typed my response, without any of those "awkward silences" that I'd have in person.” Another participant mentioned that “I find it easier to learn Spanish visually, by typing words out.” A participant in the professor interlocutor group elaborated further on the visual aspect of learning vocabulary via text-based SCMC:

I felt myself learning faster with this method. I was able to refer back to things that were said earlier in the conversation without having to ask the professor to repeat them as I would in a classroom environment. I felt more in control and confident. There were fewer ways to communicate (e.g. no facial expressions or hand gestures), but seeing the conversation displayed in the chat log in front of me was very helpful in terms of retaining vocabulary and referring to what was previously said.

Ultimately, researchers have been touting these potential benefits of SCMC for lessening learners’ anxiety for decades (e.g., Beauvois, 1992; Cote & Gaffney, 2018; Satar & Ozdener, 2008), and this dissertation offers strong support for those arguments, both quantitatively and qualitatively.
Pedagogical Implications

The findings of the present study suggest that, overall, participants engaged in similar amounts of negotiation for meaning and created a collaborative relationship with their partner regardless of their interlocutor. This led all three interlocutor groups to improve a similar amount over time in their recognition and production of the targeted lexical items. Therefore, interacting with either a peer, native speaker, or professor can lead to similar opportunities for interaction and similarly improved performance over time.

These findings have many implications for the L2 classroom, particularly in the online setting. First, it suggests that collaborating with a peer of a similar proficiency on linguistic tasks can have linguistic benefits (i.e., improved lexical recognition and production) similar to interacting with a more knowledgeable partner, such as a native speaker or teacher. Thus, these findings support the use of task-based activities in the communicative classroom for aiding in vocabulary development. In this learning environment, learners often collaborate in peer dyads, so the findings of the present study support this practice. Furthermore, these findings confirm that the L2 teacher does not necessarily have to be directly involved in these classroom interactions in order for learners to see the benefits of interaction. Instead, the teacher may act as a guide to ensure that learners are completing the task as instructed.

In terms of the use of text-based SCMC in this dissertation, the ramifications are far-reaching. This medium was likely at least part of the reason why minimal differences in LRE production, L2 lexical recognition and production, and state anxiety were observed between interlocutor groups. An informal review of participant comments on the language background questionnaire revealed that use of text-based SCMC made participants feel more comfortable participating and collaborating with their partner, provided them added time to plan their
responses, and gave them a visual history of the interaction that they could refer back to, among other benefits. Although it is not entirely realistic to consider using text-based SCMC for task-based activities in a traditional classroom, L2 teachers could certainly incorporate text-based SCMC into task-based activities to be completed at home.

Nevertheless, these findings will not necessarily have the most implications for the traditional L2 classroom, but rather for online or hybrid language courses, e-tutoring, or Teletandem. For online language courses where all assignments are completed at home and online, text-based SCMC could be a valuable language learning tool for students who do not have the opportunity to develop their L2 vocabulary with peers in person. In terms of e-tutoring, anxious learners, or learners with time or scheduling constraints could choose to interact with their teacher through text-chat and still reap the benefits of the interaction. At the same time, learners who do not have the financial means to pay for a tutor might be relieved to know that they can choose to collaborate with a peer of a similar proficiency level instead and achieve similar results. Finally, one of the logistical difficulties often encountered with Teletandem is spotty internet, which can make it difficult for Teletandem partners to maintain a working video feed. Thus, Teletandem learners will be relieved to know that they do not necessarily need to establish a video connection in order to successfully participate in communicative tasks.

On another note, the findings of this dissertation are particularly relevant given the decision by many schools and universities across the globe to transition to virtual learning environments as a result of the COVID-19 pandemic, since many students, particularly those of a lower socio-economic profile, do not have access to Wi-Fi, or do not have a reliable internet connection or sufficient bandwidth to interact with others synchronously via video chat. Researchers have a responsibility to seek empirically-supported solutions for students that will
prevent quality, effective online instruction from ultimately being reserved for those of a higher socio-economic background. The conclusions that any of the three types of interlocutors is equally advantageous for promoting L2 lexical development, and that improved L2 vocabulary performance over time is possible when relying on entirely text-based interaction are certainly comforting under the circumstances, and will help disadvantaged students in particular. Although the tasks employed in this dissertation may not have led to the accurate recognition and production of a high number of the target items by the time of the delayed posttest, these findings, when coupled with those of previous empirical studies, do support the use of communicative tasks in the online classroom.

