THE SECRET IS IN THE PROCESSING: A STUDY OF LEVELS OF EXPLICIT COMPUTERIZED FEEDBACK IN HERITAGE AND L2 LEARNERS OF SPANISH

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

By

Celia Chomón Zamora, M.S.

Washington, DC May 23, 2017
ABSTRACT

The field of Instructed Second Language Acquisition (ISLA) has expressed interest in pursuing a research agenda that expands the current heritage language (HL) strand of research to investigate how this heterogeneous population re-learns their family language, and how this experience differs from that of second language (L2) learners. This dissertation examines an unexplored aspect of this strand - Depth of Processing (DOP) - and its potential role in the development of the Spanish pluperfect subjunctive in contrary to fact conditional sentences in the past, in both HL and L2 learner participant populations, and how it is facilitated by computerized feedback in more and less explicit conditions.

The current study focused on four main issues: 1) it investigated how Spanish HL participants processed input in their heritage language, and how it was similar or different than L2 participants; 2) it examined how the Speaker profile and/or 3) Type of feedback (i.e. +Explicit feedback, -Explicit feedback, control) had an effect on subsequent performance on assessments; and 4) it addressed the potential correlation between levels of processing on performances following the experiment. Levels of processing were measured by means of think-aloud protocols collected during the experiment. Performance was examined by way of two written assessment tasks - a controlled production and a semi-spontaneous picture description task.
Results showed that HL participants primarily employed non-metalinguistic strategies (e.g. grammatical judgment, prior knowledge) to process the input, whereas L2 participants relied more on metalinguistic strategies (e.g. mentions of previous lessons and explicit grammar). Additionally, although the HL participants significantly outperformed the L2 participants on both tasks at the immediate posttests, the significant difference was not maintained on the delayed posttest. Regarding the levels of explicitness of feedback, both the +EF and -EF groups significantly outperformed the control group on both tasks. At the same time, no significant difference was found between the +EF and –EF on both the immediate and delayed posttests. Finally, with respect to DOP, the +EF facilitated more instances of higher levels of processing than the –EF condition, and DOP was positively and significantly correlated with better performances at the delayed posttest.
DEDICATION

A mis padres, Juan Carlos y Celia Chomón. Nunca les podré agradecer lo suficiente. Los adoro.
ACKNOWLEDGEMENTS

First and foremost, thank you to the person who invented coffee (especially Cuban coffee), as well as Starbucks. I could confidently say I am made up of 80% caffeine after this dissertation.

To Ron, the ultimate mentor: there aren’t enough words in the dictionary to describe how much of an influence you have been to me, or my admiration for you. Thank you for embracing my “Mother Hen” tendencies, and for always providing me the opportunity to grow as an academic, a professional, a researcher, a teacher, and most importantly—a person. Thank you for always believing in me, especially when I didn’t quite believe in myself (i.e. the never-ending statistics chapter). I look forward to our upcoming exciting projects, and lunches at the “Leow Lounge!”

To Missy, the definition of “goals”: I’ve said it time and again … I would never have reached this moment were it not for you. I am so appreciative for all that you have done for me: from pushing me to apply to Georgetown, to being my academic cheerleader, to introducing me to TBLT and heritage speakers, and especially for your crucial role helping me with my HL participants; THANK YOU for being my role model! Let’s go get some empanadas and un vinito, STAT!

To the wonderful members of my committee, Cristina and Melissa: thank you for your tiresless dedication, amazing feedback, countless letters of recommendation, and words of encouragement. I am indebted to you! Thank you for making me into a strong researcher and professional, and for always reminding me to pause and consider the sociolinguistic perspective! I don’t know where I would be without your support!

To my other professors: Dr. Héctor Campos, Dr. Elena Herburger, Dr. Tom Walsh, Dr. Alfonso Morales-Front, Dr. Lourdes Ortega, Dr. John Norris, Dr. John Davis, Dr. Paul Portner, Dr. Jennifer Sclafani, and Dr. Michael Ferreira – you are the foundation of my rigorous linguistic training and academic preparedness. Thank you for instilling in me an undying passion for
linguistics and language. I am also grateful for the fellowship and guidance of Dr. Gwen Kirkpatrick, Dr. Vivaldo Santos, Dr. Alison Mackey, Dr. Peter Janssens, and Dr. Luke Plonsky, for aiding me in my uphill battle with statistics.

To all the colleagues I have had the pleasure of calling friends throughout the last five years: thank you for always making me laugh, for encouraging me to persevere, and for making this journey one I shall never forget. You know who you are! Alexandra Martín, thank you for being my ‘guapis’ and for all of your help and support in creating (and editing) my materials, and for the adorable little notes (and candy!) you left in my office. Ángela Donate, mi guapísima, thank you for being my Asturian sister. Thank you for constantly reminding me to breathe and for always being there to lend a hand when I needed it the most.

To Allison Caras, my ‘media naranja’: Your friendship is the third best thing that came out of my time at Georgetown (besides my thesis and my son, haha!). I would never have been able to complete this journey without you. You are one of my best friends, my source of inspiration, my third ear, and my academic rock. I will always be grateful for the times we spent together in and out of Georgetown (TTB! Coffee! Beer! More coffee!) trying to get through stats, data collection, coding, and just overall surviving. We did it! I can’t wait to see what’s in store for us!

To my gorgeous friends: Lauren Bitar and Rachael Leighton-Castro (my ride or dies), thank you guys for always being a source of inspiration, motivation, and laughter. Be it through our hilarious SMMS4Lyfe conversation group, to tagging each other in hilarious things, to your encouraging messages checking in, to visiting me and providing me with a much-needed break… I love you girls always and forever! My M14 moms, who kept me sane when I questioned my parenting skills while simultaneously questioning my graduate student skills. Thank you for slapping the mommy guilt out of me, and for always offering words of encouragement and support! You are all more special to me than you’ll ever know (Andrea, Brandi, Carissa, Cindy, Danielle, Dannah, Dara, Erica, Hailey, Heather, Julie, Kate, Katie K., Katie L., Kristin, Laura,
Marissa, Meg, Sammy, and Sandy)!

To my in-laws: I wholeheartedly appreciate you all for being you! Ana and Nathalie, the sisters I gained, thank you for your companionship, support, and love! To the Zamoras, thank you for your help, support, and prayers! To the Sirus, thank you for being my dissertation guinea pigs, for always thinking of me, and for cheering me on! To my gorgeous and intellectual godchildren, Johanna Sophia and Noah Isaiah: thank you for being an effervescent source of happiness in my life.

I would also like to give a special thanks to Professors Marie Guiribitey and Aurelio Baldora from F.I.U. for their patience, as well as for allowing me to go into their classrooms to recruit participants for my study. The same goes for Colleen Moorman, Maria Moreno, Chrissy Bistline-Bonilla, and Bohumira Smidakova at Georgetown! And of course, thank you to my participants! Especially the ones that came despite Hurricane Matthew!

Jose (Titoño), I will never be able to muster up enough words to express how much your patience, jokes, support, and friendship mean to me. One of the highlights of my day has always been our phone conversations in the mornings and afternoons, where we would vent, criticize, judge, and make stupid jokes. Thank you for always, always, ALWAYS believing in me, and pushing me to reach higher than I ever thought I could (despite being your annoying little sister). Thank you for always being proud of me, and for all that you’ve contributed to ensure I finish this dissertation [from the smallest (envelopes!) to the largest (letting me crash for 3 weeks to collect data)]! I love you.

Mamá y Papá, thank you for sacrificing your affluent and comfortable lives in Venezuela in order for us to have a much better life in the U.S. Thank you for not just the gift of life, but for giving me the gift of passion, perseverance, and the gift of language! Thank you for being the original advocates for our language preservation, and for instilling in me my meticulous organization skills, work ethic, and love for life. You are my inspiration for everything I accomplish. There’s no way I would have finished this dissertation without you having moved up
here to help me take care of Jude Francis. I owe everything to you. I dedicate this to you both.

¡Los adoro!

Finally, I owe everything I am and everything I have accomplished to my core: my husband and my son. Edgard, you are my best friend, my soulmate, and everything. You didn’t even blink when I told you I got accepted into Georgetown, and you have continuously been my primary source of reassurance, encouragement, love, laughter, and motivation. Thank you for ‘soportándome’ during the times that I couldn’t even stand myself, and for picking me up every time I fell. You have not only been the world’s best husband, but are also the most wonderful human being I’ve ever met. You and Chupeets are my reason for living. Jude Francis, gracias por dejar que Mami trabajara tanto; sé que lo único que querías hacer era jugar y estar en el “cuarto de Mami Papi”. Te quiero, te amo, te adoro. ¡Ahora te prometo que pasaremos todo el tiempo junto! Mami’s home!

Thank you, God, for all your blessings. Y gracias a mi Dios, a mi Virgencita, a mi Lela, Tía Vicky, y Tío Tony por interceder por mi desde el cielo durante este tiempo.
TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION ........................................................................................................1

The Statement of the Problem .................................................................................................1

Definition of Key Terms ........................................................................................................5

Awareness ...............................................................................................................................5

Cognitive Effort .......................................................................................................................5

Depth of Processing ...............................................................................................................5

Feedback .................................................................................................................................5

Heritage Language Learners ................................................................................................5

Think-aloud Protocols .............................................................................................................6

CHAPTER 2: HERITAGE LANGUAGE SPEAKERS ..................................................................7

Overview ................................................................................................................................7

Comparison to L2 Learners .......................................................................................................14

Empirical Studies Comparing Heritage Language and Second Language Learners ..........17

Focus on Instruction ...............................................................................................................23

General Summary Limitations ...............................................................................................29

Instructed Heritage Language Acquisition ............................................................................32

The Future of the Strand .........................................................................................................34

CHAPTER 3: FEEDBACK .........................................................................................................37

Feedback and Computerized Feedback .................................................................................37

Review of Empirical Studies .................................................................................................40
CHAPTER 4: DEPTH OF PROCESSING

Overview ........................................................................................................... 53
Input Processing Stage ...................................................................................... 55
Intake Processing Stage .................................................................................... 57
Empirical Studies on Depth of Processing ....................................................... 60
Awareness in Processing .................................................................................. 64
Concurrent Data-Elicitation Procedures ......................................................... 65
Research Questions .......................................................................................... 72

CHAPTER 5: RESEARCH DESIGN AND METHODOLOGY ............................... 75
Pilot Study ......................................................................................................... 75
Overview .......................................................................................................... 75
Participants ....................................................................................................... 75
Methodological Issues ..................................................................................... 76
Current Study ................................................................................................... 78

CHAPTER 6: RESULTS ....................................................................................... 108
Statistical Analyses .......................................................................................... 108
Descriptive Statistics ..................................................................................... 108
Effect Sizes ...................................................................................................... 109
Repeated Measures Analysis of Variance (ANOVA) ...................................... 109
Correlations (r) ............................................................................................... 110
Programs .......................................................................................................... 110
Reactivity .......................................................................................................... 110
Proficiency Levels ........................................................................................... 111
<table>
<thead>
<tr>
<th>Research Question 1</th>
<th>Research Questions 2 and 3</th>
<th>Descriptive Statistics</th>
<th>Speaker Profile</th>
<th>Type of Feedback</th>
<th>Main Effect for Time</th>
<th>Effect Sizes</th>
<th>Research Question 4</th>
<th>Research Question 4.1</th>
<th>Overall Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CHAPTER 7: DISCUSSION, IMPLICATIONS, FUTURE RESEARCH, AND CONCLUSION

<table>
<thead>
<tr>
<th>Research Question 1</th>
<th>Research Question 2</th>
<th>Research Question 3</th>
<th>Research Question 4</th>
<th>Implications</th>
<th>Limitations</th>
<th>Future Research</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

APPENDIX A: Language Background Questionnaire (L2 Participants)

APPENDIX B: Language Background Questionnaire (HL Participants)

APPENDIX C: Explicit Instruction PowerPoint on Imperfect Subjunctive Distractor

APPENDIX D: Controlled Production Task, All Scenarios
LIST OF FIGURES

Figure 1: Leow’s (2015) model of the learning process in instructed second language acquisition ......................................................54

Figure 2: Correct answer + explicit feedback screen .......................................................... 89

Figure 3: Incorrect answer + explicit feedback screen ...................................................... 90

Figure 4: Correct answer – explicit feedback screen ...................................................... 91

Figure 5: Incorrect answer – explicit feedback screen .................................................... 91

Figure 6: Sample of the semi-spontaneous picture description task (SSPDT) ............... 95

Figure 7: Experimental procedure for main study ......................................................... 99

Figure 8: Type of feedback over time on the controlled production task (CPT) .......... 122

Figure 9: Type of feedback over time on the SSPDT .................................................. 127
LIST OF TABLES

Table 1: Patterns of language dominance and proficiency of heritage Spanish speakers and the parental generation .................................................................9

Table 2: Survey responses to which language was used the most in specific periods in life ..........................................................................................................10

Table 3: Overview of select heritage speaker empirical studies ..........................................................30

Table 4: Assessment scores in order of greatest to least gains .........................................................45

Table 5: Summary of selected think-aloud (TA)/reactivity empirical studies .........................69

Table 6: Participant characteristics, overall and by experimental group .................................82

Table 7: Participant self-rated proficiency in Spanish and diplomas of Spanish as a foreign language (DELE) score, overall by experimental group .........................83

Table 8: Overview of experimental groups .......................................................................................92

Table 9: Operationalization of depth of processing (DOP) grammatical items ..................102

Table 10: Overall DOP scores ........................................................................................................104

Table 11: Adapted operationalization of DOP grammatical items ..............................................106

Table 12: Repeated measures analysis of variance (ANOVA) results of reactivity ........111

Table 13: Descriptive statistics for results of reactivity .................................................................111

Table 14: Descriptive statistics for results of DELE .....................................................................111

Table 15: Independent t-test results of DELE ..............................................................................112

Table 16: Think-Aloud categories for analyzing qualitative processing research question (RQ) 1 ........................................................................................................113

Table 17: Results per category used for analyzing qualitative processing RQ1 ..........115

Table 18: Descriptive statistics for controlled production task (CPT) to answer RQ2 ...116

Table 19: Descriptive statistics for semi-spontaneous picture description task (SSPDT) to answer RQ2........................................................................................................................................117
Table 20: ANOVA summary table for CPT RQ2 for within-subject .........................118
Table 21: ANOVA summary table for CPT RQ2 for between-subject ......................118
Table 22: Means on immediate and delayed posttest for speaker profile for CPT ....119
Table 23: Independent t-test and mean results for speaker profile for SSPDT ........119
Table 24: Independent t-test results for immediate and delayed posttest for speaker profile for CPT .................................................................120
Table 25: Independent t-test results for immediate and delayed posttest for speaker profile for SSPDT .................................................................120
Table 26: Means on pretest, immediate and delayed posttest for type of feedback for CPT .........................................................................................121
Table 27: ANOVA between-subject results for CPT for immediate posttest .........122
Table 28: ANOVA between-subject results for CPT for delayed posttest ..........123
Table 29: Scheffé’s results for CPT for immediate posttest .....................................123
Table 30: Scheffé’s results for CPT for delayed posttest .........................................124
Table 31: ANOVA within-subject results for CPT for +explicit feedback (EF); immediate and delayed posttest ..................................................................124
Table 32: Paired t-test results for +explicit feedback group on CPT ..................124
Table 33: ANOVA within-subject results for CPT for -EF; immediate and delayed posttest .........................................................................................125
Table 34: Paired t-test results for -explicit feedback group on CPT ..........................125
Table 35: ANOVA within-subject results for CPT for control group; immediate and delayed posttest ........................................................................126
Table 36: Mean averages for SSPDT for pretest, immediate, and delayed posttest ..127
Table 37: ANOVA between-subject results for SSPDT for immediate posttest ..........128
Table 38: ANOVA between-subject results for SSPDT for delayed posttest ..........128
Table 39: Scheffé’s results for the immediate posttest for SSPDT .....................128
| Table 40 | Scheffé’s results for the delayed posttest for SSPDT | 129  
| Table 41 | ANOVA within-subject results for SSPDT for +EF; immediate and delayed posttest | 129  
| Table 42 | Paired t-test results for +explicit feedback group on SSPDT | 130  
| Table 43 | ANOVA within-subject results for CPT for -EF; immediate and delayed posttest | 130  
| Table 44 | Paired t-test results for -explicit feedback group SSPDT | 131  
| Table 45 | ANOVA within-subject results for CPT for control group; immediate and delayed posttest | 131  
| Table 46 | Within-subject ANOVA summary table for CPT for second language learner (L2), +EF | 132  
| Table 47 | Paired t-test results for L2 +explicit feedback group for CPT | 132  
| Table 48 | Within-subject ANOVA summary table for CPT for heritage language learner (HL), +EF | 133  
| Table 49 | Paired t-test results for HL +explicit feedback group for CPT | 132  
| Table 50 | Within-subject ANOVA summary table for CPT for L2, -EF | 134  
| Table 51 | Paired t-test results for L2 -explicit feedback group for CPT | 134  
| Table 52 | Within-subject ANOVA summary table for CPT for HL, -EF | 135  
| Table 53 | Paired t-test results for HL -explicit feedback group for CPT | 135  
| Table 54 | Within-subject ANOVA summary table for CPT for L2, control | 136  
| Table 55 | Within-subject ANOVA summary table for CPT for HL, control | 136  
| Table 56 | Within-subject ANOVA summary table for SSPDT for L2, +EF | 137  
| Table 57 | Paired t-test results for L2 +explicit feedback group for SSPDT | 137  
| Table 58 | Within-subject ANOVA summary table for SSPDT for HL, +EF | 138  
| Table 59 | Paired t-test results for HL +explicit feedback group for SSPDT | 138  
| Table 60 | Within-subject ANOVA summary table for SSPDT for L2, -EF | 139  

Table 61: Paired t-test results for L2 -explicit feedback group for SSPDT ..........139
Table 62: Within-subject ANOVA summary table for SSPDT for HL, -EF ..........140
Table 63: Within-subject ANOVA summary table for SSPDT for L2, control ..........140
Table 64: Within-subject ANOVA summary table for SSPDT for HL, control ..........140
Table 65: Mean of differences and t-test results between paired samples on CPT to answer RQ3 ............................................................................................................144
Table 66: Mean of differences and t-test results between paired samples on SSPDT to answer RQ3 ............................................................................................................144
Table 67: List of correlations between CPT and immediate and delayed posttest ....146
Table 68: List of correlations between SSPDT and immediate and delayed posttest ....147
Table 69: List of correlations between CPT and immediate and delayed posttest on experimental conditions and linguistic profiles .........................................................148
Table 70: List of correlations between SSPDT and immediate and delayed posttest on experimental conditions and linguistic profiles .........................................................150
Table 71: Results per category used for analyzing qualitative processing, RQ1 for L2 participants, separated into experimental condition ............................................................157
Table 72: Results per category used for analyzing qualitative processing, RQ1 for HL participants, separated into experimental condition ............................................................157
Table 73: Analysis of participants’ scores on immediate and delayed CPT posttest .....161
Table 74: Participants’ DOP scores based on 20/20 score on delayed posttest ..........162
Table 75: Participants’ DOP scores based on 0/20 score on delayed posttest ..........165
Table 76: Results from DOP and levels of awareness coding ....................................167
Table 77: Results from DOP and levels of awareness coding of participants who scored 20/20 on the delayed posttest .................................................................171
Table 78: Overview of performance on immediate and delayed posttest, separated by type of feedback ................................................................................................................174
Table 79: Overview of performance on immediate and delayed posttest, separated by speaker profile and type of feedback .................................................................177
CHAPTER 1: INTRODUCTION

The Statement of the Problem

The last couple of decades witnessed a growth in the Heritage Language (HL) subfield of Applied Linguistics, giving way to various facets such as its own resource center (The “National Heritage Language Resource Center” (NHLRC), sponsored by the University of California, Los Angeles; “The Alliance for the Advancement of Heritage Languages”, sponsored by the Center for Applied Linguistics (CAL)), conferences (e.g., the “Symposium on Spanish as a Heritage Language”; “International Conference on Heritage/Community Languages”), professional developments and workshops, its own journal (The Heritage Language Journal, created in 2002), and countless published articles.