Given that anxiety has consistently correlated negatively with success in L2 learning (Horwitz, 2001; Teimouri et al, 2018), arguably one of the most important findings in the present study was that the use of text-based SCMC seemed to either lessen or in some cases entirely mitigate learners’ state anxiety as it related to the interaction. There was a limited relationship between participants’ anxiety and their performance on the assessment tasks over time, and in fact, the higher participants’ state anxiety was, the less their receptive and productive performance decreased from Time 2 to Time 3. Many language learners can suffer from some form of anxiety, whether it be state or trait anxiety, and the incorporation of linguistic tasks using text-based SCMC into the language curricula could level the playing field for the most anxious of language learners. In this sense, L2 teachers would be creating a more inclusive environment that offers similar opportunities for linguistic development to all of their students.

At the same time, the use of text-based SCMC does not come without pedagogical limitations. First, while text-based interaction alone can provide learners with writing and reading practice in the L2, since there is no oral component, there is no opportunity for learners
to practice their listening and speaking in the L2. Although reading and writing practice are pivotal to learners’ L2 development, listening and speaking practice are equally vital, particularly if the goal of the learners is to be able to successfully communicate in person with another speaker of the L2. Furthermore, given the current circumstances of the COVID-19 pandemic, which have caused most people to integrate communication via Zoom or other similar platforms into their daily routine, we may see an evolution in learners’ wants and needs in the classroom. Assuming that the United States and other countries where virtual learning has been largely implemented will be able to return to in-person learning post-COVID, many learners, even those who may have previously preferred interaction via Zoom, may ultimately prefer solely in-person instruction and interaction in the L2. This could be a result of so-called “Zoom fatigue” or it could also stem from negative experiences associated with the pandemic and Zoom, a platform that was relatively unknown or at the very least used sparingly by your average learner prior to the pandemic. Based on this argument, a replication of the present study after in-person learning has resumed could potentially produce vastly different findings.

Given that the potential activation of prior knowledge likely played such a significant role in the findings on the decision-making task, it seems unwise to attempt to extrapolate any pedagogical ramifications from the findings on type of task. Both tasks employed in the present study led to gains in recognition and production that were statistically significant, but it is important to remember that accurately producing 3 out of 12 targeted lexical items would not have translated to successful performance or a passing grade in a L2 classroom (Leow, 2019). Given the relatively low percentage of lexical items learned by the delayed posttest, perhaps the number of items may be reduced to address any potential cognitive overload. In addition, the tasks may be performed more than once to allow multiple exposures (Leow, 1998) to these items.
to encourage vocabulary development in addition to providing more opportunities to practice vocabulary in communicative activities. Finally, learning outcomes could be improved by lowering the number of lexical items per experimental task. This is because lexical items have to be learned individually, compared to a grammatical item that is typically based on one underlying rule that, once learned, serves as the foundation for facilitating the production of other grammatical forms.

**Limitations**

As is the case with any empirical study, this dissertation has numerous limitations that should be recognized here and addressed through further empirical research. Furthermore, given that this study is the first attempt in the field of ISLA to address the effects of three types of interlocutors in either the FTF or SCMC modes, it should be acknowledged as exploratory in nature. At the very least, replication is warranted before generalizing this dissertation’s findings.

Arguably, one of the greatest limitations of this study is the design of the decision-making task. Although the researcher made a concerted effort to choose unknown targeted lexical items for both experimental tasks by reviewing the pilot study data and all vocabulary included in the Spanish language program at the participants’ university, many participants were able to accurately recognize some of the decision-making task’s targeted lexical forms at the time of the pretest, even if they could not produce them in writing. Given that participants were presented with 24 possible answers for each question on the recognition assessment, it is unlikely that guessing played a role. Thus, participants’ prior receptive knowledge of some of the decision-making target items arguably altered the results on Type of task significantly. The findings in this study suggest that the information gap task may have been superior for language
learning in terms of the production of LREs and the amount of talk produced on task. At the same time, the findings also suggest that the decision-making task was superior in promoting accurate recognition and production of the targeted lexical items over time. When taken together, these findings seem contradictory, given that a large body of research on interaction has argued that the greater the negotiation for meaning, the better the learning outcomes (see Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006; Spada & Tomita, 2010; Ziegler, 2016 for reviews).

If the information gap task and the decision-making task had been created equal (in other words, if participants would have had no prior receptive or productive knowledge of the lexical targets on either experimental task), it is very likely that the decision-making task would have led to a significantly higher number of LREs and significantly more talk on task than the information gap task. This argument is based on a qualitative review of participants’ interactions on each experimental task. When participants were completing the information gap task, their interactions seemed very dry and mechanical in nature. When comparing the items present in their image (i.e., image A) with those of their partner’s (i.e., image B), participants only needed to ask if the item was present in their partner’s image (i.e., image B), and provide a description of the item if it was not present. Then, it was the partner’s responsibility to either express understanding of the item found in the participant’s image (i.e., image A), or ask more questions until they understood the item. Nevertheless, in many interactions, further clarification on the item being described was never requested. It is possible that the partners understood what item the participants were referring to, but it is also possible that they said they understood without actually understanding in order to finish the task more quickly. Another possibility is that the partner may have thought they understood, but their understanding was incorrect or represented a
mismatch. Since participants did not have to synthesize any of the lexical information presented to them on the information gap task, participants could have, in theory, lacked understanding of any of their partner’s items and still completed the task successfully.

On the decision-making task, it would have been much more difficult for participants to complete the task successfully without a minimal understanding of their partner’s items. This is because participants in each dyad had to work together to decide which items would be most useful on the deserted island, which meant that they not only exchanged information on their target items, but they also synthesized it. Therefore, under ideal circumstances, it is difficult to believe that the decision-making task would not lead to both more negotiation of meaning, and ultimately, superior recognition and production of the lexical targets over time. Of course, the results show that participants did improve significantly more in their recognition and production of the decision-making lexical target items over time; however, it is nearly impossible to determine whether this is a result of the task’s design, or because of participant’s prior receptive knowledge, or both. In this aspect, the use of concurrent data methods (i.e., think aloud protocols) could have helped immensely in interpreting these findings.

Another key limitation of this study is that the professor and the native speaker interlocutor were the same person. In other words, in spite of her native-like proficiency, the native speaker interlocutor was not truly a native speaker, although she was perceived as such by the learner participants who could not see her face or hear her voice, but were told they would be interacting via text chat with a native speaker. It is important to note that a non-academic native speaker might have interacted differently with the learner participants than a non-native speaker professor. This limitation makes it difficult to draw any conclusions beyond the context of the present study regarding the nature of interactions in the native speaker interlocutor group.
Although the learners might have thought that they were interacting with a native speaker, the responses of the “native speaker interlocutor” were not necessarily the ones that a true native speaker would have produced. Furthermore, a native speaker, particularly a native speaker not trained in foreign language pedagogy, might have had different emphases when providing learners with support during completion of the tasks.

An additional potential limitation of the study is the number of participants in the peer interlocutor group. Although all three interlocutor groups were very similar in size, it would have been ideal if the peer interlocutor group had included more participants, particularly for the analysis of LREs. Since participants worked together in dyads, and both participants in the peer dyads were learner participants, their LREs could only be counted once for analysis, which meant that a group of 30 was analyzed as a group of 15 for the first part of research questions one and two. This likely weakened the analysis.

The use of only text-based SCMC may also be interpreted as a shortcoming of this study. Many arguments have been made throughout this dissertation, particularly in terms of participants’ state anxiety, that have attributed the lack of significant different between interlocutor groups or between experimental tasks to this text-based medium. It may have been ideal to include interlocutor groups that interacted FTF in order to offer concrete evidence for this conclusion. At the same time, the argument can be made that since the world is increasingly dependent on technology, particularly now that the COVID-19 pandemic has quickly pushed education towards a more virtual environment, researchers’ energies should be singularly focused on finding the ideal conditions, tasks, interlocutors, and so on for different types of language learners in the hopes of helping them achieve the best learning outcomes. Although there has been research to show that SCMC may offer similar or slightly improved language
learning outcomes when compared with FTF interaction (see Ziegler, 2016 for a review), there is likely to be a push to move some courses, or even schools, to a virtual learning environment whether or not we are empirically and pedagogically ready for it.

Another potential shortcoming of the study was its instructions. Participants were asked to describe all of their target items in order to successfully complete the task, but they were not told that they needed to describe the item even if their partner already understood what it was. In spite of this, when participants in the native speaker and professor interlocutor groups said they understood what an item was before their interlocutor explained it, the interlocutor often asked the learner participant to provide a description of the item. This would have resulted in another LRE, potentially distorting the findings on LREs. Although there was not necessarily anything wrong with the native speaker or professor interlocutor performing these comprehension checks, the learner participants in the peer dyads should have been instructed to do the same while completing the experimental tasks.

In addition, the use of LREs alone, without considering their length, learners’ engagement with the episodes, how many of these were resolved or unresolved, or any number of other factors, may not have provided an accurate picture of the negotiation for meaning that took place during the experimental tasks. For example, Fortune and Thorp (2001) argued that “analysis based on LRE counts, although valuable, fails to capture completely the complexity of the interaction” (p. 143). Some researchers have also posited that the production of LREs alone is not sufficient for the restructuring of learners’ interlanguage – identifying correct solutions is critical (Lasito & Storch, 2013; Swain & Lapkin, 1998).