Paramount studies conducted in the field include research on heritage speakers’ (HS) characteristics (e.g. Benmamoun, Montrul, & Polinsky, 2013; Montrul, 2004, 2009; Montrul & Bowles, 2010; Montrul, Davidson, de la Fuente, & Foote, 2014; Montrul & Foote, 2014; Montrul, Foote, & Perpiñán, 2008; Montrul & Perpiñan, 2011; Montrul & Potowski, 2007; Rothman, 2009; Rothman & Treffers-Daller, 2014), education (e.g. Carreira, 2012; Valdés, 2000, 2005), classroom and pedagogy (e.g., Leeman, 2015), assessments (e.g., Malone, Peyton, & Kim, 2014), interaction and/or comparisons with native speakers (NS) and second language (L2) learners (e.g., Bowles, 2011; Rothman & Treffers-Daller, 2014), and their wide-range levels of proficiency (e.g. Beaudrie & Fairclough, 2012; Polinsky & Kagan, 2007).

Although Spanish in the United States is the most prolific language of choice for these studies, numerous of studies have been carried out in other HL, including
Portuguese (e.g., Flores & Barbosa, 2014), Korean, (e.g., Shin & Kim, 2000), Arabic (e.g., Albirini & Bennamoun, 2015), Russian, (e.g., Swender, Martin, Rivera-Martínez, & Kagan, 2014), Japanese (e.g., Kanno, Hasegawa, Ikeda, Ito, & Long, 2008) and Italian (e.g., Tortota, 2014).

However, despite the increased amount of research and sources available on this heterogeneous population, only a paucity of studies currently exists in analyzing Heritage Speakers (HS) from a psycholinguistic perspective, and more specifically, no study exists, to the knowledge of the author, that has investigated how HL learners process input in their heritage language (in this case, Spanish), and whether such processing differs substantially from that of native speakers and/or L2 learners. The current study aims to address these issues.

The role of feedback in Second Language Acquisition (SLA) has been comprehensively researched in meta-analyses (e.g., Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), as well as in numerous empirical studies (e.g., Ammar & Spada, 2006; Bowles, 2008; Carroll & Swain, 1993; DeKeyser, 1993; Ellis, Loewen, & Erlam, 2006; Ellis, 2007; Leeman, 2003; Loewen & Erlam, 2006; Morgan-Short, & Bowden, 2006; Nagata, 1993; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004). The consensus has been that feedback is positively related to language development. However, conflicts that have yet to be resolved lie in the specifics, such as the best type of feedback provided (more or less explicit), the proper timing of providing the feedback (immediately or after the interaction), and quantity of feedback deemed sufficient for language learning. While the majority of previous studies have contemplated the role of type of feedback primarily in the context of Face-to-Face (FTF)
interactions (e.g., Carroll & Swain, 1993; Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Leeman, 2003; Sheen, 2007), many have also investigated the role of type of feedback in Computer-Assisted Language Learning (CALL) in the form of e-tutors (e.g., Bowles, 2008; Lado, Bowden, Stafford, & Sanz, 2013; Morgan-Short & Bowden, 2006; Nagata, 1993; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004, Stafford, Bowden, & Sanz, 2012). Whereas several of them discovered an advantage for more explicit feedback (e.g., Bowles, 2008; Cerezo, 2010; Lado et al., 2013; Lin, 2009; Nagata, 1993; Rosa & Leow, 2004; Sachs, 2011), others did not find significant differences (e.g., Hsieh, 2007; Sanz & Morgan-Short, 2004). More pressing to the current study, only two studies to date have investigated the role of feedback on HL (e.g., Montrul & Bowles, 2010; Kang, 2010), and they did not focus on the role of type of feedback on HL in comparison to their L2 counterparts.

Additionally, while the majority of these studies have looked at the “product”, the “process”, that is, what is occurring in learners’ minds while performing the treatment tasks, has not been thoroughly investigated. The concept of depth of processing (DOP) was first proposed by Craik and Lockhart (1972) in the field of cognitive psychology, designating it in terms of deep versus shallow (i.e. high vs. low) processing. It has been more recently investigated in the field of SLA (e.g. Hsieh, Moreno, & Leow, 2016; Hulstijn, 2001; Laufer & Hulstijn, 2001; Leow, Hsieh, & Moreno, 2008; Morgan-Short, Heil, Botero-Moriarty, & Ebert, 2012; Qi & Lapkin, 2001). The current dissertation defines depth of processing as the amount of cognitive effort, attention, and time spent processing (or elaborating) on the target item in the input, and/or noticing facilitated by different kinds of experimental conditions and/or tasks (cf. Hsieh, Moreno, & Leow,
2016; Hulstijn, 2001; Laufer & Hulstijn, 2001; Leow, Hsieh, & Moreno, 2008). Although the few studies thus far that have empirically addressed the concept of DOP have suggested that it has a facilitative effect on several aspects of language development, several methodological limitations remain to be addressed. In addition, it has not been investigated within the field of heritage language learners, or under more or less explicit types of feedback. Additionally, Leow et al. (2008) proposed that this thread of investigation needs to utilize tasks and target items that are more conducive to DOP (instead of, for example, reading comprehension tasks, that do not necessarily facilitate deeper processing).

In summary, as demonstrated by the brief overview above (and the literature review in the proceeding chapters), there are numerous areas that merit additional research in Instructed SLA (ISLA). More specifically, the strand of heritage language learners (from a psycholinguistic perspective), and the strand of DOP taking into account more or less type of feedback in a computerized setting with different speaker profiles (i.e. HL and L2 learners).

This dissertation is organized into seven chapters. The following chapter presents the concept of Heritage Speakers (i.e. Heritage Language Learners) and discusses what has been done, what is being done, and what needs to be done. The third chapter outlines the history of feedback, and reviews empirical studies that have operationalized +/- explicit feedback. Chapter four introduces theoretical approaches to the construct of Depth of Processing (DOP) and reviews findings in previous research, as well as theoretical implications for its use. The chapter concludes with the research questions that guided the current study. Chapter five provides an overview of the design of the main
study, as well as explains what occurred in the pilot study. It further presents the details of the participants, materials, methodology, and coding scheme. The sixth chapter presents the statistical analyses employed in the study, as well as the results. The final chapter, chapter seven, completes the study by discussing the qualitative and quantitative findings, as well as emphasizing pedagogical and methodological implications of the outcomes for the field. It concludes by further addressing the study’s limitations, and future areas of research.

**Definition of Key Terms**

**Awareness:** "A particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus" (Tomlin & Villa, 1994, 193).

**Cognitive Effort:** The amount of attention that a learner pays to input (e.g., McLaughlin, 1987).

**Depth of Processing:** In this study, it is used to refer to the amount of cognitive effort and/or time spent processing the target item as revealed in concurrent think aloud protocols (cf. Hulstijn, 2001; Laufer & Hulstijn, 2001; Leow, Hsieh, & Moreno, 2008).

**Feedback:** In this study, feedback refers to the instructive information participants received concerning the answers they chose when completing a task. It can be delivered in a +/- explicit manner, based upon the information that was provided about the target form.

**Heritage Language Learners:** “A student who is raised in a home where a non-English [or the majority] language is spoken, who speaks or merely understands the heritage language, and who is to some degree bilingual in English and the heritage language”
(Valdés, 2005, 412).

Think-aloud Protocols ("think alouds", "TAs"): Audio recordings performed in an experimental setting in which a participant thinks their thoughts aloud while completing a task. Used to measure processing constructs such as awareness, depth of processing, and cognitive effort, among others. Protocols may be metacognitive, that is, provide an explanation for what was just thought aloud, or non-metacognitive.
CHAPTER 2: HERITAGE LANGUAGE SPEAKERS

Overview

Historically, the context of Spanish education in the United States has been that of teaching a foreign language, “as evidenced by the focus on instructional methods designed for monolingually-raised English speakers learning Spanish as adults” (Bowles, 2011: 32). However, much has changed since, particularly regarding the target population enrolling in Spanish courses and their needs. According to a 2015 report by the Instituto Cervantes (a prestigious government agency and non-profit organization responsible for promoting the study and teaching of the Spanish language), the United States is the world’s second largest Spanish-speaking country in the world, with approximately 41 million native Spanish speakers living in the United States, plus an additional 11.6 million simultaneous English-Spanish bilingual speakers. These bilingual speakers are normally the children of immigrants, if not immigrants themselves, who usually no longer have access to explicit Spanish instruction in a formal setting, as they did in their country of origin. As such, it is not uncommon to find these simultaneous English-Spanish bilinguals matriculating in a Spanish as a foreign language course to remedy this lack of formal Spanish instruction, along with the monolingual English speakers wanting to learn Spanish as a foreign language. The vast discrepancy in learners’ proficiency and exposure to the target language, as well as background, obviously led to pedagogical issues in the Spanish foreign language classroom. Yet this discrepancy was not just occurring between the bilinguals and the foreign language learners, but within the bilingual population as well. Depending on various factors, which will be discussed
below, these bilinguales may also fall into the category of what is called “a heritage speaker”.

Heritage language learners are bilingual individuals whose native language (in the case of this study, Spanish) is acquired in a familiar, informal setting through “interaction and naturalistic input” (Rothman, 2009: 34). For the purpose of this study, Valdés’ (2005) definition will be used, namely, “a student who is raised in a home where a non-English [or the majority] language is spoken, who speaks or merely understands the heritage language, and who is to some degree bilingual in English and the heritage language” (p. 412). The characteristics and proficiency of heritage language learners vary greatly due to a plethora of factors such as the age of arrival to and length of residence in the host country (in this instance, the USA), context of acquisition and formal learning condition, cultural background, amount and type of exposure, proficiency level, and motivation to learn and/or maintain the heritage language. As Polinsky and Kagan (2007) note, “their speaking abilities fall within a continuum, from rather fluent speakers, who can sound like competent native speakers, to those who can barely speak the home language” (p. 5). Therefore, they are a relatively heterogeneous group of learners (Bowles, 2011; Kanno et al., 2008; Kondo-Brown, 2005; Valdés, 1995).

A heritage language speaker in the United States may be a first-, second-, or third-generation immigrant to the country who speaks a language other than English. Valdés (2001) developed a typology to particularly categorize Spanish heritage learners, which corresponds to the aforementioned. The lowest proficiency group is usually English-dominant, and is typically comprised of the third- or fourth-generation immigrants. The intermediate level group is usually composed of first- or second-generation immigrants
with varying proficiency levels in both English and Spanish. Lastly, the highest proficiency level group consists of immigrants who “grew up at least partially in a Spanish-speaking country, received some schooling in Spanish, and are Spanish-dominant” (Bowles, 2011: 33). Table 1 below, adapted from Montrul (forthcoming), summarizes the possible linguistic profiles discussed above.

Table 1: Patterns of language dominance and proficiency of heritage Spanish speakers and the parental generation. (adapted from Montrul forthcoming)

<table>
<thead>
<tr>
<th>Generation (of immigrants)</th>
<th>Potential language characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>First generation (parents)</td>
<td>Dominant in Spanish</td>
</tr>
<tr>
<td>Second generation (children)</td>
<td>Dominant in English</td>
</tr>
<tr>
<td>Third generation (grandchildren)</td>
<td>Dominant in English</td>
</tr>
</tbody>
</table>

To add to the heterogeneity of this population, a heritage language speaker, regardless of their age of arrival, may have received formal education in the language for several years (for example, in after-school or weekend community programs), or may have received language input solely in familiar settings (Montrul, 2010a; Rothman, 2009; Silva-Corvalán, 1994). Furthermore, the amount of exposure to the heritage language may fluctuate based on the age of the speaker. Carreira and Kagan (2011) conducted a comprehensive national survey on university-level heritage language learners across different geographic regions and languages (22¹ in total) in order to facilitate in the design of curricula, professional development, and materials for this population. Within this survey, participants were asked, for example, which language they used the most in

¹ 23.1% of the population was specifically heritage language learners of Spanish.
particular periods of their lives. Interestingly, across all of the heritage languages, the primary period where respondents reported to have used their heritage language the most was during early childhood (0-5 years old) (70.2%), compared to adulthood (18+ years old) where they reported they used the heritage language the least (1.3%). These figures are represented in Table 2.

Table 2: Survey responses to which language was used the most in specific periods in life. (Adapted from Carreira & Kagan 2011)

<table>
<thead>
<tr>
<th>AGE</th>
<th>ENGLISH (%)</th>
<th>HERITAGE LANGUAGE (%)</th>
<th>BOTH (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5 YEARS OLD</td>
<td>11.2%</td>
<td>70.2%</td>
<td>18.7%</td>
</tr>
<tr>
<td>6-12 YEARS OLD</td>
<td>27.5%</td>
<td>18.9%</td>
<td>53.6%</td>
</tr>
<tr>
<td>13-18 YEARS OLD</td>
<td>44.0%</td>
<td>4.0%</td>
<td>51.9%</td>
</tr>
<tr>
<td>18+ YEARS OLD</td>
<td>44.4%</td>
<td>1.3%</td>
<td>54.3%</td>
</tr>
</tbody>
</table>

The language proficiency of even those educated in the heritage language, however, has been found to often dwindle as they assimilate to U.S. culture, language, and schooling. Additionally, those who only speak the language at home and are not formally educated in it fail “to reach full competence in the family language in early childhood, probably because they began to use the other language more frequently early in life” (Montrul, 2004: 128).

While currently deemed controversial, the aforementioned quote by Montrul represents ‘incomplete acquisition’, one of two major explanations originally utilized in the field as to why heritage language learners are so dissimilar to native Spanish speakers in their fluency. Language attrition or, in other words, the dwindling of competence in the language, is the other.
Montrul (2002) examined whether the age of onset of bilingualism had an effect on the incomplete acquisition of the tense and aspect morphology of the preterit and imperfect verb tenses in Spanish. Her study consisted of four groups of Spanish-speaking participants: 1) adult monolinguals (n=20); 2) early simultaneous bilinguals (early child second language (L2) learners who learned English before age 3 (n=16); 3) early bilinguals who learned English between the ages of 4-7 (n= 15); and 4) late child L2 learners who learned English after the age of 8 (n=8) where the second, third, and the some of the fourth group would be considered the typical heritage Spanish language learner. Participants underwent various offline production and comprehension tasks, including: a written morphology recognition task, an oral production task, a sentence conjunction judgment task, and a Truth Value Judgment Task. 

Results showed that the later the participants were exposed to English, the less likely they were to have difficulty discerning between the Spanish preterit and imperfect aspect. In other words, the heritage speaker groups appeared to have the most problem with the preterit/imperfect distinction in the following order: the simultaneous bilinguals, followed by the early child L2 learners, and then late child L2 learners. Overall as a group, there were no large differences between the late child L2 learners and the monolingual group. However, the simultaneous bilingual and early child L2 groups significantly differed in their results from the monolingual group. This led Montrul to suggest that the age of onset of bilingualism and the number of years of exposure to the majority language do play a role in the ultimate attainment of the heritage language (Spanish). For the few participants in the late child L2 learner group who differed in their production results from the monolinguals, it was concluded that they must have
undergone language attrition, as they were all assumed to have undergone a complete acquisition of Spanish. Montrul went on to add that overall, based on the interpretation and production tasks, it could be established that: “incomplete acquisition in the case of simultaneous bilinguals and early child L2 learners, and attrition in the case of late child L2 learners, affect more profoundly stative verbs in the preterit.” (Montrul, 2002: 57)

There were, however, some limitations to this study. First and foremost, it is difficult to argue that a speaker has an incomplete acquisition of a language, or more specifically in this study, Spanish tense and aspect, when it is almost impossible to guarantee that the participant was exposed to the same amount and type of input as the other participants. Additionally, out of the four tasks included in the study, only one (the oral production task) elicited implicit knowledge, which is what many consider to be the primary knowledge source for heritage speakers:

Written tasks may have tapped into the L2 participants' explicit (metalinguistic) knowledge, which is typical of instructed L2 acquisition. In contrast, the oral task[s] involve more automatic processing and the kind of implicit knowledge that is the result of naturalistic language acquisition, which would be the norm for heritage speakers" (Bolger & Zapata, 2011: 5)

Coincidentally, the oral production implicit task was the only task in which the overall frequency of production of both the imperfect and preterit forms was almost identical across the groups. Although there were minor discrepancies between the groups in terms of accuracy, the overall rate of accuracy per group was above 90%. Moreover, tasks 2-4 (e.g. the oral production, sentence conjugation, and truth value judgment tasks) suggested that overall the areas of most difficulty for the participants were the preterit stative verbs, followed by the achievement verbs in the imperfect tense, both of which are marked (i.e. uncommon) forms, utilized only under certain, specific conditions. Therefore, it is
difficult to claim that lacking knowledge in these marked forms is either incomplete acquisition or language attrition, as the participant may have never been exposed to them, regardless of proficiency level.

As mentioned briefly above, heritage Spanish language learners are assumed by researchers to acquire the language through a different internal learning/acquisition process than their non-native and native Spanish-speaking counterparts (Montrul, Foote, & Perpiñan; 2008). However, there is no empirical evidence to prove this. Nevertheless, it is true that, similar to native speakers, heritage speakers are first exposed to the heritage language (Spanish) at home in a familiar setting (i.e. implicitly). However, unlike the native speakers, heritage speakers seldom receive formal classroom grammatical (i.e. explicit) instruction in Spanish, and instead become immersed in the majority language (English). This causes interference from the majority language, and the grammar “ends up stabilized and incomplete, and native-like attainment of the heritage language in these speakers when they reach adulthood is not guaranteed” (Montrul et al., 2008: 127).