An additional limitation of this study was that, overall, participants did not seem to improve very much from the time of the pretest to the time of the delayed posttest in their
accurate production of the targeted lexical items from both tasks. The long-term gains in recognition were sizeable, with participants improving an average of 45-50% from Time 1 to Time 3 on lexical targets from either experimental task. On the other hand, the long-term gains in written production were much smaller, with participants improving an average of only 20-30% in their accurate production from Time 1 to Time 3. This means that participants were only able to accurately produce 2 or 3 out of the 12 targeted lexical items from each experimental task at the time of the delayed posttest. Thus, although the statistical analyses demonstrated that significant improvements were made over time in recognition and production on both tasks, from a pedagogical perspective, this result may not be ideal for a classroom setting. As Leow (2019) explains,

   Crucially, given the curricular aspect of ILL with the ultimate achievement of its learning goals, any pedagogical extrapolations from ISLA > ILL need to be grounded in robust learning outcomes or gain scores that would qualify for approximating a passing grade in the instructed setting and not based solely on statistical data such as p values and effect sizes (p. 489).

   A final important limitation of this study is that because the treatment only consisted of one session and the delayed posttest took place only a week after the experiment, it is difficult to make generalizable conclusions about whether or not actual L2 lexical development took place. Although participants’ performance over time on the assessment tasks can be discussed, the findings in terms of L2 development must be interpreted with caution due to the design of the study and the short treatment phase.
Future Directions

First and foremost, any researchers aiming to replicate this research would need to address the concerns stemming from the lexical target items on the experimental tasks. Either the participants for the study would need to be of a lower proficiency level, which would not be ideal since they may struggle to describe their items in the target language or to make decision with one another, or lower frequency lexical items would need to be chosen for the experimental tasks. In addition to piloting the lexical items and confirming the absence of these items from the language curricula, researchers could present learner participants with a list of 100 lexical items two to four weeks prior to the pretest and ask them to circle any of the words they recognize. Then, the researcher could use any of the words that were not circled by any participants as lexical targets for the experiment, thus avoiding any of the effects from the activation of prior knowledge that were potentially seen in this dissertation.

Although research on language learning in SCMC has been gradually moving towards addressing a wide variety of variables solely in an online environment, researchers who would like to further investigate the role of interlocutor type might consider comparing FTF and SCMC interaction. Although the lack of significant difference between the three interlocutor groups was often attributed to the SCMC mode in the present study, the reality is that there have been no studies that have addressed these interlocutor types all together in the FTF mode that the findings of this dissertation can be compared to. Therefore, attributing the lack of significant differences in anxiety, L2 performance, and so on to the SCMC environment feels a bit tenuous. Additionally, considering the growing use of video chat (e.g., Zoom, Skype) as a principal or supplemental platform for classroom learning, future research might involve comparing text-based interaction to video-based interaction.
Another potential avenue for future research would be a within-subject comparison of type of interlocutor. In other words, each participant would interact with a peer, a native speaker, and a professor. This could provide meaningful and unique insight into the way the dependent variables of production of LREs, the amount of talk on task, L2 performance over time, and anxiety interact with the independent variable of interlocutor type. In addition, the study design might include a variety of different professor and native speaker interlocutors, since different interlocutors might vary in the way they interact with learner participants even if they are trained before the experiment. Since the present study only used one person to act as both the native speaker and professor interlocutor, it is possible that these findings cannot be generalized to what it would be like to interact with any other professor or native speaker that the learner does not know.

In terms of type of task, since this is the first study of its kind to address these three interlocutor types together, a future study might consider incorporating a variety of different task designs to see if minimal differences would be observed between tasks or between interlocutor groups. Also, a future analysis of task-related data might consider whether participants only improve in the recognition and production of the target items on their own versions of the task, or if they are also improving on the target items that appear on their partner’s version of the task. This would provide further insight into the efficacy of interaction in promoting L2 lexical development.

In addition, because learners did not seem to realistically retain written productive knowledge of that many vocabulary items from Time 1 to Time 3, future research might consider a more longitudinal approach, where perhaps a larger group of lexical items are employed, but learners are exposed to them repeatedly over a longer period time. If stimulated recall were also
employed in this research design, it would have shed some light on whether type of task or activation of prior knowledge plays a more important role given participants’ repeated exposure. This would also allow researchers to discuss the findings in terms of L2 development rather than L2 performance.

There are those who might argue that presenting participants with a list of vocabulary and asking them to memorize it would prove more successful. Given the valid, but untested (at least in the present study), argument for using rote memorization to promote L2 vocabulary learning, a future replication of this study could also include two more groups: one that is provided a list of the lexical items in English and Spanish to study during length of the treatment but that does not participate in the treatment, and one that is provided with a list to study but also participates in the treatment. This would allow researchers to test whether rote memorization, rote memorization plus task-based interaction, or only task-based interaction may be the superior method for promoting L2 vocabulary learning.

Finally, future research might consider incorporating heritage language learners into the research design. Heritage language learners tend to have higher confidence in their listening and speaking skills while simultaneously reporting lower confidence in their reading and writing skills in the heritage language. At the same time, the opposite is often the case with L2 learners, who tend to have higher confidence in their reading and writing skills in the L2 when compared with their listening and speaking abilities (Hedgcock & Lefkowitz, 2011).

**Conclusion**

This dissertation employed three types of interlocutors (professor, native speaker, and peer) along with two types of tasks (decision-making and information gap) in text-based SCMC
to investigate whether 1) production of LREs on task, 2) amount of talk on task, and 3) L2 lexical recognition and production over time are related to type of interlocutor, type of task, and learners’ state anxiety.

In terms of both LREs and amount of talk on task, the findings of this study indicated that a combination of Type of interlocutor and Type of task differentially affected their quantity. While participants from all interlocutor groups produced a similar number of LREs and a similar amount of output on the information gap task, participants from the peer interlocutor group produced significantly less LREs and significantly less output on the decision-making task. This would suggest that interacting with a native speaker or professor interlocutor can lead to more negotiation of meaning and output; however, this may have been the result of the native speaker and professor interlocutor initiating comprehension checks, which often resulted in additional LREs. In addition, this finding implies that the two-way information gap task may be superior for promoting negotiation of meaning and output. Nevertheless, potential activation of prior knowledge on the decision-making task likely played a significant role in both the findings on Type of interlocutor and Type of task, so these findings should be interpreted with caution.

In terms of participants’ lexical recognition and production performance over time, Type of interlocutor mostly did not play a role. The lack of a main effect for Type of interlocutor can likely be attributed to the collaborative patterns of interaction observed in all three interlocutor groups, which in turn may have been a result of the medium of interaction: text-based SCMC. However, the exception to these findings was that the lexical recognition performance of the professor interlocutor group did not significantly decrease from the immediate posttest to the delayed posttest. This may have been a result of a combination of participants’ anxiety and their depth of processing during the experimental tasks.
With regard to the effects of Type of task on participants’ recognition and production, the results indicate that participants were significantly more accurate in their recognition and production of the lexical target items from the decision-making task. As with the findings on LREs and amount of talk on task, this finding was likely related to participants’ potential activation of prior knowledge. Overall, participants improved significantly over time in their accurate recognition and production of the targeted lexical items from both experimental tasks. Yet, although participants’ gains over time were statistically significant, from a pedagogical perspective, it should be reiterated that on average they could only accurately produce 2 or 3 out of 12 lexical items from either experimental task on the delayed posttest. In the L2 classroom, this would have translated to a failing grade.

In terms of state anxiety, few significant correlations were found between participants’ state anxiety and the production of LREs or the amount of talk on task. The lack of significant correlation between overall state anxiety and the production of LREs or amount of talk on either experimental task seems to contradict a large body of research that has established a negative relationship between FLA and L2 learning. However, the fact that a negative relationship was not detected in the present study is likely due to the medium of interaction: text-based SCMC. These findings support a small number of studies that have addressed learners’ anxiety in SCMC (Satar & Ozdener, 2008; Cote & Gaffney, 2018).

For the most part, no correlations were detected between participants’ L2 lexical recognition and production gains and their self-reported state anxiety levels. However, the correlation analyses did reveal that the more anxious participants felt while completing the information gap task, the less their recognition and production of the target items improved from the pretest to immediate posttest. By the same token, the more anxious participants felt while
completing the information gap task, the more their accurate recognition and production of the target items was retained from the immediate to delayed posttest. These findings are again attributed to potential differences in learners’ DOP, highlighting the need to incorporate stimulated recall (see Gass & Mackey, 2000) in the design of any replication study.

Finally, it was determined that, in the context of the present study, Type of interlocutor and Type of task were not predictors of participants’ overall state anxiety. This and the other findings on state anxiety emphasize that text-based SCMC can be an invaluable tool to mitigate the state anxiety of learners, for the most part regardless of interlocutor or task type.

In conclusion, this dissertation is the first empirical attempt to compare the effects of professor, native speaker, and peer interlocutors on the production of LREs, amount of talk on task, lexical recognition and production performance, and state anxiety. The findings of this study demonstrated that interaction in text-based SCMC can add value to the L2 classroom, whether in-person or at a distance, given that it can help mitigate learners’ anxiety and offers a more inclusive alternative to interacting via video chat for learners who have logistical, temporal, or financial limitations. Pedagogically, given the relatively low percentage of lexical items learned by the delayed posttest, it was recommended that the number of items be reduced to address any potential cognitive overload and the tasks performed more than once to allow multiple exposures to these items in addition to providing more opportunities to practice vocabulary in communicative activities.
APPENDIX A: PILOT DECISION-MAKING TASK

PLEASE READ CAREFULLY: Imagine you are on a sinking ship and there are twelve (12) total items on board that you and your partner can choose to take with you onto a nearby deserted island. However, you only have access to six (6) of these items, which are different from the 6 items your partner has access to on the other side of the boat. You can’t carry a lot with you while swimming to shore, so you need to narrow it down to 6 total items for the two of you to bring on shore.