Due to this alleged “incomplete” attainment (e.g. Montrul, 2002, 2004, 2005, 2009, 2010b; Montrul & Bowles, 2010; Montrul & Foote, 2014; Montrul et al., 2008; Montrul & Perpiñan, 2011) of certain grammatical structures of the heritage language, some heritage speakers exhibit linguistic structures that differ from those used by native speakers of the language. These include morphosyntactic errors in gender agreement (e.g. Montrul, Foote, & Perpiñan, 2008; Polinsky, 2008), the overt (and therefore redundant) use of the subject in the pro-drop language (e.g. Montrul, 2004; Otheguy, Zentella & Livert, 2007; Polinsky, 2006), word order in interrogative constructions (Cuza & Frank, 2011); tense, aspect, and mood distinctions (e.g. Mira, 2009; Mira & Isabel, 2009;
Mikulski, 2010; Montrul, 2002, 2007, 2009; Polinsky, 2006; Silvá-Corvalán, 1994), dative and accusative case marking (e.g. Montrul, 2004; Montrul & Bowles, 2010), lexicon (e.g. Montrul, 2010a), the copulas *ser* and *estar* (e.g. Garavito & Valenzuela, 2006; Silva-Corvalán, 1994), subject pronoun expression (e.g. Montrul, 2004; Otheguy & Zentella, 2012; Silva-Corvalán, 1994) and advanced grammar, such as the subjunctive and pluperfect forms, (e.g. Montrul, 2004, 2007, 2010a, b; Silva-Corvalán, 1994, 2003). However, despite these patterns in the production of heritage Spanish speakers, they still possess various skills that non-native Spanish speakers may perhaps not acquire regardless of the number of hours of formal instruction, such as native pronunciation, cultural and pragmatic knowledge, and grammatical judgment (Valdés, 1995, 2000).

**Comparison to L2 Learners**

As stated by Montrul in the majority of her studies on heritage Spanish speakers (e.g., Montrul, 2002, 2004, 2007, 2009, 2010b, Montrul & Bowles, 2010; Montrul et al., 2008), this heterogeneous population shares linguistic characteristics with not only their native speaker counterparts because of their initial, naturalistic acquisition of Spanish, but also with L2 learners because of their varying degrees of proficiency in the target language. Montrul, Davidson, de la Fuente, and Foote (2014), for example, make the case that “many of the non-native patterns displayed by heritage speakers resemble the grammatical patterns typical of adult L2 learners who are either in the process of learning the L2 or have ceased development (and fossilized)” (p. 119). Furthermore, various studies have found that like Spanish L2 learners, heritage speakers of Spanish also make transfer errors between Spanish and English, and because of possible fluctuations in their
input and exposure, their full attainment and development of their linguistic abilities may not always be guaranteed, as opposed to native speakers (e.g. Montrul, 2002, 2010b; Polinsky, 1997; Silva-Corvalán, 2003). Montrul (2010a) further adds that “like L2 learners, heritage language learners need strong motivation to maintain and learn the heritage language, and issues of identity are very important” (p. 12).

Moreover, previous studies have shown no discrepancies between L2 and heritage speaker participant groups, in regards to language production results. In Swender, Martin, Rivera-Martinez, and Kagan’s (2014) study on exploring the Oral Proficiency Interview (OPI) from the American Council for Teaching a Foreign Language (ACTFL), data showed that the features of heritage intermediate level speakers of Spanish and Russian were similar to the L2 learners of the same level, insofar as their lack of structural control, limited vocabulary, and linguistic breakdown when producing higher level grammar. Montrul’s (2004) study on the usage of the preterit and imperfect verb tenses found that both the L2 learners and heritage language learners both had difficulty with the stative verbs in the preterit tense.

A similar pattern arose in Montrul (2005) when examining unaccusativity. 28 Spanish monolingually-raised native speakers (who served as the control group), 71 English-speaking L2 learners of Spanish, and 36 Spanish heritage speakers, all ranging from low-intermediate to advanced proficiency as per an excerpt from the Diplomas of Spanish as a Foreign Language (DELE), completed a grammaticality judgment test (GJT) in order to measure their perception of acceptability of different types of verbs. The test consisted of 55 grammatical and 55 ungrammatical sentences. Participants were asked to express degrees of acceptability for these sentences using a 1-5 scale. The 110 sentences
included 9 unaccusative verbs, 9 unergative verbs, and 10 transitive verbs, which served as distracter items. Quantitative differences between the two groups emerged after the proficiency levels were factored into the results, and this only applied to the low-proficiency groups, where the advantage was found for lower proficiency heritage speakers over L2 learners. However, overall results indicated that heritage speakers and L2 learners did not differ qualitatively in their linguistic competence, and that both groups displayed “a robust syntactic knowledge of unaccusativity in Spanish” (Montrul, 2005: 235).

Regardless of this outcome, numerous studies ascertain that L2 learners should be considered a separate entity from the profile of the heritage speaker due to numerous crucial differences, which contradict the above results. These include the heritage speakers’ lack of academic literacy and orthographic skills (Bowles, 2011; Montrul, 2010a; Montrul et al., 2008), formal explicit instruction and metalinguistic knowledge (Bowles, 2011; Montrul, 2010a; Montrul et al., 2008), as well as higher levels of grammar. These include structural errors and incorrect linguistic formulations when using the subjunctive and conditional moods (Swender et al., 2014). Conversely, when compared to traits commonly associated with L2 learners, heritage speakers are found to possess more fluency and confidence, native pronunciation, varied strong command of lexical (albeit specific in context) and syntactic structures, as well as the knowledge of the cultural implications of language usage (Rothman, 2009; Swender et al., 2014; Valdés, 1995, 2000). As Montrul (2010a) notes:

Heritage language learners tend to outperform L2 learners in oral tasks that minimize metalinguistic knowledge (and conscious memorization of rules), while L2 learners outperform heritage language
learners on written tasks that require high levels of metalinguistic awareness. (p. 17)

A multitude of empirical studies have investigated not only the comparisons between L2 learners and heritage speakers, but also the outcome of the interactions between these two groups in the same classroom. Selected studies will be discussed below.

Empirical Studies Comparing Heritage Language and Second Language Learners

Bowles (2011) analyzed interactions between Spanish heritage language (HL) and L2 learner dyads enrolled in the same Spanish university-level class. Specifically, she asked whether one type of learner (HL or L2) would engage in more language-related episodes (LRE) than the other; whether one type of learner’s LRE would be resolved more than the other (and in a more target-like manner); and whether the amount of initiation and resolution of the LRE correlated with modality of the task (oral or written). All participants \( n=18 \) filled out a detailed language background questionnaire (LBQ) and took a written Spanish proficiency test (the vocabulary and cloze sections of the DELE, as previously used in various studies), and were found to be statistically similar overall in their level of proficiency. Participants were placed in HL-L2 pairs, and were given four tasks to complete together: an oral imitation task, an oral narrative task, a timed grammaticality judgment task (GJT), an untimed GJT, and a metalinguistic knowledge test.

Approximately one-third of the LREs focused on orthography \( n=64, 31\% \), one-third of the LREs focused on vocabulary \( n=67, 32\% \), and one-third focused on grammar \( n=72, 35\% \). A little less than 1\% \( n=3 \) focused on pronunciation. Additionally, 15\% \( n=31 \) of the LREs occurred in the oral tasks, while the majority of the LREs \( n=171, 85\% \) occurred on the written tasks. Notwithstanding, Bowles reported finding various
patterns within the interactions that distinguish the types of learners, particularly in the written tasks. For one, 67% of the orthography-focused LREs were initiated by the HL learners, whereas the majority of grammar and vocabulary-focused LREs were initiated by the L2 learners. Results showed that both HL and L2 learners instigated a similar amount of LRE on both oral and written tasks, and that there was no noticeable difference in the amount of resolution (overall and target-like) of LREs initiated by both types of speakers.

These results suggest that a) as seen in previous studies, HL learners had more doubts and issues with the written modality (e.g. spelling and accent placement) than their L2 counterparts, and b) L2 learners relied on their HL partners for grammatical intuition and lexicon. Bowles concluded that due to their different strengths, when placed in the same classroom and paired for pedagogical purposes, L2 and HL learners could mutually benefit from each other.

Similarly, Bowles, Adam, and Toth (2014), one of the few HL empirical studies that incorporated concurrent data elicitation methods, utilized Language Related Episodes (LREs) from interactions of Spanish L2-L2 and L2-HL dyads in two-way informational gap tasks to see if they differed in their focus on form (e.g. quantity, focus, outcomes), self-repairs, amount of talk, number of words, and number of turns. 26 intermediate-level dyads (L2-L2, N= 13, L2-HL, N= 13) were first given a Language Background Questionnaire (LBQ), a warm-up picture narration task, performed the two-way information exchange task, and were lastly presented with a perception questionnaire.

Results revealed that the two types of dyads (L2-L2 and L2-HL) were very similar, although occurrences of focus on form were most likely to be target-like in the
L2-HL dyads rather than the L2-L2 pairs. Furthermore, there were significantly more instances of target-like language interactions in the mixed pairs than the L2-L2 dyads. Additional results showed that L2 participants utilized the target language considerably more with the HL participants than with other L2 participants, possibly suggesting that there exist varied interaction and conversational standards within the dyads, which corroborates Blake and Zyzik’s (2003) study, where HL participants were more likely to provide rather than receive linguistic assistance. Bowles et al. conclude that, “nonetheless, these results are unique in showing greater amounts of talk among L2-HL pairs, a greater initiation of LREs from L2 learners working with HL learners, and a greater perception among L2 learners that their HL partners' abilities were stronger than their own” (Bowles et al., 2014, p. 506).

In order to assess whether heritage speakers have advantages over post puberty L2 learners in the context of written comprehension and oral production, Montrul, Foote, and Perpiñan (2008) tested low- and intermediate-proficiency heritage speakers (n=69), L2 learners (n=72), and monolingual (n=22) Spanish speakers on gender agreement. In order to attempt to control for varying degree of proficiency, a section of the DELE was given prior to the experiment as a proficiency test, as well as a language background questionnaire. Participants performed three computerized, untimed tasks: 1) a written picture identification task, 2) a written gender recognition task, and 3) an oral production description-naming task. The tasks were untimed.

The overall results of the three tasks showed statistical differences between the L2 and HL participants. There were greater advantages for L2 learners in both written tasks and a greater advantage for heritage speakers in the oral task. On the first task, the data
show that both L2 learners and heritage speakers made more gender errors with the identification of feminine nouns than masculine nouns. Additionally, a significant correlation was found between the score on the proficiency tests and the score on the written picture identification task for the heritage language participants (not for the L2 learners). However, a greater number of L2 learners performed above 80% accuracy than heritage speakers. The second task, the written gender recognition task, likewise showed an overall advantage for L2 learners than heritage speakers. It also displayed that the L2 errors with gender tended to be systematic, as “feminine is more affected than masculine, and agreement with adjectives is more affected than agreement with determiners” (Montrul et al., 2008, p. 528). Moreover, proficiency scores were also positively correlated with accuracy on gender agreement in this task.

Conversely, on the oral production task, the heritage speakers were significantly more accurate on gender agreement than the L2 learners (more HL participants reached a score higher than 80% in accuracy). Also, there was a significant relationship between those HL participants that achieved a higher score in accuracy to their score on the proficiency test. Positive correlations between the proficiency scores and overall accuracy on the oral task were significant for both HL and L2 learner groups. As such, Montrul et al. (2008) interpreted the accuracy scores on the oral task by the heritage speakers as “being more representative of fast, implicit, and automatically processed knowledge (typically acquired early in childhood)” (p. 541). Similarly, they interpreted the accuracy scores on the written tasks to be a reflection of the “ability with metalinguistic, explicit knowledge (typically acquired later)” (Montrul et al., 2008: 541).
In order to reexamine the claim that heritage speakers are linguistically superior to their L2 counterparts only in phonology but not in morphology, Montrul and Perpiñan (2011) compared 60 instructed HL and 60 L2 learners of Spanish, ranging from low to advanced proficiency on their performance in four written tasks featuring tense, aspect, and mood (TAM) morphology (i.e. the imperfect tense and the subjunctive). Participants’ proficiency was measured using an excerpt of the DELE, and were then given a two elicited written morphology recognition tasks to act as a sort of pretest to assess whether the participants recognized the appropriate inflectional morphology in the required contexts. In order to test for aspect, participants were asked to read a short passage narrating a story in the past, where they were required to choose the correct form of the verb. Next, to test for mood, they wrote a one-page letter from a patient to a doctor discussing advice, recommendations, and opinions. Participants then underwent two sentence conjunction judgment tasks, which were purposely created to test receptive and interpretative knowledge of TAM morphology. To assess aspect, the task consisted of 15 minimal pairs, and participants were asked to judge whether each sentence was logical or conflicting based on a 5-point scale. A similar task was then given to assess mood. The two sessions lasted between one to one-and-a-half hours, and the order of the tasks was randomized.

Overall, L2 learners were generally more accurate than the HL learners across the proficiency levels in the two morphological tasks. However, the two sentence conjunction tasks showed that low- and intermediate-proficiency HL learners discriminated more of the aspect (i.e. preterit and imperfect tense) contrast than the L2 learners. Be that as it may, the difference diminished in participants with advanced level
proficiency. The L2 advanced learners were far more accurate in their discrimination of the mood morphology (i.e. subjunctive and indicative) compared to the advanced HL learners. Interestingly, the low- and intermediate-proficiency level HL and L2 learners were not found to differ much from each other when they did not discriminate between the subjunctive and indicative mood in the task. Montrul and Foote (2014) concluded that

The results of both tasks are consistent with developmental schedules reported in the literature on L1, L2, and bilingual acquisition of tense-aspect and mood. Moreover, contrary to [previous studies], they also suggest that early language experience also brings advantages in some, but by no means all, aspects of morphosyntax. (p. 121)

Lastly, Montrul, Davidson, de la Fuente, and Foote (2014) used spoken-word recognition tasks, gender monitoring tasks, repetition tasks, and a GJT on intermediate and advanced leaners of Spanish (both L2 and HL participants) in order to look at grammaticality effects and the effect of levels of explicitness of tasks on gender agreement and processing of canonical and non-canonical ending nouns (morphosyntax). Participants were native speakers (N= 23), HL (N= 29), and L2 (N= 33). The study consisted of: 1) proficiency tests (DELE and oral proficiency test used in previous study by O’Grady (2009); 2) Language Background Questionnaire; 3) Spanish Picture Naming Task (PNT); 4) word repetition task (participants were asked to listen to a series of determiner-adjective-noun phrases and then asked to repeat the last word of the phrase as quickly and accurately as possible), 5) Gender Monitoring Task (participants were asked to listen to a series of three-word phrases (300 total) and push a button to identify the noun as either masculine or feminine, and reaction times were measured); 6) GJT (asked to judge as grammatical/ungrammatical); 7) an eye-tracking task, and 8) an oral production task.
Results indicated that the native speaker and HL participants showed a grammaticality effect overall and with canonical nouns. However, grammaticality did not have an effect for non-canonical nouns within these two linguistic profile groups. The L2 participants showed accurate progression and learning with canonical nouns, but displayed a grammaticality effect in the opposite direction with non-canonical nouns. HL participants exhibited more native-like patterns than the L2 participants regarding grammatical knowledge and native-like processing on less explicit tasks. Additionally, native speakers were perceptive to gender violation effects in all the tasks or, in other words, they responded faster to grammatical items compared to ungrammatical items). Furthermore, both HL and L2 participants were slower and less accurate than the native speaker participants on all the tasks. Lastly, although there were no differences between the L2 and HL participants on the GMT and GJT tasks, the HL had an advantage over L2 learners in the WRT. Moreover, despite being slower and less accurate than the L2 participants, the HL showed sensitivity to the grammaticality effect and mostly perceived gender violations in all 3 tasks and with the canonical and non-canonical noun types, similar to the native speaker participants.

**Focus on Instruction**

In order to investigate whether different types of instruction affect learning outcomes in different types of learners (HL and L2), Potowski, Jegerski, and Morgan-Short (2009) randomly assigned beginner and intermediate-level Spanish HL (n=127) and L2 (n=22) learners into processing instruction (PI) or traditional instruction (TI) groups. Only some of the HL learners were also randomly assigned into a control group, which did not receive instruction, but were subjected to the same assessments nonetheless.
Within the experimental groups, participants were exposed to individual instruction on the Spanish imperfect subjunctive in adjectival clauses with nonexistent or indefinite referents. The PI treatment consisted of explicit instruction of the grammatical rule following the five guidelines for processing instruction put forth by Lee and VanPatten (2003), which are: 1) only one grammatical section was presented at a time; 2) input was presented in both modalities (written and oral); 3) all activities were contextualized; 4) the input was designed in such a way as to deter learners from using processing strategies that could negatively interfere with learning and prevent them from making critical connections; and 5) activities must require that learners respond to the input. The TI treatment focused on the same grammatical structure and contained the same number of activities as the PI treatment; however, it contained output-focused production activities modeled after those found in most heritage Spanish grammatical textbooks.

Participants were first given proficiency and pretests the day before the treatment, were then given the instructional treatment in the span of two regular class sessions, and took the posttest the day after the treatment. Learning gains were measured by three tests, given before and after the treatment: interpretation, production, and grammaticality judgment tests. Results showed that both the HL and L2 learners made significant gains on the interpretation and production posttests compared to the pretests after receiving either type of instruction. However, despite having both groups achieve significant gains, it was the L2 participant group that displayed greater net gains. On the GJT, however, only the L2 learners in either instructional group showed significant gains. Notwithstanding, it was the L2 learners overall that made greater learning gains from the pretest to the posttest on all three measures over the HL group in the same instructional
treatment, as evidenced by the Time X Group interaction and analyses. Furthermore, they also had a higher overall level of accuracy, as revealed by the posttest scores (which were higher than those of the HL group). Additionally, the control group did not demonstrate improvement on any assessment. As such, Potowski et al. (2009) concluded that the results suggest that heritage speakers’ language development may differ from that of L2 learners, although they also suggest that heritage speakers can benefit from focused grammar instruction. The finding that heritage speakers respond differently than to L2 learners to PI and TI instructional treatments draws attention to the need for future research with heritage learners to determine what specialized types of instruction may be the most beneficial for heritage learner language development, as well as to understand the cognitive processes that underlie such development. (p. 565)

They further add that a possible explanation for the difference in gains for the two types of learners may be that they were both “engaging in very different processes regarding this [linguistic] form” (Potowski et al., 2009: 563). However, because there were no concurrent procedures in place, it would be impossible to assess this claim. Another limitation in this study deals with the lack of a delayed posttest to assess long-term effects, as well as the fact that although the control group did not receive any form of instruction, they also did not receive any exposure to the target form whatsoever.

Montrul and Bowles (2010) extended their 2009 study on intermediate Spanish L2 learners’ reactivity to the instruction on the learning of the Differential Object Marking (DOM) and gustar-verbs to Spanish HL participants. Participants were 45 second-generation heritage speakers, and 12 native speakers of Spanish, to serve as the baseline. The participants were separated into two experimental groups: the uninstructed group (N= 16), which consisted of participants who completed the pretest, but not the instructional module or the posttest, and the instructed group (N=16), who were learners that only completed the instructional intervention and the posttest (not the pretest). The
pretest-posttest design consisted of a self-rating and proficiency test (DELE), the pretest (an elicited written production task and a written grammaticality judgment test (GJT)); followed by the instructional intervention one week after the pretest (ann explicit grammatical explanation of the a-personal with transitive verbs, indirect objects with ditransitive verbs, and dative experiencers with gustar-verbs. There were three practice exercises, where participants received immediate, explicit corrective feedback); followed by a 20-item practice exercise online for each of the three structures; concluding with the posttest.

Montrul and Bowles found that both the explicit instruction and the immediate, corrective feedback were beneficial to the HL participants. Results from both written tasks indicated that, overall, instruction improved the HL participants’ sensitivity to the a-marking overall, insofar as their improvement on the GJT from pre- to posttest. However, it was also noted that instruction appeared to be less effective with animate, specific DOM with the preposition a. Compared to their original study on L2 learners, the HL participants showed larger gains on both the written GJT and elicited written production task, revealing highly significant gains by HL in both intuition and production, thereby suggesting that instruction, and both positive and negative evidence, facilitates HL learning in the classroom. However, based on the experimental design (particularly that there was no delayed posttest), this can only be applied to the short-term. Montrul and Bowles also noted that since the instruction in their study “contained a variety of sources of information about the targeted structures, including positive and negative evidence, explicit rule presentation, and explicit metalinguistic feedback, it is not possible to determine the individual contributions of each to the learners' gains" (p.
As this was a longitudinal classroom empirical study, there were various limitations, and thus results should be taken with caution. Significant attrition from the initial pool of participants occurred, as many HL participants who completed the posttest did not complete the pretest. Furthermore, as it was a classroom experiment, some of the participants either did not finish the module (which was assigned as a homework assignment), or did not achieve at least an 80% accuracy score. Moreover, there were other participants who did complete both the pretest and instructional module, but did not complete the posttest. Due to this, there was no control group.

Bylund and Diaz (2012) examined the effects of HL instruction, and whether it promoted development of the heritage language proficiency. Participants HL of Spanish who were in their twelfth-grade year of high school in Sweden. Participants were separated into two instructional groups: 1) a group that received HL instruction up to the previous year (N= 26), but were no longer able to attend HL classes (and were thus only receiving instruction in the majority language, Swedish, and 2) a biweekly HL instruction group that received two hours’ worth of instruction per week at the time of the study (N= 28). Results showed the HL instructed group that was receiving HL instruction at the time of the study performed significantly better on the grammaticality judgment tasks (GJTs) and the cloze tests than those HL participants that were not enrolled in HL instruction at the time.

Bylund and Diaz attempted to control for as many variables as possible, including age of arrival (AOA) and length of residence (LOR) in Sweden, as well as the amount of Spanish and Swedish used by the participants, and the instruction received in the HL. The primary difference was whether the HL instruction took place at the time of the testing, or
the year prior. The results suggested that once heritage learners cease receiving HL instruction, the beneficial effects on their L1 proficiency that accompany it begins to dwindle. However, care must be taken with these findings, as the both authors acknowledge that the assessments utilized in the study were both metalinguistic and only in the written modality, and that perhaps, while literacy skills may dwindle after stopping HL instruction, oral and other skills may not have decreased to the same effect.

In an unpublished dissertation, Torres (2013) explored the response of HL and L2 learners exposed to a task-based pedagogical intervention. HL (N= 34) and L2 (N= 49) participants were randomly assigned to either +/- complex instruction groups, or a control group, and asked to complete a task on the subjunctive in adjectival clauses. The experimental task was computerized, and was comprised of pictures, written prompts, and written corrective feedback (sans metalinguistic information). Participants in the control group did not receive instruction. All participants were given oral and written production assessments in a pre-, post-, and delayed posttest design, which consisted of a complete and an incomplete sentence, whereby the participant was required to complete the latter using an adjectival clause either in the present subjunctive or indicative.

Torres reported that, similar to Potowski et al.’s (2009) results, the L2 participants displayed greater gain scores than the HL participants overall. Additionally, Torres noted that the two linguistic profiles may have approached the tasks in different manners, as per their responses on the exit questionnaire. L2 learners appeared to have been more focused on form, and were able to recognize that the study was looking at the grammatical distinction between the subjunctive and indicative forms, whereas the HL participants focused primarily on the content of the task, and made no mention of metalinguistic
information on the exit questionnaire. Furthermore, based on their comments, it appeared as though HL participants did not interpret the written feedback provided (in the form of recasts) as corrective. Although these findings cannot be extrapolated due to the information being non-concurrent, utilizing TAs during the experiment could provide more information as to how L2 and HL participants respond to tasks, as well as how they interpret feedback.

**General Summary Limitations**

These selected studies all show how Spanish heritage speakers are both similar and different from their Spanish L2 learner counterparts. However, there exist some limitations to these studies, which make it difficult to extrapolate the findings (see Table 3).
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Proficiency</th>
<th>Prof. Test</th>
<th>Task(s)</th>
<th>Target Structure</th>
<th>Concurrent Measures</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Montrul (2004)</td>
<td>20 NS</td>
<td>Intermediate &amp; Advanced</td>
<td>DELE</td>
<td>Oral Production Task</td>
<td>Morphosyntax; subject &amp; objects</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Montrul, Foote, &amp; Perpiñan (2008)</td>
<td>22 NS, 69 HL, 72 L2</td>
<td>Advanced</td>
<td>DELE</td>
<td>Oral Production; written comprehension; written recognition</td>
<td>Morphosyntax- gender agreement</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Potowski, Jegerski, &amp; Morgan-Short (2009)</td>
<td>127 HL, 22 L2</td>
<td>Beginner &amp; Intermediate</td>
<td>Class placement (None)</td>
<td>Interpretation; production; &amp; GJT</td>
<td>Past subjunctive w/ indefinite referents w/ different instruction</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Montrul &amp; Bowles (2010)</td>
<td>12 NS, 45 HL</td>
<td>Intermediate &amp; Advanced</td>
<td>DELE</td>
<td>Oral picture naming; written recognition; written comprehension</td>
<td>DOM on gustar verbs w/ explicit instruction</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bowles (2011)</td>
<td>9 HL, 9 L2 (dyads of HL/L2)</td>
<td>Intermediate</td>
<td>DELE</td>
<td>Oral imitation; oral narrative; timed &amp; untimed GJT; metalinguistic knowledge test</td>
<td>Various (orthography, grammar, lexicon); oral interaction</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Montrul &amp; Perpiñan (2011)</td>
<td>60 HL, 60 L2</td>
<td>Low to advanced</td>
<td>DELE</td>
<td>2 elicited written morphology recognition &amp; 2 sentence conjunction judgment tasks</td>
<td>Tense, aspect, mood (TAM)- morphology</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Bylund &amp; Diaz (2012)</td>
<td>54 HL</td>
<td>Not mentioned</td>
<td>Cloze &amp; GJT</td>
<td>Cloze test (written); GJT (written)</td>
<td>Looked at HL instruction</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Torres (2013)</td>
<td>34 HL, 49 L2</td>
<td>Not mentioned</td>
<td>DELE</td>
<td>Oral Production Task; Written Production Task</td>
<td>Present subjunctive in adjectival clauses</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Bowles, Adam, &amp; Toth (2014)</td>
<td>13 HL, 39 L2</td>
<td>Intermediate</td>
<td>Not mentioned</td>
<td>Language Related Episodes (LRE) in dyads through 2-way informational gap task</td>
<td>Focus on Form through interaction</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Montul, Davidson, de la Fuente, &amp; Foote (2014)</td>
<td>23 NS, 29 HL, 33 L2</td>
<td>Intermediate &amp; Advanced</td>
<td>DELE</td>
<td>Picture naming; word repetition; gender monitoring; GJT; oral production</td>
<td>Morphosyntax- gender agreement</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Proficiency Measures- One common instrument used to assess the proficiency of HS and L2 learners in these studies has been an excerpt of the DELE (i.e. CLOZE and a multiple-choice section) adapted by Montrul. While it is important to have a reliable instrument that could be compared across the board with other studies in the same strand, it begets the question as to whether these two sections truly encompass a true proficiency assessment that is fair for both profiles of learners, particularly in its modality. Since it is exclusively written (rather than oral), with which heritage learners are assumed to be at a disadvantage.

Lack of Delayed Posttest- By including a delayed posttest in the experimental design (when it is applicable), it would allow the researcher to make better claims as to the retention and perhaps ‘learning’ of the target item, especially if the posttest is given the day after the treatment (e.g. Potowski et al., 2009). Additionally, the longer the delay, the more informative the results (see also Norris & Ortega, 2000).

Absence of Concurrent Measures/Manipulation Checks- It is difficult, if not impossible to make claims as to what the participants were doing/ selecting/ thinking/ processing without some kind of concurrent measure. These measures (e.g. Think Aloud (TA) protocols, Eye Tracking (ET)) not only enable the researcher to assess whether learners paid attention to certain forms, followed the experiment directions accurately, were on task throughout the entire session, and allow for qualitative analyses, but also provide more substantial evidence of what is being measured (when applicable), which all raise the internal validity of the study (Leow, 2000). Aside from Bowles’s (2011) communicative task interaction study that analyzed LREs, currently, there exist very
limited studies, to the knowledge of the author, in the HS strand that utilize concurrent measures. If concurrent measures are used, it is not utilized to elicit type of processing.

Interpretation of Results- All things considered, one important factor to consider when interpreting the results of these assessments is best summarized by Valdés (2005):

What will not be immediately clear from superficial assessments is whether flawed production is due to interrupted acquisition, individual language attrition, or "full" acquisition of a contact variety of the heritage language that is now quite different from the varieties of the heritage language originally brought to the community. (p. 417)

Additionally, a group of select heritage language researchers have taken up the cause to argue that consistently comparing heritage speakers to L2 learner and native speakers, to then classify HS as a different entity (i.e. not native speakers) is a grave mistake (Rothman & Treffers-Daller, 2014). As such, care should be taken when interpreting results to not necessarily classify heritage speakers as nonnatives of their heritage language, and when utilizing native speaker participants as a control for the study, to refer to them as monolingual speakers of language X (Rothman & Treffers-Daller, 2014). Rothamn and Treffers-Daller continue by adding

There is no question that heritage speakers often display significant differences from monolinguals in competence and/or performance and that HS acquisition across individuals is characterized by levels of variation aren’t normally seen in monolinguals. This begets the question of whether one can be considered native if one fails to display the outcomes expected of monolingualism. Many monolinguals are not speakers of the standard variety, displaying difference from that baseline as well. No one would dare label non-standard variety monolingual speakers as anything less than natives of at least a vernacular variety of language X (p. 94-95).

**Instructed Heritage Language Acquisition**

Although the push for separate classes for heritage speakers is not a new one, it is nonetheless still a relevant issue. The last decade has witnessed an exponential growth in
the engendering of new syllabi, courses, materials, and textbooks specifically tailored for this heterogeneous population. From this, the field of instructed heritage language acquisition (IHLA) was created, aimed at gathering systematic research on the previously mentioned facets of heritage language education. The aforementioned empirical studies and varied strengths and weaknesses in linguistic competence are why researchers advocate for separate Spanish classes for L2 and heritage language learners (e.g., Bowles, 2011; Colombi & Alarcón, 1997; Potowski, 2002; Valdés, Fishman, Chávez, & Pérez, 2006). This is mainly because these researchers have argued that curricula designed for L2 learners are inappropriate for heritage language learners (Carreira & Kagan, 2011; Peyton, Carreira, Wang, & Wiley, 2008). Furthermore, others have suggested that when individual classes are made available for HL learners of Spanish, separate objectives are normally created; frequently concentrating on literacy skills and explicit grammar (e.g., Colombi & Alarcón, 1997; Valdés et al., 2006).

According to Montrul (2011), for instance, placing intermediate- or high-proficiency level heritage speakers in introductory or low-level Spanish classes with non-native novice speakers puts them at a disservice, since they already possess native-like phonetic ability and lexicon. Similarly, misplacing lower-level heritage learners in an advanced Spanish class may lead to a decrease in motivation due to the course material being too difficult (Potowski et al., 2009). However, although universities across the country are creating and offering more courses geared for these learners, the reality of the situation is that most institutions do not offer specially-tailored classes for heritage speakers, and instead opt for mixed classrooms of HL and L2 learners (e.g., Bowles, 2011; Lynch, 2008). Although this may be true, mixed classrooms, such as those that
have been reported in studies such as Bowles (2011), could be beneficial to both learners if care is taken to provide adequate tasks and instruction to assist them both. However, if there are substantial differences in the levels of proficiency between both groups, and/or, as Montrul and Bowles (forthcoming) note, “when tasks are not carefully structured to maximize the learning potential for each, it appears that learning opportunities may be unbalanced, and often in favor of L2 learners” (p. 18)

**The Future of the Strand**

A paucity of studies, if any, addresses the topic of heritage speakers within a psycholinguistic framework. Much instructed second language acquisition (ISLA) research has been accomplished without “attention to the fact that learners are not homogenous in their learning styles or language learning abilities, but rather differ in terms of numerous cognitive measures” (Montrul & Bowles, forthcoming: 17). As such, the same line of thinking should be brought to IHLA.

Although various empirical studies exist describing the characteristics of heritage learners, and their profiles in comparison to their L2 learner equivalent, as has been previously discussed, further research needs to be done to not just analyze their productions, but how they produce. Several researchers have called for this, such as Montrul (2010), who said, “we need more psycholinguistically-oriented studies of adult HS to find out how they process input in the HL and in different skills" (p. 19). As such, the future of the heritage speaker strand needs to address empirically HS’s processing and processes via the use of concurrent data elicitation measures.

Traditionally, the studies that have investigated heritage speaker grammar have relied on offline, non-concurrent measures, such as grammaticality judgment tests.
However, these kinds of tests do not necessarily encompass much information about how language is being processed in ‘real time’, which, as Bolger & Zapata (2011) state, “is how language functions must normally be carried out” (p. 2). Furthermore, studies such as Potowski et al. (2009), which were reviewed previously, made a note that if they had included concurrent measures, they would have been able to analyze how some of their participants had processed the target structure, which would have allowed them to provide another possible explanation for their results.

Investigating how heritage language speakers process the input, such as a target grammatical structure, or different types of feedback, hold important pedagogical implications. Instead of focusing solely on attempting to create separate classes and materials for these learners, one would be able to also emphasize forms in which to adapt and utilize differentiated instruction to aid heritage learners in processing specific structures or information when these particular courses for heritage learners are unavailable. This would be possible by knowing how these learners process information, to what particular information they pay attention (and how), and so forth. Another serious limitation currently in the research field is that of how heritage speakers’ processing compares to that of an L2 learner. Are the similar? Different? If so, how, and how would these differences or similarities affect the classroom setting (in other words, dyads such as the ones seen in Bowles (2011))?

To summarize, while there is a saturation of studies that offer insight as to the interactions and the grammar of this group of speakers, what is still so far under-researched is how this heterogeneous population processes the input in the first place. By studying how HL process linguistic information, and comparing it to how L2 and
monolinguals process the same information, it could provide researchers and teachers with important pedagogical information as to the best practices to teach to such mixed classes.
CHAPTER 3: FEEDBACK

Feedback and Computerized Feedback

Feedback, a construct originating from the cognitive psychology field, is defined as “[a]ny indication to the learners that their use of the target language is incorrect” (Lightbown & Spada, 1999: 171). Although initially understood to be either positive or negative (i.e. positive reinforcement as opposed to highlighting an error), the majority of SLA research has focused on negative evidence, as can be seen in the aforementioned definition. The role of negative feedback (more commonly referred to in the field as ‘error correction’ or ‘corrective feedback’) has been researched extensively in the SLA field, and various meta-analyses have resulted in an overall consensus on its efficacy for promoting L2 learning (e.g. Mackey & Abbuhl, 2005; Mackey & Goo, 2007). These studies, particularly in the field of oral interaction (e.g. Carroll & Swain, 1993; Mackey, Oliver, & Leeman, 2003; Mackey and Philp, 1998; McDonough & Mackey, 2006), have provided substantial support for the notion that L2 learners benefit from these “explicit or implicit means of drawing their attention to linguistic form” (Sachs & Polio, 2007: 68).

Despite this general agreement regarding the benefits of feedback, a debate regarding which type of feedback is superior has ensued. Some of these types of feedback include the level of explicitness of negative evidence (+/- explicit, or implicit vs. explicit), type of evidence (positive or negative), timing of feedback (preemptively or reactively), and the modality of the feedback (e.g. computer-mediated communication (CMC) or face-to-face (FTF); oral or written). Of particular relevance to this study is the construct of +/- explicit feedback.
Previous studies have stated that feedback could be provided either explicitly, where it is overtly stated that there is an error (i.e. giving the learner specific corrections relevant to an error) or implicitly, where the learner has to infer the existence of an error (i.e. through indirect discourse strategies, such as clarification requests or recasts) (e.g. Carroll & Swain, 1993; Mackey & Abbuhl, 2005). However, rather than creating a dichotomous label for these terms, authors such as Carroll (2001) have discussed degrees of explicitness, and have claimed that these degrees can be determined based on how it aids the learner in 1) detecting the purpose of the feedback; 2) spotting the location of the error; and 3) recognizing the nature of the error (i.e. lexical, syntactic, phonetic, etc.).

Most of the research that has investigated the degrees of explicitness of feedback has been conducted in the traditional FTF interactions (e.g. Carroll & Swain, 1993; Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Leeman, 2003; Loewen & Nabei, 2007; Lyster & Mori, 2006; Lyster & Ranta, 1997; Sheen, 2007), and in terms of meta-analyses (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), only two (Li, 2010; Mackey & Goo, 2007) included computerized feedback. Out of these four meta-analyses, none were able to conclusively speak to the effectiveness of the degree of explicitness of feedback. Li (2010), in her meta-analyses, categorized types of feedback as either implicit (e.g., repetitions, clarification requests, elicitations, and generally any feedback without intent to explicitly draw the learner’s attention to the error) or explicit (e.g., any metalinguistic feedback, explicit correction of an error, and generally any feedback with the explicit intent to draw the learner’s attention to the error). Results showed that there was a medium overall effect for corrective feedback, and that the effect was sustained over time. Nonetheless, the effect of the implicit feedback was sustained
better than the effect of the explicit feedback. Additionally, Li looked at the mode of delivery (i.e. FTF or computerized feedback) as a variable, and was the only meta-analysis to do so. Her results did not indicate any difference between the two modalities in affecting L2 development. Lyster and Saito (2010) investigated types of feedback such as recasts, prompts (e.g. clarification requests, repetition of error, metalinguistic clues, elicitation), and explicit correction in a continuum (versus a dichotomy), where recasts were considered more implicit, and explicit correction was most explicit. Their results revealed larger effects for prompts over recasts, in other words, a larger effect for feedback closer to the implicit continuum.