Work with your partner via iMessage to decide which 6 items to bring with you to the island. If your partner mentions a vocabulary word that you are not familiar with, you must ask him/her to describe the item until you know what he/she is describing. Once you reach a decision on which items to bring, please type it in the text box below. Your final decision may not be different from your partner’s.

All discussion with your partner must be conducted in Spanish.

¡Buena suerte!

Images A

La linterna
Las pastillas
El bronceador

La cubeta
El cepillo de dientes
El encendedor
Images B

El paraguas  La comida enlatada  El sartén

El móvil  El reloj  Las tijeras
APPENDIX B: PILOT INFORMATION GAP TASK

PLEASE READ CAREFULLY: The image below includes a number of common household items, but only some of them are labeled with their equivalent in Spanish. Your partner holds a similar image that includes six (6) items that do not appear in your image. Work with your partner via iMessage to make a list of the items that appear in one person’s image but not the other’s, and vice versa. This will require you to describe the unlabeled items as best you can in your own words. By the same token, if your partner mentions a vocabulary word that you are not familiar with, you must ask him/her to describe the item until you know what he/she is describing.

Please note: When you have finished this activity, you should have a list with twelve (12) total items that both you and your partner will type into the text box below. All discussion with your partner must be conducted in Spanish.

¡Buena suerte!

Image A
APPENDIX C: PILOT OPEN-ENDED ASSESSMENT TASK

INSTRUCTIONS: Your friend, Karina, is finally graduating and she needs to move out of her campus apartment. When she calls you and asks if you can come over to help, you agree. When you get there, you find a big mess in the kitchen. Karina is feeling frustrated and can’t seem to find anything that she’s hoping to pack.

Based on the image below, try to help Karina get organized by making a list in Spanish of as many of the items you see as you can. If you don’t know the word in Spanish for an item, try using other words to describe it or try guessing. Type your answer in the text box below.

Your answer must be entirely in Spanish.

¡Buena suerte!
APPENDIX D: PILOT FILL-IN-THE-BLANK ASSESSMENT TASK

Please complete the following activities the best you can. If you do not know the answer, make your best guess.

Instructions: Read the following sentences in English and write in the blanks the Spanish equivalents for the underlined words. Do not forget the article (el, la, un, una)

E.g., the house la casa

Ronald was tired and wanted to make an easy late lunch, so he decided on a canned food ______________. Eventually, he narrowed down his options to a bowl of soup, but he couldn’t find a can opener ______________. Once Ronald found the can opener, he needed to look for a pot ______________. Finally, he found a pot in the cabinet ______________ but it was dirty. In order to save some time, he decided he would heat up his soup in the microwave ______________. That was when Ronald poured the soup into a bowl ______________ and placed it in the microwave. Ronald was afraid of burning himself, so he used an oven mitt ______________ to take the soup out of the microwave. He was also thirsty, so he took the pitcher ______________ out of the refrigerator. The refrigerator was covered by the magnets ______________ he had bought yesterday. There were also the photos ______________ of his family hanging on the fridge. Ronald liked his water extra cold so he also took the ice cube tray ______________ out of the freezer. He went to grab a glass from the cabinet when he realized that all of his glasses were in the dishwasher ______________. What a dirty kitchen! He poured his water into a mug ______________ instead.

As Ronald was about to eat, his infant daughter woke up from her nap so he heated up a bottle ______________ for her. Ronald put a bib ______________ on the baby and gave her the bottle. He realized his phone was dying, so he plugged it into the charger ______________. Ronald’s son came home from school hungry so he opened the drawer ______________ in the kitchen. He took out an apron ______________ so that he could start making dinner. The apron was very wrinkled but he couldn't do anything about it because his wife had broken the iron ______________. Ronald grabbed a bottle ______________ of their most expensive olive oil. He wanted dinner to be special for his wife. He went to turn on the gas stove, but it wasn’t lighting on its own so he used a cigarette lighter ______________. He put some olive oil in the frying pan ______________ and waited for it to heat up. At that moment, his wife came into the kitchen and asked for a cup of tea, so he turned on the kettle ______________. Ronald also wanted to make smoothies for his kids, so he took out the blender ______________. He didn’t want to use the knives ______________ his wife had bought since he had a tendency to cut himself. Therefore, he decided to use the scissors ______________ to cut up some fruit for the smoothies. When he was done cutting the fruit, he used a towel ______________ to clean off the counter. He washed off the scissors with the soap ______________ he liked to use and water.