Although Mackey and Goo (2007) did include computerized feedback as part of their meta-analyses, the number of studies that were included in the analysis (as per their inclusion criteria) were only two (cf. Ayoun, 2001; Sagarra, 2007). Within the computerized vs. FTF mode of feedback, they did not make any claims, as they felt that there needed to be “a greater theoretical specificity or practical innovations in making claims about the superiority of one feedback over another” (Mackey & Goo, 2007, p. 440). Lastly, Russell and Spada (2006) examined the effects of corrective feedback (CF) on L2 development. Out of 56 studies that examined CF’s effect on L2 learning, only 15 were included in the study (as per their inclusion criteria). Although their meta-analysis did find support for the benefits of CF in general, they did not draw any definitive conclusions, as there were a paucity of studies included.

A number of computer-based empirical studies have surfaced recently, examining the role of the different types of feedback, and particularly regarding their degree of explicitness (e.g. Loewen & Erlam, 2006; Yilmaz, 2012). Similarly, other studies suggest
that explicit negative feedback seems to have an observable effect on classrooms (e.g. Lightbown & Spada, 1990) and on individuals (e.g. DeKeyser, 1993). While various of these empirical studies investigating computerized feedback have yielded positive results for groups exposed to explicit feedback over implicit (e.g. Bowles, 2005; Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004), others found no difference between implicit or explicit feedback groups from pretests to posttests (e.g. Sanz, 2004; Sanz & Morgan-Short, 2004). A comprehensive look at these studies follows.

**Review of Empirical Studies**

Nagata (1993) and Nagata and Swisher (1995) examined whether traditional feedback without metalinguistic information (TCALI) or intelligent feedback with metalinguistic information (ICALI) was more effective on the learning of three types of Japanese passive constructions. Second-year undergraduate Japanese students ($n=32$) were randomly assigned to one of these two conditions, where feedback in the TCALI group indicated what was wrong with their answer, and feedback in the ICALI group consisted of the former, as well as a metalinguistic explanation. Participants read grammatical explanations about the linguistic target forms, and then completed 90-item production exercises of Japanese passive sentences. Performance was measured by an achievement test and a retention-delayed test. Results showed that participants in the ICALI group performed significantly better than the TCALI group on both delayed posttests, particularly regarding complex sentence-level errors (particle errors). However, concerning word-level errors (vocabulary and verb conjugations), there were no significant differences between both groups.

The authors concluded that metalinguistic feedback (ICALI) is more effective
than traditional feedback, particularly when the grammatical system is complex. These studies, however, contained a few limitations. First, of course, was the small sample size, which makes it difficult to extrapolate the findings. Furthermore, as far as the studies’ internal validity (cf. Leow, 1999), it was not explicitly stated as to how the participants were randomized into one of two conditions (if at all). Additionally, there was no mention of manipulation checks during the treatment, nor was there a mention of a true experimental control group (i.e. a control group that was exposed to the target items but did not receive any feedback). Lastly, it was not mentioned if in between the posttest and the delayed posttest, external influences or exposure to the linguistic target forms were controlled.

100 5th semester L2 advanced Spanish college students completed a multiple-choice jigsaw problem-solving task in Rosa and Leow’s (2004) study in order to investigate the effects of different task conditions on participants’ recognition and production of the Spanish contrary-to-fact past conditional constructions. The study compared six conditions that varied in their degree of explicitness: 1) EPEFE (explicit pretask, +explicit feedback); 2) EPIFE (explicit pretask, +implicit feedback; 3) EFE (explicit feedback); 4) EP (explicit pre-task); 5) IFE (implicit feedback); and 6) CO (control). The computerized jigsaw puzzle contained 18 critical items and 10 distractors. Participants had to select one out of four options that best presented the contrary-to-fact meaning in the past in the subordinate clause of conditional sentences. Task-essentialness was part of every group condition, whereas the type of feedback changed based on the experimental group (i.e. the explicit feedback group received a metalinguistic explanation as to why the answer provided was correct or incorrect; the implicit feedback group was
only informed of the correctness of the response). Learning was measured by recognition
and production of old and new items in the posttest (which took place immediately after
the experiment) and delayed posttest (which took place 3 weeks after the treatment).

The results were mixed. The data revealed that the explicit feedback group (EFE)
significantly outperformed the implicit feedback group (IFE) in the recognition of new
items and the production of old items on the immediate and delayed posttest. However,
no difference was found between groups in any of the assessments over time for the
recognition of old items or the production of new items. Furthermore, as all the
experimental groups outperformed the control group, which was the only condition
without task-essentialness, it could be argued that task-essentialness had a differential
effect on the recognition and production of old and new target items. However, feedback
was not teased out from task-essentialness. An important limitation that affects the
results of this study, as noted by Sanz and Morgan-Short (2004), is that the implicit
feedback condition only received exposure to the linguistic target form in the 18 items of
the task, which may have not been enough to adequately grasp the structure; particularly
in such a learning condition (Ellis, 1993; 2006). Moreover, all of the experimental
groups, in one manner or another (i.e. in the form of explicit/implicit feedback or through
the pre-task), received some kind of input with the target structure sans the control group,
which could help explain why all of the experimental groups outperformed the control
group. Therefore, it is imperative to take note of this when attempting to separate the
contributions of each variable.

attempted to address this issue, as well as investigate whether exposure to structured-
input (oral and written) has a differential effect on L2 learning of the Spanish preverbal direct object pronouns, based on the presence or absence of metalinguistic information, explicit feedback, and task-essentialness. The participants (n=69) were randomly separated into one of the following groups: a) grammar explanation and explicit feedback [+E, +F]; b) grammar explanation and no explicit feedback [+E, -F]; c) no grammar explanation and explicit feedback [-E, +F]; and d) and no grammar explanation and no explicit feedback [-E, -F]. Participants were exposed to different activities (oral and written) with 56 practice items (compared to Rosa and Leow’s (2004) 28 items), and the assessments included interpretation and production tests.

Results suggested that the amount of interaction with the target item (i.e. in the task-essential practice) was enough to promote language learning, as the authors found that all of the groups improved significantly; thereby showing that metalinguistic information (i.e. explicit feedback) was not necessarily critical for learning. However, this claim needs to be taken cautiously, as even the groups that did not receive explicit feedback [-F] still received some form of implicit feedback, since they were informed whether their answers were correct or incorrect. It is a possibility that participants became metalinguistically aware of a pattern, and formed a hypothesis about the underlying rule of the target structure. As there were no concurrent measures (e.g. think alouds), such claims are unable to be proven. Especially relevant to this dissertation was the aforementioned, as well as that the retention of any learning that took place was not measured (i.e. no delayed posttests).

In order to investigate the claims presented by both Rosa and Leow (2004) and Sanz and Morgan-Short (2004), Cerezo (2016) sought to determine whether the
differences between the amount and type of input-based practice provide an explanation of the conflicting results in these two studies, in order to mediate the effectiveness of computerized right/wrong feedback. 52 intermediate learners of Spanish were placed in one of four experimental groups that differed in a) the amount of input-based practice (completing 28 or 56 items in a task) and b) the type of input-based practice (choosing between 2 or 4 options for each question). The study replicated Rosa and Leow’s (2004) task, in which participants would need to discern between the imperfect and the pluperfect subjunctive in contrary-to-fact conditional sentences in a computerized task.

The 28-item group consisted of 18 target items and 10 distractor items, while the 56-item group contained twice the number of items (36 target and 20 distractor items). In the groups with two multiple choice options, participants were required to choose between: a) the target structure, the pluperfect subjunctive (e.g. “no hubiera dicho” [“hadn’t said”]) or b) the non-target structure, the imperfect subjunctive (e.g. “no dijera” [“didn’t say”]), whereas in the four-option multiple choice group, two additional distractor options were added: c) the Spanish simple conditional (e.g. “no diría” [“wouldn’t say”]), and d) the Spanish perfect conditional (e.g. “no habría dicho” [“wouldn't have said”]). All of the experimental groups received concurrent feedback consisting of either a correct (“Cool!”) or incorrect (“Oops! Try again!”) response. The pre-, post-, and delayed assessment tasks were comprised of a 34-item multiple-choice recognition task and a 32-item controlled production task, each with 20 critical items. The recognition task had the same structure as the experimental task, while the production task required participants to conjugate the verb (provided in the infinitive tense) in the correct manner in the blank space provided, hence not offering options for
the correct answer. The tests included the same items, in randomized order. All experimental groups were asked to think aloud during the experiment.

Based on the results from 2x2 repeated-measures ANOVAs, as well as standardized mean difference effect sizes, the different amount of options (2 or 4) in the computerized tasks of Rosa and Leow (2004) and Sanz and Morgan-Short (2004) did not appear to have a differential effect on the improvement of scores in the target form. Cerezo (2016) thus concluded that this study did not provide empirical evidence to support Sanz and Morgan-Short’s (2004) previous assertion that the 4-option task in Rosa and Leow’s (2004) study may have hindered participants “from noticing the target form” (Cerezo, 2016, p. 116). Furthermore, the non-significant interactions between time and amount of practice in both the recognition and production assessments showed that both groups (28- and 56-items) experienced comparable and significant learning of the target item. However, it is interesting to note that in both the production and recognition assessments, the 56-item groups (both the 2- and 4-option) produced the greatest gains (see Table 4 below).

<table>
<thead>
<tr>
<th>Number of Options X Number of Items</th>
<th>Production Assessment Score</th>
<th>Recognition Assessment Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 X 56</td>
<td>14.15</td>
<td>14.69</td>
</tr>
<tr>
<td>4 X 56</td>
<td>11.92</td>
<td>14.54</td>
</tr>
<tr>
<td>4 X 28</td>
<td>10</td>
<td>13.93</td>
</tr>
<tr>
<td>2 X 28</td>
<td>9.69</td>
<td>12.69</td>
</tr>
</tbody>
</table>

Consequently, although there was no significant interaction between time and the
amount of input-based practice, the greater score gains of the 56-item groups, and the small and medium effect sizes for the 2-option groups “suggested that the higher amount of practice items completed by the right/wrong feedback group in Sanz and Morgan-Short (2004) vis-à-vis Rosa and Leow (2004) might have added some edge in the development of the targeted form, particularly in production” (Cerezo, 2016, p. 115). However, it is interesting to note that based on information acquired in informal post-debriefing interviews, it could be postulated that the desired and expected effects of the extra practice items was not reached because it produced fatigue and/or boredom in some of the participants. “Indeed, the think-alouds in the 56-item groups provided many instances of participants who complained about the large number of items, both during the treatment and posttests” (Cerezo, 2016, p. 116).

Although Cerezo did not find a significant interaction between time and amount of input-based practice, this could be due to his relatively low cell size per experimental group, which could explain as to why the 56-item group achieved greater gain scores without achieving significance. Additionally, it should be noted that the target items in Cerezo’s study were similar to those in Rosa and Leow’s (2004), but dissimilar from those in Sanz and Morgan-Short (2004); in order to truly compare these two studies, both target items should be included, as well as the assessments utilized by both studies, in order to truly compare the results. Lastly, as the assessments were controlled production and recognition tests in multiple-choice format, it is difficult to control for participants guessing. This limitation is remedied in this dissertation.

Stafford, Bowden, and Sanz (2012) investigated the effects of pre-practice grammatical explanation and explicit corrective feedback while learning the Latin
assignment of thematic functions through an interactive computer program by Spanish-English students. Participants \( n=65 \) were placed into 4 experimental groups that varied in the degree of explicitness \(+/-\ EF\) of the instructional treatment, and varied in the provision of the grammatical explanation prior to the treatment \(+/-\ GE\). The four groups were as follows: 1) \([+\ GE, +\ EF]\); 2) \([-\ GE, +\ EF]\); 3) \([+\ GE, -\ EF]\); and 4) \([-\ GE, -\ EF]\).

Participants first underwent a Latin vocabulary lesson and quiz to ensure that they were naïve learners of Latin, as well as were given vocabulary reviews (of Latin words) and pretests. A week later, participants completed the experimental treatment. When provided (for the two \(+\ GE\) groups), participants completed a self-paced computer lesson, which provided explicit metalinguistic information in English of the target structure, and other relevant Latin grammar. Furthermore, onscreen feedback was provided for the interactive examples.

All participants were provided with task-essential practice, which consisted of six tasks (3 oral, 3 written). Two of the tasks were picture-matching tasks, where participants were asked to match the photograph to the Latin sentence they had just read or heard. Two other tasks had participants read or listen to a Latin sentence and then interpret it by choosing the correct English translation (from two options). The other two tasks required participants to read or listen to a sentence in Latin and choose which sentence best described the photograph on the screen. Feedback was provided to the all of the participants, and varied in its level of explicitness based on the group. If the experimental group was \([-\ EF]\), they were told whether or not the answer was correct. The \([+\ EF]\) included the aforementioned, as well as item-specific metalinguistic information.

Immediately following the treatment, four Latin language tests were given in
order to assess “initial language learning”. Two interpretation tests were modeled after
the interpretation tests in the treatment. A grammaticality judgment test (GJT) was
included, as well as a production test. 3 weeks after the last session, a delayed posttest
was given to assess retained knowledge. Results revealed that pre-practice grammar did
not appear to have any clear benefit for language learning apart from what the
participants received through the practice in the treatment and the feedback. Additionally,
although the [+GE] groups performed significantly better in the immediate posttests in
the areas of GJT, interpretation (specifically noun case morphology) and written
production (especially verb markings and noun case), it was not retained over the three-
week interval between the immediate and delayed posttests, except when both explicit
feedback and explicit grammatical instruction were coupled. In other words, [+GE] was
only effective with [+EF]. The results between the [-GE, -EF] and [-GE, +EF] were
particularly relevant for this dissertation. The GJT, for instance, showed a better
performance for the [-GE, -EF] group, whereas the written production test showed a
better performance for the [-GE, +EF] group. Stafford et al. (2012) went on to add that
only the [-GE, -EF] group exhibited continued improvement on the GJT, even on the
delayed posttest, which would be attributed to differing types of instruction leading to
qualitatively different manners of input processing. Therefore, it is difficult to tease out a
clear conclusion in regards to the effectiveness of +/- explicit feedback.

A further limitation which affects this dissertation involves the outcome of the
above result of the [-GE, -EF] group. Were there concurrent measures (such as TAs)
throughout the treatment, the researchers would have been able to assess why in
particular this group performed better long-term (Hsieh, Moreno, & Leow, 2016).
Finally, and more relevant to the current proposed study, it was reported that within the participant pool, there were some heritage speakers of Spanish who were highly proficient in English. However, proficiency in Spanish was only self-assessed by the speakers themselves. From what is known about the HS population, it may very well be that these speakers possess a high proficiency in English but not necessarily in Spanish.

While numerous studies exist investigating the effects of +/- explicit feedback on foreign language learners (L2, L3, and so forth), there is a lack of studies conducted on its effect on heritage speakers. At the time of this dissertation, only two major empirical studies have been conducted looking at explicit feedback within this heterogeneous population. Montrul and Bowles (2010) examined the effects of instruction on the acquisition of the differential object marking (DOM) a on gustar verbs in Spanish. 45 second-generation Spanish HS enrolled in the same HS class, as well as a baseline of 12 Spanish monolinguals were placed into two groups- the instructed group and the uninstructed group. Despite an initial pool of 86 HL participants, 45 participated in the study, and only 32 were included in the analysis. This was largely due to the expected subject attrition that is commonly found in classroom research, as well as some technical difficulties faced during the course of the study. Additionally, because the instructional module was an online component assigned for homework, various participants either did not complete it, or did not achieve 80% accuracy. Furthermore, several participants did not fulfill all of the requirements of the study, particularly the pre- and/or post assessments. Also, as only two HL participants remained from the initial group that completed both the pre- and posttest (and not the instructional module), Montrul and Bowles were unable to include a HL control group to assess the effects of
the instructional treatment as originally planned.

The uninstructed group (n=16) was made up of participants who completed the pre-test but not the instructional module (or the posttest), whereas the participants in the instructed group (n=16) only completed the instructional module and the post-test (not the pre-test). During the phase of the study, the a-personal and dative experiences with gustar-verbs were neither taught in class nor highlighted in any related coursework.

Participants were given a pretest, which was comprised of an elicited written production task, and a written GJT. One week after the pretest was given to the corresponding group, the participants completed the instructional module, which was appointed as a homework assignment. It consisted of an explicit grammatical explanation of the uses of the DOM, followed by three practice exercises, where participants received immediate, explicit feedback (both positive and negative). As it was assigned for homework, the module was untimed and self-paced. Lastly, the posttest was another elicited written production task, and a written GJT.

Results from both the production and GJ task revealed that the instruction aided in the improvement of HL’s sensitivity to the a-marking in general, and that the instructional treatment was effective overall, due to the significant improvements on the different tasks from pre- to posttests. Additionally, it was suggested by the significant gains from the HL in their posttest results that both positive and negative evidence facilitates classroom heritage language learning, at least in the short-term.

Montrul and Bowles (2010) admit that their study was heavily flawed. For one, there was no control group, and in addition, several participants (in the uninstructed group) did not complete the posttest. The instructed group did not complete a pretest. As
the treatment was completed at home for homework, it is impossible to ascertain that it was in fact the participants themselves completing the module, or that participants did not consult with outside resources. Furthermore, because the instruction contained a variety of variables, such as explicit metalinguistic information, positive and negative feedback, and others, it is not possible to accurately determine which factor (or factors) contributed to the learner gains. Lastly, the study could have benefited, once again, from online measures to analyze how HS respond to feedback (positive or negative) and explicit metalinguistic information, as well as how they process this, in order to measure any correlation to the former with their posttest results. Also, a delayed posttest would have aided the study in investigating if these results held true long-term.

Kang (2010) investigated the function of negative evidence in Korean heritage language learners. Participants were asked to perform two communicative tasks with the research (a story sequencing task, and a spot-the-difference task), and were assigned to one of four experimental groups (including a control group), which varied in the explicitness of feedback (explicit or implicit), as well as the positioning of the feedback (proactive or reactive) on the use of Korean past tense morphology. The explicit group received negative evidence which ranged from metalinguistic feedback to providing the correct form (or instructions on how to do so), whereas the implicit group received feedback in the form of reformulation to repeating the erroneous production of the participant. Assessments included a GJT and a picture-description task. Although the results showed that both explicit and implicit feedback groups (along with the reactive positioning) performed better on the GJT than the proactive positioning and control group, there was no discernible difference between the feedback groups. Additionally,
results from the picture-description task showed that all three groups outperformed the control group, but again, results were not able to tease apart any difference between the feedback groups. Kang suggested, however, that negative evidence, particularly explicit negative feedback, may benefit HL learners’ development of morphosyntax in the classroom setting. Interestingly, Kang noted that “although the benefits of negative evidence faded over an extended period, the provision of negative evidence led the heritage learners to retreat from the non-target-like form in their learner language and approximate the target from overall” (p. 596). However, unlike previous studies that employed a delayed posttest, Kang’s delayed assessment was given on week 6, 4 weeks after the treatment (versus the typical two-week delay). Although this may not make a difference per se, it could explain as to why HL learners in Kang’s (2010) study showed a dwindling of the performance scores.

Overall, the results in the computerized feedback strand are inconclusive regarding the effect of the degree of explicit feedback in SLA. For the purposes of the current study, and in order to facilitate in operationalizing this construct, feedback will be classified as +/- explicit feedback. Explicit feedback [+EF], in this study, can be categorized as the provision of metalinguistic information (e.g. providing a rule, or directing the learner to a specific form) to the learner when interacting with the input in regards to a certain structure or form, whereas less explicit feedback [-EF] can be operationalized as providing information regarding the accuracy (or inaccuracy) of the production by the learner (i.e. no metalinguistic information is provided) (cf. DeKeyser, 1995).
CHAPTER 4: DEPTH OF PROCESSING

Overview

Depth of processing (DOP) 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 grammatical or lexical items in the input” (Leow, 2015: 204). DOP is accredited to the cognitive psychology field, more specifically, to Craik and Lockhart’s (1972) study, which first presented a ‘levels of processing’ framework. This framework used this concept to refer to conceptual or semantic processing (i.e. deep processing) and perceptual processing (i.e. shallow processing). Deep processing is defined as when the learner decodes a word, for instance, in relation to its significance and possible relationship with previously existing words in the current knowledge system. This category of processing involves elaboration rehearsal, which in turn encompasses deeper, more meaningful analysis of the item, such as the activation of prior knowledge. This ultimately leads to greater recall of the item. Shallow processing, on the other hand, may either be structural processing, which involves encoding the physical features of an item (such as the physical appearance of a word), or phonemic processing (the sound of the word). Unlike deep processing, the possibility of retention of items that were processed shallowly is not strong.

This framework was eventually adapted into the SLA field, where it has been employed to investigate attention (e.g. Shook, 1994; Gass, Svectics & Lemlin, 2003); shallow vs. deep processing (e.g. Bird, 2012); levels of awareness (e.g. Hsieh et al., 2015; Leow, 2012), mental effort and production (e.g. Calderón, 2013; Leow et al., 2008); and
quality of noticing (e.g. Qi & Lapkin, 2001). However, one SLA strand which has not employed DOP analysis is the HS strand. As such, in order to describe and to monitor how (deeply) heritage Spanish learners are processing feedback and input, Leow’s (2015) Model of the Learning Process in Instructed SLA will be utilized (see Image 1 below).

This model, “premised on the role of attention in the process of learning an L2” (Leow, 2015: 241), is supported by most theoretical perspectives in the field of SLA (e.g., Gass, 1997; McLaughlin, 1987; Robinson, 2005; Schmidt, 1990; Tomlin & Villa, 1994). It is separated into 6 stages, and emphasizes that learning consists of both processes and products. The model is further divided into three paramount processing stages: 1) the input processing stage, 2) the intake processing stage, and 3) the knowledge processing stage. As this study will focus on the first two stages, particularly on the second stage (the intake processing stage), the first two stages will be elaborated below.

Figure 1: Leow’s (2015) model of the learning process in instructed second language acquisition
**Input Processing Stage**

The first processing stage in Leow’s model is the input processing stage. This process is situated between the input (the target language to which the learner is being exposed) and the intake held in working memory. Leow (1993) defines intake as

That part of the input that has been attended to by the second language learners while processing the input. Intake represents stored linguistic data which may be used for immediate recognition and does not imply language acquisition (p. 334).

Depending on the level of attention paid to the input by the learner (peripheral, selective, or focal), this stage may also be accompanied by cognitive registration, awareness, or depth of processing, thereby dividing the input process stage into three distinct phases: attended, detected, and noticed intake.

The first phase, attended intake, is accompanied by peripheral attention to the input, in other words, a very low level of processing, lack of cognitive registration and awareness of the linguistic items; where cognitive registration is the detection of some stimuli in the input, and detection is “the process that selects, or engages, a particular or specific bit of information.” (Tomlin & Villa, 1994: 192). Because this particular low-level intake will unlikely be processed any further, it is most likely to be discarded from working memory.

Detected intake is a second phase of the input processing stage, and assumes that a certain amount of attention (selective) was minimally paid to the linguistic target input, and is also accompanied by a very low level of processing, and cognitive registration with no awareness of the target information. In accordance with Tomlin and Villa’s (1994) notion of detection, whereby detection is the sole necessary function of attention
that will lead to further cognitive processing and the possibility of learning, detected intake has a possibility for further processing; more so than attended intake. However, this is contingent on the learner’s working memory, and “whether a higher level of processing or cognitive effort is allocated subsequently to the detected intake” (Leow, 2015: 243).

noticed intake is the third intake phase of the input processing stage. Although it comes with a low level of awareness, defined as “a particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus” (Tomlin & Villa, 1994: 193), and includes cognitive registration, it is still accompanied by a low level of processing. This phase is in line with Schmidt’s (1990) notion of noticing, which unlike Tomlin and Villa’s (1994) notion of detection, involves attention with a low level of awareness. Due to its comparatively higher level of processing and cognitive effort (in comparison to attended and detected intake), noticed intake holds the highest potential to be stored in the learner’s working memory and made available for further processing, which may lead to the incorporation of the target item into the learner’s grammatical system. This has been reported in various studies addressing the effects of noticing in L2 development, particularly utilizing think aloud (TA) protocols (i.e. concurrent verbal protocols) (e.g., Leow, 1997, 2000; 2001a; Leow et al., 2003; Rosa & Leow, 2004) and more recently, eye-tracking (ET) (e.g., Godfroid et al., 2013). The model also underscores that although all phases of intake, including to a lesser degree, attended intake, may be kept in working memory, they can all be discarded if they are not minimally processed further (e.g., Bowles, 2005; Hama & Leow, 2010; Leow, 2000, 2001b).
**Intake Processing Stage**

The second stage of Leow’s model deals with the processing of the preliminary intake, which refers to the three types of intake (attended, detected, or noticed). The processing may be done in one of two ways, contingent on the depth or level of processing, and/or the cognitive effort. Data-driven processing (cf. Robinson, 1995), also known as item learning (cf. Gass, 1997) allows for the data to enter the L2 learner’s internal system as chunks of language that have not been systemized without exhausting much cognitive effort in its processing. If further examples follow suit (also without higher levels of processing), they may be processed in a similar fashion, “thus forming a collection of encoded discrete data or entities lodged in learners’ L2 developing system” (Leow, 2015: 243).

Chaudron (1985) defines the concept of intake as “the mediating process between the target language available to learners as input and the learners’ internalized set of the L2 rules and strategies for second language development” (p. 1). Chaudron further adds that the learner is the one who yields an active role in the L2 learning process, as s/he is the one who controls the intake. In his view, Chaudron states that intake should not be considered a singular process; as intake consists of three stages of processing of information, which moves from the preliminary stage of intake to the final (discussed below). According to Chaudron, there is first a perception of the input, followed by the recodification and encoding of the message into the long-term memory. Finally, there is a series of phases in which the learner incorporates and assimilates the linguistic information into the developing grammar. This entire process description occurs as a
continuum from preliminary to final intake.

Preliminary intake can also be processed through conceptually-driven processing (cf. Robinson, 1995) that involves a higher level of processing or cognitive effort. Due to the higher levels of awareness that may accompany this type of processing, the linguistic data are kept in working memory facilitating its prospective entry and assimilation into the learner’s systemized grammatical system (cf. Leow, 1997, 2015; Rosa & Leow, 2004; Sachs & Suh, 2007).

There are two ways in which systemized learning may occur: 1) linking the linguistic data to prior knowledge or 2) through data-driven processing. Through the linking of prior knowledge, the linguistic data are connected cognitively to some prior knowledge of related linguistic data that are utilized to assist the encoding and decoding of the linguistic information that is present in the preliminary intake. By doing so, it couples higher levels of awareness with higher levels of processing. Subsequent data that reactivate the same prior knowledge will result in a reduction of the depth of processing and the level of awareness necessary to process the linguistic data.

Another manner in which systemized learning may occur is through data-driven processing, whereby the linguistic exemplars have been encoded and lodged in the learner’s L2 developing system un-systemized. When consequent data come into the system to be analyzed, intake processing may occur through the activation of old or new data (data that had been previously stored in the internal system unsystemized). Based on the depth of processing, amount of cognitive effort, and/or the levels of awareness, this activation may lead to either implicit or explicit learning, in which case the data would be stored in the grammatical system within the corresponding system learning component.
For instance, a low level of processing could eventually lead to an implicit restructuring of the linguistic data and the implicit systemized learning. However, this type of processing is contingent upon myriad factors, including the occurrence of a great amount of exemplars in meaningful contexts, as well as ample time allotted for processing and internalizing the data, thereby having the data available for later use.

Advanced levels of depth of processing normally include hypothesis testing and rule formation (Leow, 1997; Rosa & Leow, 2004). The higher the depth of processing, the higher the potential level of awareness: from awareness at the level of noticing to awareness at the level of reporting to awareness at the level of understanding (Leow, 1997). However, it is imperative to note that while it follows that higher depths of processing may result in higher levels of awareness, the relationship may not always hold true. Although it is reasonable to note that a higher depth of processing would involve a greater amount of cognitive effort when trying to form a rule or a hypothesis about a certain grammatical rule, awareness at the level of understanding is only reached when the correct underlying rule is obtained, or the form-meaning connection of the lexical item is attained.

As linguistic data becomes automatized through subsequent multiple exposures and meaningful practice, the roles of awareness and depth of processing may change. It may lead to

a sharp reduction in the depth required to process the relevant linguistic data in the L2 input, which may lead to a less important role for awareness and depth of processing during intake processing of previously learned linguistic data. (Leow, 2015: 245)

However, it should be noted that even during the same exposure, depth of processing may vary.
Empirical Studies on Depth of Processing

Craik and Tulving (1975) examined a series of experiments in order to investigate levels of processing and memory. In every experiment, participants (between \( n=12 \) and \( n=36 \)) were visually exposed to a word, and were then asked a question in order to prompt them to process the word in one of three levels of analysis, varying in complexity. The questions asked induced a lower (analyzing the word’s physical structure), medium (analyzing at the phonemic level), or higher (analyzing the semantic category) level of processing. The participants then completed a free recall, cued recall, or recognition retention test of the words in the experiment. They found that when deeper-level questions were asked about a specific word, the consequent retention of the word was higher. Overall results showed that DOP is not a continuum, and that processing at a shallow level is sufficient to encode new items in the semantic memory, whereas the processing of items at a deeper level encodes them in the episodic memory.

Qi and Lapkin (2001) explored what learners notice, the role of processing when writing an L2 task, as well as what effects this has on their textual production. The two participants were Mandarin speakers learning English as an L2 with different levels of proficiency. Throughout the three sessions, participants wrote a text, compared it with a reformulated version (from the researchers), and revised it without access to the draft in session 2. TAs were used and language-related events (LRE) were analyzed. Results revealed that the participant with a higher proficiency resolved more LREs, and revealed that instances of deeper processing resulted in more corrections, which ultimately led to a better rewrite. Therefore, DOP appeared to be correlated with proficiency levels. This led
Qi and Lapkin (2001) to claim that noticing without understanding, or noticing with no purpose does not hold the same influence as noticing with understanding. The authors further concluded that the reformulation proved to be a better pedagogical tool than the traditional ‘error correction’ technique. However, caution must be taken with these results, as results cannot be extrapolated to the general population from only 2 participants.

Leow, Hsieh, and Moreno (2008) employed concurrent measures (e.g. TAs) to determine whether participants were processing simultaneously meaning and form, and to also investigate whether limited attentional resources during input processing compete to be assigned to either meaning or form. 72 2nd semester L2 students of Spanish were randomly assigned to one of five groups that either a) read for meaning or b) read for meaning with further instructions to circle instances of a lexical item (the word *sol*), a feminine article (*la*), a masculine article (*lo*), or a verbal morpheme (*-n*), which was an extension of VanPatten (1990) who had postulated that paying attention to grammatical form would negatively affect attention to (processing for) meaning. Immediately following the task, a 10-item multiple-choice comprehension test was administered.

The findings, which revealed no significant differences in comprehension between any of the conditions, showed a clear contradiction from those in VanPatten (1990). Leow et al. (2008) postulated that this could be due to the difference in modality (VanPatten’s input was aural whereas this study’s was written). In regards to the TA data, three levels of processing were engendered (levels 1, 2, 3), where participants in Level 1 merely circled the target form, Level 2 provided some additional proof of attending to the target form, and the participants in Level 3 either translated or interpreted the target form.
Overall, deeper processing was observed for the lexical item sol (73%), than for la (45%), lo, or –n (31%). However, the small number of participants in the study precluded the use of statistical analyses. An additional limitation aside from the aforementioned is the lack of a non-TA group in order to omit doubt of potential reactivity. Leow et al. (2008) concluded that one possible explanation for the results in comprehension across the experimental groups could be the relatively low level of processing of the target forms.

Morgan-Short, Heil, Botero-Moriarty, and Ebert (2012) replicated Leow et al.’s (2008) study with the intention of discovering the potential relationship between comprehension and depth of processing. Although the materials and procedures remained the same as the Leow et al. (2008) study, modifications included increasing the participant size from 72 to 308, as well as the inclusion of a non-think-aloud (silent) group, thereby addressing the limitation of the previous study. Morgan-Short et al. (2012) reported a small yet significant correlation between comprehension scores and the level of depth of processing, reporting that the results “demonstrated that the deeper the learner’s level of processing, the better his or her comprehension score of the written text” (p. 676). This is in line with Leow et al.’s (2008) results that participants who showed a deeper level of processing of the target structure did not suffer lower scores in comprehension (compared to participants that did not process the target structure deeply), thereby also contradicting VanPatten’s Primacy of Meaning Principle (2004), which states that processing the written input in an L2 for both meaning and form concurrently may be detrimental to comprehension.

Calderón (2013) sought to test a few of the postulations of Leow’s Model,
focusing on the early stages of the L2 learning process, particularly by addressing the possible existence of the different levels of intake and the role of depth of processing. Her study further addressed if the type of linguistic item (grammatical versus lexical) played a role. She utilized concurrent measures such as eye-tracking and think aloud protocols on 96 beginning learners of Spanish. The participants read a text in the L2 and then were asked to complete production, recognition, and comprehension assessments in a pre- and posttest design. Calderón’s (2013) results revealed that 1) no reactivity existed in regards to utilizing concurrent verbal reports in her study, 2) different levels of intake seemed to exist, as documented in the larger recognition scores of the noticed intake group compared to the attended intake group, and 3) there appeared to be differences in how different linguistic items were processed. As Calderón (2013) noted, “regarding depth of processing, as grammatical depth of processing increased, so did accurate production and comprehension” (p. 254). Furthermore, regarding lexical depth of processing, “as depth of processing lexical item increased, recognition also increased” (Calderón, 2013: 254). Additionally, there were significantly higher recognition scores for the grammatical forms than for the lexical items. Lastly, her final result showed that 4) depth of processing may not only play a role in the successive processing of the intake, but also appeared to expedite the deeper level of processing needed for the integration of the intake into the L2 developing system.

There were a few limitations with Calderón’s study. Firstly, the task, which she utilized was a reading task, which warrants further exploration of different types of experimental tasks. Second, given the few (if any) prior studies that have attempted to distinguish differential processing of grammatical forms and lexical items, the descriptors
would need to be more finely-tuned and tested elsewhere. Third, while her study was the first and only one to test the tenets of Leow’s Model with regard to the input-to-intake stage, further research is needed to test the remaining stages of the model.

**Awareness in Processing**

Various studies in the field of SLA have investigated the construct of awareness with and without levels of processing. A majority have found, through the use of concurrent measures (i.e. TAs) that awareness facilitates L2 learning (e.g. Rosa & Leow, 2004; Robinson, 1995, 1997). Similar to the levels of depth of processing, levels of awareness have been reported and categorized into three levels: awareness at the level of noticing, awareness at the level of reporting, and awareness at the level of understanding (e.g. Leow, 2001; Rosa & Leow, 2004; Rosa & O’Neill, 1999). It stands to reason that depth of processing is comparable to the levels of awareness (e.g. de la Fuente, 2015; Hsieh et al., 2015; Leow, 2001; Medina, 2015; Rosa & Leow, 2004; Rosa & O’Neill, 1999; Sachs & Suh, 2007), as protocols from the aforementioned studies have shown that protocols coded as ‘awareness at the level of understanding’ are typically associated with the same criteria as the highest level of processing (e.g. hypothesis testing, rule formulation, conscious activation of prior knowledge). Additionally, the lower levels of awareness tend to align with the other levels of processing, whereby cognitive effort is applied, however, the target rule/form is not reached, which would promote it to the highest level of awareness (at the level of understanding).

Leow (2015) revisited his previous studies to analyze TA protocols of his participants (from his Leow, 1997 and Rosa & Leow, 2004 studies), and noted:

As can be seen, concurrent data reveal that as the level of awareness rises, so too does the level of processing or amount of cognitive effort and
elaboration. Awareness at the level of noticing is usually correlated with a low level of processing and cognitive effort, at the level of reporting a medium level of processing and cognitive effort, while at the level of understanding a very high level of processing and instances of hypothesis testing and rule formulation are usually reported...depth of processing, then, may be closely tied to levels of awareness and can be used to account for the statistically superior performances reported for participants evidencing higher levels of awareness or learning explicitly when compared to lower levels (e.g. de la Fuente, 2015; Leow, 1997, 2000, 2001; Martínez-Fernández, 2008; Medina, 2015; Rosa & Leow, 2004; Rosa & O’Neill, 1999; Sachs & Suh, 2007). (Leow, 2015, p. 220)

**Concurrent Data-Elicitation Procedures**

If one is to argue that the learning of a language takes place in the internal system of the learner, then it stands to reason that the tools utilized in empirical studies should be ones that are able to show (as much as possible) the learner’s cognitive processes while interacting with the L2. Concurrent verbal reports, referred to as Think Aloud (TA) protocols from henceforth, are arguably the primary methodological tool of choice in the attention and awareness strand of SLA to measure these processes (cf. Leow, Grey, Marijuan, & Moorman, 2014). By utilizing TA, the researcher is not only able to glean insights on how the participants interact with the L2 data, and what roles attention and awareness play, but is also able to garner insights on the levels of awareness, types of processing (data-driven vs. conceptually-driven), depth of processing, and levels of cognitive effort. According to Leow et al. (2014), data gathered from TA serve at least two paramount purposes: 1) allowing researchers to operationalize and measure the roles of cognitive processes postulated to play a role in the learning process; and 2) providing data that establish the representativeness of participants in each experimental cell (p. 114). This, in turn, raises the level of internal validity (cf. Leow, 2015 for further elaboration of this issue), which is critical for the legitimacy of the study and its results.
The data that has been collected from participants using TA have been utilized to support various degrees of several cognitive concepts, such as levels of awareness (e.g., Leow, 2001a, 2001b; Rosa, 1999; Rosa & Leow, 2004; Sachs & Suh, 2007; Schmidt, 2001), the construct of (un)awareness (Hama & Leow, 2010; Leow, 2000), levels or depth of processing (e.g., Hsieh, Moreno, & Leow, 2015; Leow et al., 2008; Morgan-Short et al., 2012; Qi & Lapkin, 2001), activation of prior knowledge (e.g., Leow, 1998), different types of processing, in other words, conceptually-drive vs. data-driven (e.g., de la Fuente, 2015; Leow, 1998a), and so forth. Additionally, TAs have aided researchers in ensuring that their participants have fulfilled what was required of them in an experiment (i.e. followed instructions) (e.g., Alanen, 1995; Leow, 2000). Furthermore, although widely used in the attention and awareness strand of SLA, as previously mentioned, TA protocols have also been utilized in other strands of research, such as L2 reading and writing (e.g., Cohen & Cavalcanti, 1990), L2 test-taking strategies (e.g., Cohen, 2000), interlanguage pragmatics (e.g., Kasper & Blum-Kulka, 1993), translation (e.g., Jaaskelainen, 2000), and so on.