When he turned around, he tripped and fell on a toothbrush ______________ that his daughter had left on the kitchen floor. He banged his head, so he needed the pills ______________ from the medicine chest to help with his headache. His day only got worse from there. Suddenly, a
torrential downpour began, the power went out, and he couldn’t find a flashlight. He also had to put a bucket in the bathroom because the ceiling was leaking. Ronald went outside to check on the roof and he brought an umbrella with him. When he got back inside, he checked the clock and couldn’t believe what time it was. He made a quick call to a friend on the cell phone he had bought last week, and then he went to sleep. The next morning, it was so sunny that he needed the sunscreen his wife used when he walked out the door.
APPENDIX E: PILOT LEARNER STATE ANXIETY QUESTIONNAIRE

Instructions: Please circle if you strongly agree (SA), agree (A), neither agree nor disagree (N), disagree (D), or strongly disagree (SD) with each statement.

1. I was not bothered by my partner communicating quickly.
   SA   A   N   D   SD

2. I feel more anxious in class than I did while completing this task.
   SA   A   N   D   SD

3. I got flustered when my partner communicated things I did not understand.
   SA   A   N   D   SD

4. I felt like I didn’t have enough time to think before I had to respond.
   SA   A   N   D   SD

5. I felt rushed during the task.
   SA   A   N   D   SD

6. I felt confident in my ability to quickly learn new things in Spanish.
   SA   A   N   D   SD

7. I felt tense having to communicate with my partner.
   SA   A   N   D   SD

8. This task did not make me anxious.
   SA   A   N   D   SD

9. I was relaxed and comfortable completing the task.
   SA   A   N   D   SD

10. This task made me less anxious than I feel in class.
   SA   A   N   D   SD

11. I feel like I had enough time to complete the task.
   SA   A   N   D   SD

12. This task was fun and enjoyable.
   SA   A   N   D   SD

13. It did not bother me when I did not understand everything my partner was saying.
   SA   A   N   D   SD

14. I enjoyed communicating with my partner during this task.
   SA   A   N   D   SD
15. This task was stressful for me.
   SA  A  N  D  SD

16. It did not bother me that my partner was a peer.
   SA  A  N  D  SD

17. I feel more anxious interacting with my peers in class than I did while completing this task with a peer.
   SA  A  N  D  SD

18. I would have felt more relaxed and comfortable completing this task face-to-face with my partner.
   SA  A  N  D  SD

19. I feel less anxious interacting with my peers in class than I did while completing this task with a peer.
   SA  A  N  D  SD

20. This task would have been more stressful if I had to complete it face-to-face with my partner.
   SA  A  N  D  SD

21. I would have felt more relaxed and comfortable completing this task with a professor.
   SA  A  N  D  SD

22. It bothered me that my partner was a peer.
   SA  A  N  D  SD
APPENDIX F: DECISION-MAKING TASK

PLEASE READ CAREFULLY: Imagine you are on a sinking ship and there are twelve (12) total items on board that you and your partner can choose to take with you onto a nearby deserted island. However, you only have access to six (6) of these items, which are different from the 6 items your partner has access to on the other side of the boat. You can’t carry a lot with you while swimming to shore, so you need to narrow it down to 6 total items for the two of you to bring on shore. In order to successfully complete the task, you must do the following:

1) Assume that your partner is not familiar with the vocabulary in Spanish for your 6 items. In Spanish, explain each item and its purpose until your partner understands what item you are describing.

2) Work with your partner to decide which 6 items to bring with you to the island. **Your chosen 6 items may not be different from your partner’s.** You will submit your choices as the response to this question.

An example of how to properly complete the task:

Participant A: Tengo el cepillo de dientes.
Participant B: ¿Qué es eso?
Participant A: Es un instrumento que usas para limpiar la boca. Lo usas dos veces al día, cuando te despiertas y cuando te acuestas. ¿Entiendes?
Participant B: Sí, entiendo. ¿Lo llevamos?

An example of how you may NOT complete the task:

Participant A: Tengo el cepillo de dientes, la silla, y la linterna.
Participant B: Tengo el tostador, el teléfono, y las pastillas.

**All discussion with your partner must be conducted in Spanish.**

Take your time, and ¡buena suerte!
Images A

La licuadora
El sartén
La pala

La cubeta
El tazón
El encendedor

Images B

La comida enlatada
El bronceador
Las cerillas
El colchón hinchable

La hielera

El cargador
APPENDIX G: INFORMATION GAP TASK

PLEASE READ CAREFULLY: The image below includes twelve (12) common household items which are labeled with their equivalent in Spanish. Your partner holds a similar image, but six (6) of his/her items are different than yours. In order to successfully complete the task, you must do the following:

1) Assume that your partner is not familiar with the words in Spanish for the 6 items that are unique to your image. In Spanish, explain these items and their purpose until your partner understands what household item you are describing.

2) Work with your partner to make a list of the items that appear in one person’s image but not the other’s. This will give you a list of twelve (12) total items. You will submit your list as the response to this question. By the time you submit your list, you should both know which household item each vocabulary word represents.

An example of how to properly complete the task:

Participant A: Tienes el cepillo de dientes?
Participant B: No, qué es eso?
Participant A: Es un instrumento que usas para limpiar la boca. Lo usas dos veces al día, cuando te despiertas y cuando te acuestas.
Participant B: Sí, entiendo.