TA protocols may vary from participant to participant given that not all participants think, talk, or react the same way. Similarly, TA protocols themselves vary because not all TAs are created equal. A distinction is found, for instance, between introspective and retrospective TA. Introspective verbal reports are collected while the participants are performing the task at hand. A benefit of this is that the verbalizations are not “constrained by memory” (Leow, 2015: 141). A disadvantage of introspective reports, however, is the issue of reactivity (cf. Bowles, 2010 for a review of the issue of reactivity). Retrospective verbalizations are collected immediately after processing has
taken place, either during specified times throughout the task (online) or immediately after the task has been completed (offline). Although reactivity would not be an issue in retrospective reports, the disadvantage of veridicality (i.e. not reporting accurate information due to memory decay) is introduced. Moreover, Ericsson and Simon (1993) warn on relying heavily on retrospective verbalizations, as it is impossible to ascertain that the information gathered from the participants is in fact what occurred at the moment of processing within the task.

An additional distinction is made between metacognitive and non-metacognitive verbal reports. In a non-metacognitive TA protocol, the participants are asked to focus solely on the task at hand, and to verbalize every thought going through their mind as they complete the task without worrying about explaining their thoughts. In essence, they are asked to solely think aloud. These are considered Type 1 and Type 2 verbalizations. In metacognitive verbalizations, however, the learners may be asked to provide specific information in their TA, such as reasoning or explanations, and learners provide a metacognitive report on their assumptions of the processes (Leow, 2015). These are categorized as Type 3 verbalizations. It has been suggested that in order for the verbalizations to truly reflect the processes of the learner, an introspective, non-metacognitive TA verbalization should be elicited (Cohen, 2000; Ericsson & Simon, 1993).

Although the possibility of reactivity is usually cited as a disadvantage of using TA protocols, only a handful of studies have reported positive reactivity in their studies (e.g., Rossomondo, 2007; Sanz et al., 2009, Experiment 2; Yanguas & Lado, 2012) and one study (Sachs & Polio, 2007) has reported negative reactivity. At the same time,
several studies have reported no significant reactivity in their studies (e.g., Bowles, 2008; Bowles & Leow, 2005; Calderón, 2013; Egi, 2008; Leow & Morgan-Short, 2004; Sachs & Suh, 2007; Sanz et al., 2009, Experiment 1; Stafford, Bowden, & Sanz, 2012; Yoshida, 2008). In addition, as pointed out in both Bowles (2010) and Leow et al. (2014), the issue of reactivity may also be associated with type of task, namely, problem-solving, which does not pose much potential for reactivity (Bowles, 2008; Lass, Klettke, Lüer, & Rhulender, 1991; Rhenius & Deffner, 1990; Russo, Johnson, & Stephens, 1989) versus reading, which posits some potential for reactivity (Bowles & Leow, 2005; Ellis, 2005; Rossomondo, 2007). Moreover, explicit tasks and assessments (such as controlled written production tests, and untimed GJTs) fare better in regards to limiting the issue of reactivity over tasks and assessments created to elicit implicit knowledge, such as timed GJT (Ellis, 2005). In order to remedy the potential issue of reactivity, it has been suggested to include a –TA (no Think Aloud) group in the study to ascertain whether there are reactivity effects, and to determine to what extent they play a role (Bowles, 2010; Leow & Morgan-Short, 2004). Table 5 summarizes some of the aforementioned studies.

Bowles (2010) conducted a meta-analysis addressing the issue of reactivity. She gathered data from various published studies in the field of SLA, which reported reactivity, including data from one unpublished empirical study, with the task of determining if concurrent verbal reports (e.g. Think Alouds (TA)) were reactive with regards to accuracy and/or latency when employed with verbal tasks. Results from her meta-analysis concluded that in general, thinking aloud while performing a verbal task has, if anything, only a small effect on post-task performance (i.e. those in TA groups
performed only slightly better or worse). However, the results on time-on-task were more significant: Bowles found that thinking aloud increases time on task, particularly during reading comprehension tasks. Furthermore, due to the discrepancy of all of the studies analyzed in the meta-analysis (a violation of the assumption of homogeneity), Bowles was unable to provide a definitive answer on the matter of reactivity and TAs. However, she went on to explain that several factors accounted for a significant amount of variance between the studies analyzed, including the proficiency level of the participants, the explicitness of the instruction provided in the study, as well as the type of TA (metacognitive vs. non-metacognitive).

Table 5: Summary of selected think-aloud (TA)/reactivity empirical studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Materials</th>
<th>Type of TA</th>
<th>Assessment Tasks</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bowles &amp; Leow (2005)</strong></td>
<td>45 fifth semester college students enrolled in Spanish language program</td>
<td>Magazine article</td>
<td>Both types (metacognitive &amp; non-metacognitive with a silent control)</td>
<td>Comprehension: 10 item multiple choice about information expressed with the TS; production: two fill-in-the-blanks with 20 items each</td>
<td>No significant effect for type of TA; TA group took longer to complete task</td>
</tr>
<tr>
<td><strong>Bowles (2008)</strong></td>
<td>194 students in two private US universities enrolled in a first-semester Spanish course</td>
<td>Computer maze program</td>
<td>Both types (metacognitive &amp; non-metacognitive with a silent control)</td>
<td>Controlled written production tests with target structure and distractors</td>
<td>Non TA groups performed better for previously encountered exemplars, but no significant difference for novel exemplars</td>
</tr>
<tr>
<td><strong>Egi (2008)</strong></td>
<td>44 participants</td>
<td>Japanese structures: numeral classifier and the morpheme –te</td>
<td>Not concurrent TA (stimulated recall)</td>
<td>Recall stimuli, and recall stimuli + verbalization and control group had posttest, and late post test</td>
<td>Stimulated recall group did not significantly differ from control group</td>
</tr>
<tr>
<td><strong>Leow &amp; Morgan-Short (2004)</strong></td>
<td>77 adult, college-level students enrolled in a first-year Spanish language program</td>
<td>Article to read</td>
<td>Non-metacognitive with a silent control</td>
<td>Comprehension assessment with short and multiple choice answers; intake assessment with multiple choice recognition task; controlled written production assessment</td>
<td>TA does not have a detrimental effect on comprehension, intake, or controlled written production in a reading task</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Materials</td>
<td>Type of TA</td>
<td>Assessment Tasks</td>
<td>Results</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>-----------</td>
<td>------------</td>
<td>-----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Morgan-Short, Heil, Botero-Moriarty, &amp; Ebert (2012)</td>
<td>308 college students in a third-semester Spanish course</td>
<td>Participants read a written passage</td>
<td>Non-metacognitively with a silent control</td>
<td>Includes a control group that only reads for comprehension</td>
<td>Non TA groups performed slightly better on comprehension test; however, the size of this effect was minimal</td>
</tr>
<tr>
<td>Rossomondo (2007)</td>
<td>161 learners in one-semester accelerated elementary Spanish program course at a large American university</td>
<td>Reading passage either with LTIs or without LTIs</td>
<td>Non-metacognitively with a silent control</td>
<td>All received multiple choice comprehension test; some received cloze passage production test while others received multiple-choice recognition test</td>
<td>No significant difference between groups for comprehension, but the TA group performed better in recognition and production</td>
</tr>
<tr>
<td>Sachs &amp; Polio (2007)</td>
<td>2 studies: 1) 15 adult students of English, varying native languages enrolled in a high-intermediate ESL program and living in the US for between 1 mo. and 1 yr.; 2) 54 ESL students, different levels with varying native languages</td>
<td>Composition, comparison, revision task to gauge ability to revise accurately - some given error corrections, others native-like version (reformulation), and others thought aloud in the L2 after receiving reformulations</td>
<td>Non-metacognitively with a silent control</td>
<td>Revision of written work</td>
<td>Error correction led to better performance than reformulation or reformulation + TA and more accurate revisions found in reformulation than in reformulation + TA - in composition stage, the TA appear to distract the learner</td>
</tr>
<tr>
<td>Sachs &amp; Suh (2007)</td>
<td>30 college-age Korean ESL students in the intermediate to high-intermediate level; ages 19-27; 19 female, 11 male</td>
<td>Short story in L1 (Korean), 20 picture and vocabulary cards in English (14 relevant ones and 6 distractors)</td>
<td>Non-metacognitively with a silent control</td>
<td>Recognition task with multiple choice questions with 4 options; production task using interactive, computer-mediated story-retelling</td>
<td>No strong conclusion can be drawn, no significant interaction effect found</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Materials</td>
<td>Type of TA</td>
<td>Assessment Tasks</td>
<td>Results</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sanz, Lin, Lado, Bowden, &amp; Stafford (2009)</td>
<td>24 college students (between ages of 18-35) native speakers of English without knowledge of case-marking languages</td>
<td>Experiment 1: Computer-administered lessons on case markers for Latin nouns consisting of (a) vocabulary presentation, (b) grammar lesson, (b) task-essential practice session, and (d) explicit feedback, but no explicit grammar lesson</td>
<td>Non-metacognitive with a silent control</td>
<td>Aural interpretation test, a written grammaticality test, and a sentence production test.</td>
<td>There was reactivity on latency (on GJT); no reactivity on posttest accuracy (on all tests) and latency (on aural interpretation and sentence production posttests)</td>
</tr>
<tr>
<td>Sanz, Lin, Lado, Bowden, &amp; Stafford (2009)</td>
<td>24 college students (between ages of 18-35) native speakers of English without knowledge of case-marking languages</td>
<td>Experiment 2: consisted of task-essential practice (as described for the first experiment) and explicit</td>
<td>Non-metacognitive with a silent control</td>
<td>Aural interpretation test, a written grammaticality test, and a sentence production test</td>
<td>Showed positive reactivity (on aural interpretation test); no reactivity on posttest accuracy (on GJT and sentence production tests)</td>
</tr>
<tr>
<td>Stafford et al (2012)</td>
<td>65 Spanish-English bilingual adults living in the US (including heritage speakers &amp; late bilinguals)</td>
<td>Latin morphosyntax related to the assignment of thematic agent/patient roles to nouns at the sentence level</td>
<td>Does not specify</td>
<td>2 interpretation tests: 1 aural, 1 written; grammaticality judgment test; production test</td>
<td>All groups significantly improved their ability to interpret written and aural Latin sentences; +grammar explanation, +explicit feedback group learned and maintained through 3-week delayed posttest. A subset of participants' performances did not reveal any evidence of reactivity.</td>
</tr>
</tbody>
</table>
Of particular interest to this dissertation is Yanguas and Lado’s (2012) study of the effect of reactivity on bilingual writing on heritage learners of Spanish. Participants \(N=37\) enrolled in a Spanish for HL speakers class were asked to develop their own story in a free-writing picture-description activity in the past tense; 20 were asked to think aloud in English while performing the task. Results showed that thinking aloud while performing the writing task benefited the HL participants’ fluency and accuracy. The authors supported their results by considering that verbalizing (nonmetacognitively) provided additional chances for HL to be aware of linguistic forms during production and to “monitor their own writing processes and acquire helpful strategies” (Yanguas & Lando, 2012, p. 393). Therefore, although reactivity was found in this study, it was positive reactivity, and thus provided an advantage for these learners. However, these results should be taken cautiously, as the participant size was very small, the participants were asked to think aloud in English (instead of being given an option between English and their heritage language), no assessments were provided to see if their fluency and accuracy did in fact improve by thinking aloud (i.e. did these participants have a higher
or lower working knowledge of these structures before the task), and various levels of HL’s proficiency should be studied before extrapolating the results to all HL.

**Research Questions**

The current study aims to address the gaps in light of the previous research discussed above identify how heritage Spanish speakers process linguistic data in Spanish as well as to investigate whether the type of feedback (i.e. +/- explicit) affects performance in a pretest, posttest, and delayed posttest design, in comparison to L2 learners of Spanish with a similar proficiency level. While doing so, an additional component will be to test the postulations made in the Model of the L2 Learning Process in Instructed SLA (Leow, 2015), particularly at the Intake Processing Stage (Stage 3). The following research questions will inform the present investigation:

1. How do heritage language learners of Spanish process incoming data in their heritage language?
   a. Is this process similar or different than L2 learners of Spanish?

2. Does the Speaker profile of participants have an effect on their subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task?)
   a. If so, does the effect last for 2 weeks?

3. Does the type of feedback (+explicit, -explicit, or control) have an effect on participants’ subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task?)
   a. If so, does the effect last for 2 weeks?
4. Is there a relationship between the levels of processing (low, medium, high) as measured by Think Aloud protocols and subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task?
CHAPTER 5: RESEARCH DESIGN AND METHODOLOGY

The current chapter outlines the pilot study conducted prior to the main data collection, as well as a detailed overview of the research methodology proposed to answer the research questions.

Pilot Study

Overview

The pilot study phase of the dissertation was conducted in the spring of 2016, prior to the main data collection. The pilot study was conducted on a much smaller scale and thus was not for the purpose of attaining any statistically significant findings. Rather, it was conducted to test the validity of the materials, assess the average time of each task, and to investigate the prospect of obtaining significant results in the major study. As such, it only included some of the methodological components of the proposed major study.

Two research questions guided the pilot:

1) Is there an interaction between type of feedback (+/- explicit), in terms of
   a) their performance on consequent assessments (pretest, posttest, and delayed posttest)?
   b) how they process (as defined by their level of processing)?
2) Is there a relationship between the levels of processing (low, medium, high) as measured by Think Aloud protocols and subsequent performance based on a battery of tests?

Participants

Participants were 30 Spanish L2 learners enrolled in a 5th semester (advanced level I) Spanish class at a Mid-Atlantic university. 27 of the participants were L1 English
speakers, 2 were L1 Chinese speakers, and 1 was an L1 Russian speaker. Participants were recruited towards the end of the semester prior to receiving instruction on the target linguistic item. However, out of the 30 original participants, 1 was excluded from the analysis due to achieving a perfect score on the pretest, signifying knowledge of the target linguistic item. Of the 29, 8 (27%) identified as male and 21 (73%) identified as female. On average, participants had been enrolled in a Spanish course for 6 years, (ranging from 1.5 to 14 years). None had studied abroad in a Spanish speaking country. 6 of the 29 (21%) were not born in the United States, and 7 of the 29 (24%) spoke a language other than English natively (Yoruba, Mandarin Chinese, Russian, Creole, Patois, and French). However, post hoc analyses did not show any differences in performances. No heritage Spanish speakers were present in the pilot study.

Methodological Issues

The purpose of this pilot was to test the validity and reliability of the materials prior to the principal study. Throughout the pilot, various issues arose that, although minimal, served to provide insight as to how the overall study might be improved upon.

The short explicit instruction given during the first session to ensure baseline knowledge of the imperfect subjunctive in contrary-to-fact conditional sentences proved to be popular with participants, as it was colorful and interactive. However, because it was presented on Microsoft PowerPoint, the researcher was unable to control for some participants clicking on the ‘next’ button rather than on the correct answer in some of the final slides assessing their knowledge. This led to some participants not receiving the appropriate feedback as to whether they were right or wrong. Regardless, because the
explicit instruction was presented in a classroom setting by the researcher, this was no longer an issue.

A potential concern that was addressed in this pilot study was the lack of different distractor items. During the treatment and immediate posttest, the only distinction present was that of the imperfect and pluperfect subjunctive. An analysis of the TA protocols revealed that a few of the participants realized that if the second clause had haber + participle, then the first clause must be hubiera + participle. If not, then it would be the default ‘other’ conjugation. Although the majority of participants were able to score highly on the immediate posttest based on discerning between these two linguistic forms, the researcher felt that it might still not serve as accurate proof that the structure was learned. Therefore, for the delayed posttest, 12 additional sentences were added to each of the scenarios in the preterit tense (since all the scenarios took place in the past). There were fewer participants who scored highly on this test; however, it might have been a potential limitation as to whether these scores were a reflection of the processing level, or the fact that new distractors were added. As such, these new distractors were included in all the assessments.

Apart from the distractor items, there was a possibility that the participants would be able to infer patterns from their previous answers in the controlled production task, as there were several instances present on one page. To remedy this, the scenarios were separated into individual sentences labeled with a number, cut into strips, and placed into an envelope in numerical order. The participants were instructed to place the sheets of paper face-down next to them as they completed each sheet, and were not allowed to refer back to any completed sheet.
Additionally, the recognition task did not set to measure any form of learning, and served no purpose in answering any of the research questions, and was therefore discarded. In its place, a semi-spontaneous picture description task was added, which aided in 1) seeing how these forms were produced by the participants before and after the experiment, and 2) providing additional opportunities to produce the form in a different format.

During the pilot, SuperLab software was used for the experiment. However, given that one of the universities in which the study was conducted was not able to obtain the software in time for the data collection, another online platform was used (SurveyGizmo).

While not initially a concern, an analysis of the TA protocols revealed that some participants did not know the definitions of some lexical items (e.g. *sobrino* - nephew, *tirar* – to throw). Though this did not affect the ability to complete the task, and these participants’ scores or levels of processing did not appear to suffer due to not knowing these items, for the principal study, these words were glossed.

A last concern was with the inability to address the paramount research questions regarding heritage speakers, as there was a paucity of HS at the time of the pilot. Additionally, a control group was not incorporated in the pilot due to lack of participants. In the larger study, a control group was included.