An example of how you may NOT complete the task:

Participant A: Tienes el cepillo de dientes?
Participant B: No. Tienes la linterna?
Participant A: No. Tienes la silla?

**Remember, all discussion with your partner must be conducted in Spanish.**

Take your time and ¡buena suerte!
APPENDIX H: RECOGNITION ASSESSMENT TASK

Immediate Posttest - Recognition

Please complete the following task to the best of your ability. Remember that this is not for a grade - your performance will not affect you in any way.

* Required

Participant number *

Your answer

Choose the word in Spanish for the item pictured below. If you don't know, you can leave it blank and move on to the next question.

- Las cerillas
- El babero
- El bronceador
- La comida enlatada
Los cuchillos
La hieiera
La licuadora
Los imanes
La batidora
La botella
La tetera
La medida
El colchón hinchable
Los gabinetes
El cargador
El encendedor
La plancha
El jabón
El delantal
La cubeta
La mamila
El molde
El tazón
La toalla
Las tijeras
El abrelatas
La pala
El sartén
La olla
La maceta
APPENDIX I: WRITTEN PRODUCTION ASSESSMENT TASK

Immediate Posttest - Production

Please complete the following task to the best of your ability. If you don't know the answer for an item, you can leave it blank and move on to the next item.

Remember that this is not for a grade - your performance will not affect you in any way.

* Required

Participant number *

Your answer

In the blank, write the word in Spanish for the item pictured below. Please include the article (el, la, los, las).

![Scissors](image-url)

Your answer

Never submit passwords through Google Forms.
APPENDIX J: TARGET LEXICAL ITEMS AND DISTRACTORS

Decision-making targets (12):
El tazón
El encendedor
La cubeta
La licuadora
El sartén
La pala
El bronceador
Las cerillas
El colchón hinchable
La hielera
La comida enlatada
El cargador

Information gap targets (12):
La maceta
Los imanes
La medida
El babero
La plancha
La tetera
El abrelatas
La olla
El delantal
La batidora
La mamila
El molde

Information gap distractors (6):
El gabinete
La botella
El jabón
La toalla
Los cuchillos
Las tijeras
APPENDIX K: LEARNER STATE ANXIETY QUESTIONNAIRE

Instructions: Please circle if you strongly agree (SA), agree (A), neither agree nor disagree (N), disagree (D), or strongly disagree (SD) with each statement.

1. I was not bothered by my partner communicating quickly.
   
   SA   A   N   D   SD

2. It did not bother me that my partner was a peer.
   
   SA   A   N   D   SD

3. I feel more anxious in class than I did while completing this task.
   
   SA   A   N   D   SD

4. I feel more anxious interacting face-to-face with a peer than I did while completing this task with a peer.
   
   SA   A   N   D   SD

5. I got flustered when my partner communicated things I did not understand.
   
   SA   A   N   D   SD

6. I felt like I didn’t have enough time to think before I had to respond.
   
   SA   A   N   D   SD

7. I felt rushed during the task.
   
   SA   A   N   D   SD

8. I felt confident in my ability to quickly learn new things in Spanish.
   
   SA   A   N   D   SD

9. I felt tense having to communicate with my partner.
   
   SA   A   N   D   SD

10. It bothered me that my partner was a peer.
    
    SA   A   N   D   SD

11. I would have felt more relaxed and comfortable completing this task face-to-face with my partner.
    
    SA   A   N   D   SD

12. This task did not make me anxious.
    
    SA   A   N   D   SD

13. I was relaxed and comfortable completing the task.
    
    SA   A   N   D   SD
14. I would have felt more relaxed and comfortable completing this task with a professor.
   SA A N D SD

15. This task made me less anxious than I feel in class.
   SA A N D SD

16. I feel less anxious interacting face-to-face with a peer than I did while completing this task with a peer.
   SA A N D SD

17. I feel like I had enough time to complete the task.
   SA A N D SD

18. This task was fun and enjoyable.
   SA A N D SD

19. I would have felt more relaxed and comfortable completing this task with a native speaker.
   SA A N D SD

20. It did not bother me when I did not understand everything my partner was saying.
   SA A N D SD

21. I enjoyed communicating with my partner during this task.
   SA A N D SD

22. I would have felt more anxious completing this task with a professor.
   SA A N D SD

23. This task was stressful for me.
   SA A N D SD

24. This task would have been more stressful if I had to complete it face-to-face with my partner.
   SA A N D SD

25. I would have felt more anxious completing this task with a native speaker.
   SA A N D SD
REFERENCES


the cognition hypothesis of language learning and performance (pp. 287–306).


New Zealand.


Synthesizing research on language learning and teaching (pp. 133–164). Amsterdam: John Benjamins.


265