**Current Study**

**Participants**

Participants were 199 adult speakers of English and Spanish (either L2 or heritage language[HL] learners) drawn from two East coast American universities. The 105 L2
participants were enrolled in a Spanish as a Foreign Language 5th semester Advanced I class at a private university in the Washington, D.C. area. Participants were placed into this level either by a) taking a placement test, or b) successfully completing the Intermediate II course. Participants met three times a week for 50-minute class sessions. Both the imperfect and pluperfect subjunctive structures had not been formally covered in class, and would not be taught until the following semester. The 94 HL participants were enrolled in a Spanish for Heritage Spanish speakers course at a public university in Miami, Florida. Participants would also meet three times a week for 50-minute-long sessions, and other than self-identifying as a heritage speaker of Spanish, no other placement criteria were required to enroll in the class. Like the L2 participants, these HL participants had not received formal instruction on the past subjunctive structure at the time of the study. The programs at both institutions were communicative in nature, and placed an emphasis on the holistic instruction of listening, speaking, reading, and writing. The Advanced I course in the Washington, D.C. university utilized the book *Puntos de Encuentro*\(^2\), and the Heritage Spanish course in the Miami, FL university utilized two books: *Cómo aumentar su vocabulario*\(^3\) and *Gramática para la composición*\(^4\)

Of the original 199 participants, 24 were excluded from the statistical analysis. In the case of the L2 participants, it was due to: 1) not scoring above a 15 (out of 20) on the


controlled production task in the pretest for the distractor items, in order to ensure a baseline; 2) not scoring a 0 (out of 20) on the controlled production task in the pretest for the target structures, in order to control for prior knowledge of the target structure; or 3) not following instructions. Participants who attended both the pretest and delayed posttest sessions (but missed the experiment and posttest session) were grouped into a maturational control group. Therefore, comparisons could only be made on the pretest and delayed posttest. The same criteria were applied to the HL participants, with slight alternations, to account for their different experience in acquiring and/or learning Spanish, as well as to account for possible (albeit limited) exposure to the target structure. HL learners who showed a low level of prior knowledge of the target item (fewer than 5 out of 20) were allowed to remain in the study; the cut-off was decided upon averaging the pretest score. Additionally, participants not fitting the profile of a heritage speaker, as per Valdés’s 2001 definition, were also eliminated. This last criterion was included in the study because it was found during the study that there were two participants in the heritage speaker class who were monolingual Spanish speakers, having recently emigrated from Cuba. Additionally, another participant, also enrolled in the same class, was a native speaker of Haitian Creole, and was using the heritage Spanish class as his introduction to Spanish, due to scheduling conflicts. The final group of participants for the study was 175 in total; 85 heritage language learners, and 90 L2 learners.

For the L2 group, participants were between the ages of 17 and 22 (average age 19.42, SD= 1.19). 91% (N= 82) of the participants were born in the United States. The remaining participants’ countries of origin included South Korea (1), Canada (1), Belgium (1), Czech Republic (1), India (1), Italy (1), China (1), and Greece (1). For the
participants not born in the United States, their age of arrival to the US ranged from 3 months - 6 years old (average age 3.75). The length of residence in the United States for these participants ranged from 12-16 years. Only 11 participants were not native speakers of English: the L1s included Greek (1), Mandarin (1), Italian (1), Hindi (1), Yoruba (1), Korean (2), German (1), French (1), Armenian (1), and Bengali (1). However, all 11 participants identified speaking English before the age of 6. Seven participants also acknowledged having experience learning a foreign language apart from Spanish: Hindi (1), French (3), Arabic (1), German (1), Korean (1). 8 participants studied abroad in a Spanish-speaking country (Spain, Nicaragua, and Costa Rica), with the programs ranging from 2 weeks to a month in length. Most participants had studied Spanish since high school, with the mean overall length of study being 4.97 years. All participants (N= 90) self-identified their listening, reading, writing, and speaking skills in English with a 3 (out of 3). The average self-rating scores for their skills in Spanish for the overall L2 participant pool was as follows: listening (1.41), speaking (1.42), reading (1.75), writing (1.6).

The L2 participants’ biographical information is summarized in Table 6 below. Their self-rated proficiency and DELE scores are provided in Table 7.

For the HL group, participants were between the ages of 17 and 30 (average age 20.52, SD= 2.9). 31% (N=26) were born in the United States, whereas the rest were born in Central or South American countries, and emigrated to the US between the ages of 1 and 10 years old (median age or arrival: 3.30 years). The remaining participants’ country of origin included: Cuba (25), Venezuela (15), Colombia (9), Nicaragua (5), Peru (3), and Dominican Republic (2). The average length of residency in the United States was 17.23
years (ranging from 10-24 years). All the heritage participants identified Spanish as their native language, with 10 participants also identifying themselves as native speakers of English. 10 participants also acknowledged having experience learning a foreign language: Russian (1), Portuguese (1), Italian (2), French (2), Japanese (2), Mandarin (1), and German (1). All participants ($N = 85$) self-identified their listening, reading, writing, and speaking skills in English with a 3 (out of 3). The average self-rating scores for their skills in Spanish for the overall HL participant pool was as follows: listening (2.9), speaking (2.5), reading (2.3), writing (1.75).

The HL participants’ biographical information is summarized in Table 6 below. Their self-rated proficiency and DELE scores are provided in Table 7.

Table 6: Participant characteristics, overall and by experimental group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Age M(SD)</th>
<th>Gender M, F, Other</th>
<th>Mean Length Formal Spanish Study (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heritage Language (HL) Learners</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Explicit Feedback</td>
<td>20.65(2.6)</td>
<td>16, 21, 1</td>
<td>1.06 years (1.24)</td>
</tr>
<tr>
<td>-Explicit Feedback</td>
<td>20.21(3.0)</td>
<td>4, 14, 0</td>
<td>0.96 years (1.22)</td>
</tr>
<tr>
<td>Control Group</td>
<td>20.95(3.0)</td>
<td>11, 9, 0</td>
<td>1.91 years (1.45)</td>
</tr>
<tr>
<td>Maturational Control Group</td>
<td>20.11(3.9)</td>
<td>4, 5, 0</td>
<td>1.33 years (1.41)</td>
</tr>
<tr>
<td><strong>Second Language (L2) Learners</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Explicit Feedback</td>
<td>19.55(1.34)</td>
<td>12, 19, 0</td>
<td>5.68 years (2.41)</td>
</tr>
<tr>
<td>-Explicit Feedback</td>
<td>19.31(1.05)</td>
<td>10, 14, 1</td>
<td>4.89 years (1.66)</td>
</tr>
<tr>
<td>Control Group</td>
<td>19.29(1.33)</td>
<td>6, 17, 0</td>
<td>4.43 years (1.09)</td>
</tr>
<tr>
<td>Maturational Control Group</td>
<td>19.64(1.03)</td>
<td>1, 10, 0</td>
<td>4.40 years (1.01)</td>
</tr>
<tr>
<td><strong>OVERALL:</strong></td>
<td>19.96(2.15)</td>
<td>64, 109, 2</td>
<td>3.09 years (1.44)</td>
</tr>
</tbody>
</table>
Table 7: Participant self-rated proficiency in Spanish and diplomas of Spanish as a foreign language (DELE) score, overall and by experimental group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Listening</th>
<th>Oral</th>
<th>Reading</th>
<th>Writing</th>
<th>DELE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heritage Language (HL) Learners</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Explicit Feedback</td>
<td>2.48(0.51)</td>
<td>2.45(0.51)</td>
<td>2.23(0.56)</td>
<td>1.81(0.65)</td>
<td>30.71(9.64)</td>
</tr>
<tr>
<td>-Explicit Feedback</td>
<td>2.47(0.51)</td>
<td>2.42(0.51)</td>
<td>2.21(0.63)</td>
<td>1.68(0.67)</td>
<td>33.32(9.96)</td>
</tr>
<tr>
<td>Control Group</td>
<td>2.43(0.51)</td>
<td>2.35(0.49)</td>
<td>2.35(0.65)</td>
<td>1.61(0.66)</td>
<td>34.04(8.79)</td>
</tr>
<tr>
<td>Maturational (CG)</td>
<td>2.33(0.50)</td>
<td>2.25(0.46)</td>
<td>2.50(0.53)</td>
<td>1.33(0.50)</td>
<td>34.78(10.06)</td>
</tr>
<tr>
<td><strong>Second Language (L2) Learners</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Explicit Feedback</td>
<td>1.32(0.48)</td>
<td>1.42(0.67)</td>
<td>1.74(0.44)</td>
<td>1.58(0.72)</td>
<td>30.81(10.43)</td>
</tr>
<tr>
<td>-Explicit Feedback</td>
<td>1.31(0.53)</td>
<td>1.32(0.53)</td>
<td>1.83(0.38)</td>
<td>1.66(0.54)</td>
<td>33.66(9.33)</td>
</tr>
<tr>
<td>Control Group</td>
<td>1.64(0.57)</td>
<td>1.56(0.58)</td>
<td>1.56(0.51)</td>
<td>1.60(0.50)</td>
<td>31.64(9.11)</td>
</tr>
<tr>
<td>Maturational CG</td>
<td>1.41(0.53)</td>
<td>1.56(0.62)</td>
<td>1.55(0.50)</td>
<td>1.63(0.51)</td>
<td>30.55(8.79)</td>
</tr>
<tr>
<td>OVERALL:</td>
<td>1.92(0.52)</td>
<td>1.92(0.49)</td>
<td>2.06(0.37)</td>
<td>1.61(0.13)</td>
<td>31.06(3.76)</td>
</tr>
</tbody>
</table>

*Self-rated proficiency (scale 0-3); **DELE (out of 50 points)

**Linguistic Target Item**

Similar to Rosa and Leow (2004), the target structure for the current proposed study was the Spanish contrary to fact conditional sentences in the past. Conditional sentences may be categorized as real/factual and unreal/contrary to fact. Factual conditional sentences are those that may actually come about or at least are viewed as a possibility, whereas contrary to fact conditional sentences are those that will not come about or are viewed as being completely hypothetical. These sentences may refer to past, present, or future time frames. In this case, the ‘if’ clause is typically found in a past subjunctive tense, and a variation of the conditional tense was used for the main verb. Following Rosa and Leow (2004) and for the purpose of this study, the experimental treatment featured two different types of contrary to fact conditional sentences in Spanish:
1. those that result in the present or future with the main clause expressed in the conditional tense

“Si yo fuera rico, compraría una casa.”
[If I were rich, I would buy a house.]

where the imperfect subjunctive is found in the subordinate clause, and the conditional is used in the main clause, and

2. those that refer to the past with a result in the past tense

“Si hubiera estudiado más para el examen, no habría estado tan nerviosa.”
[If I were to have studied more for your exam, I wouldn’t be so nervous.]

where the pluperfect subjunctive is used in the subordinate clause, and the perfect conditional is used in the main clause.

These types of conditional sentences are considered to be highly complex and problematic structures for both native speakers and L2 learners of Spanish (e.g. Collentine, 2003; López Ornat, 1994) and are usually not taught in the first couple of years of language study (Rosa & Leow, 2004). According to Collentine (2003), this structure presents a complex challenge to learners of Spanish, as not only is this structure not found in English, but it is also not as salient (or widely-used) in Spanish.

Furthermore, Collentine identifies the subjunctive mood as a two-fold difficulty for Spanish learners, as it requires these learners to not only accurately employ a morphological system, but to also accurately use a complex syntactic system (particularly subordinate clauses), which is required for the subjunctive to be produced. Furthermore, the subjunctive morphemes are not as salient as other structures, and learners find that native speakers are able to comprehend them despite not producing the subjunctive
correctly. Additionally, as it is not as salient as other structures, there is usually not enough sufficiently rich input employing the subjunctive for the learners to be able to pay attention to it. Although Collentine (2003) was referencing the difficulties particularly for L2 learners of Spanish, this could be extrapolated to heritage language learners of Spanish as well.

Silva-Corvalán (1994) reported on the loss of the subjunctive mood in bilinguals (Spanish and English) from Los Angeles based on sociolinguistic interviews she conducted. These interviews exposed examples in the data of the subjunctive mood being replaced by the indicative mood in optional contexts (i.e. where the speaker’s perception and/or interpretation of the events guide the selection of the mood), but not necessarily in obligatory contexts. The data showed that these results were mostly applicable to heritage speakers of Spanish who attended school in the United States.

Due to its lack of saliency, contrary-to-fact conditional sentences, particularly those in the past, face a lot of variation in Hispanic communities. Montolío (1999) notes that this structure is usually seen, if at all, in textual forms (i.e. newspapers and literature), and when it is seen, it appears in its standard form (see the aforementioned examples). Colloquially, however, the structure takes on various manifestations:

1) Si+ [formal] pluperfect subjunctive + conditional

   _Si hubiese estudiado más para el examen, no estarías tan nerviosa._

2) Si+ pluperfect subjunctive + pluperfect indicative

   _Si hubiera estudiado más para el examen, no había estado tan nerviosa._

3) Si+ pluperfect indicative+ pluperfect indicative

   _Si había estudiado para el examen, no había estado tan nerviosa._
4) Si+ present indicative + present indicative.

Si estudio más para el examen, estoy nerviosa.

The semi-spontaneous picture description task utilized in this study also provided a glimpse into how this structure is expressed, specifically in the Miami, Florida population:

5) Si+ pluperfect subjunctive + pluperfect subjunctive

Si hubiera estudiado más para el examen, no hubiera estado tan nerviosa.

6) *Si+ present indicative + ir + a+ infinitive

Si estudio más para el examen, no voy a estar tan nerviosa

7) *Si+ present indicative + future

Si estudio más para el examen, no estaré tan nerviosa.

8) Si+ conditional + conditional

Si estudiaría más para el examen, no estaría tan nerviosa.

9) *Si+ imperfect subjunctive + conditional

Si estudiara más para el examen, no estaría tan nerviosa.

The sentences with an asterisk, although grammatically correct, are not semantically equivalent to the proposed structure.

Materials

Explicit Review Session

During the first session, prior to the pretest and before starting the treatment, all participants were presented with a review of the contrary to fact conditional sentences with a result in the present (the distractor item) in a classroom setting. Participants were asked to view the PowerPoint presentation providing explicit rules and input of this
structure, as well as its uses. The tenses reviewed were the imperfect subjunctive in the main clause and the conditional in the subordinate clause. After the review, participants were asked to conjugate the verbs for the contrary to fact conditional sentences within the PowerPoint, and were provided with the correct answers so that they could check their progress (see Appendix C).

**Think-Aloud (TA) Practice**

A TA practice task was provided to participants prior to the experimental treatment in the second session as a warm-up to practice thinking aloud during the procedure. A small math word problem was utilized, as previous studies have mentioned that it facilitates thinking aloud (Bowles, 2010). The practice task asked participants to tally up their bill at a local café, and to include tax and tip (see Appendix F).

**Treatment**

Participants were randomly assigned to one of three experimental conditions: +Explicit Feedback, -Explicit Feedback, or Control Group. Participants in the +/- explicit feedback groups were asked to think aloud throughout the treatment, in the language in which they felt most comfortable. Although there was no time limit, participants’ time on task was recorded. Both conditions received the following instructions:

- This activity consists of 2 short stories that you must help complete.

- You will first read a brief paragraph that will introduce you to the scenario. Please keep this scenario in mind as you continue the task.

- You will then see a screen with a sentence. You will notice that one of the two parts of each sentence is missing. At the bottom of the screen, you have four sentence fragments. ONLY ONE OF THOSE FRAGMENTS IS APPROPRIATE TO COMPLETE THE SENTENCE. Your task is to try to complete each sentence by selecting the sentence fragment that you think best completes it.

- You will know whether you have completed the sentence correctly based on the feedback you receive. If the feedback screen informs you that you have made the
wrong choice, then you choose another answer until you receive the feedback that your choice was the right one.

In this experiment we are also interested in what you think about as you complete this task. In order to find out, we are going to ask you to THINK ALOUD from the time you start the task to when you finish the task. You may choose to think aloud in the language in which you feel most comfortable. We would like you to talk CONSTANTLY. We don't want you to try to plan out what you say or try to explain to us what you're saying. Just act as if you are alone in the room speaking to yourself. What's most important is that you keep talking, and talk clearly and loudly enough into your microphone. We will not be able to help you in any way.

Press any key to begin. ¡Buena suerte! :

In order to be conscientious of the sociolinguistic factors that play a role with heritage learner research, and additional paragraph was added to the instruction introduction page for the HL participants:

What you are about to see is a grammatical lesson on how to formulate a specific structure in standard Spanish. There are many varieties of Spanish, and therefore, there are many different ways of saying the same thing (think about English- you could order a soft drink, a soda, or a pop, and they're all correct). If you have a different way of expressing this thought, it doesn’t necessarily mean that your way is wrong. This is just an alternative way in which to do so.

The experimental groups were as follows:

+Explicit Feedback: Participants were provided with a brief scenario in English, which served as an introduction to the sentences in the experiment. This was done to consistently remind the students of the timing of the story and the subsequent sentences, as the timing (either past or present/future is a primary factor in determining the conjugation of the main clause). After reading the scenario, participants were presented a computerized task via SurveyGizmo.com, in which they were shown a large sentence in the middle of the screen, with a blank line and a sentence fragment following it. 4 sentence fragments were listed below it. The first blank represented the main clause of the sentence. Participants were asked to continue the story from the scenario by selecting the appropriate main clause that fits the subordinate clause. If a participant selected the
incorrect answer, they received a message saying that the answer was incorrect, and to consider whether the sentence was in the present or past tense (see Figure 2 below). They were then redirected to the previous screen to try again. Participants were provided with an unlimited number of tries until they reached the correct answer, in which they then were presented with a message saying that they chose the right answer, and provided an explicit grammatical rule (see Figure 3 below). There were 2 scenarios, each with 20 sentences - 10 distractors (using the imperfect subjunctive, as well as the preterit and imperfective indicative) and 10 target items (the pluperfect subjunctive).

**Treatment 1**

Muy bien! The condition expressed in this sentence refers to the present or future. In sentences like these, you use the imperfect subjunctive— which is formed by taking the 3rd person plural [the ellos form] preterit [the past tense] form of the verb, and replacing the -ron with -ra.

Please press the **NEXT** button.

![Figure 2: Correct answer +explicit feedback screen.](image)
Within the HL +Explicit group, one additional group was created without the requirement to think aloud (HLMEF) in order to serve as a control condition to ascertain whether or not there were reactivity effects, and to determine to what extent they played a role.

**Explicit Feedback:** Participants in this treatment were presented with the same task as the +explicit group, with the exception of the type of feedback. If they selected the incorrect answer, they were presented with a screen stating that their answer was wrong (“Oops! Try again!”), and then redirected to the previous screen to try again. When they chose the correct answer, they were only told their answer was correct (see Figures 4 and 5 below).
Control Group: The control group was provided with the same materials as the +/- explicit feedback groups, with the exception that the distractor and target items were already included in the story. Furthermore, participants in the control group were asked to
read the two short stories, and answer comprehension questions on a sheet of paper (see Appendix H). As the control group task was a reading comprehension task, participants were not asked to think aloud.

**Maturational Control Group:** Since the first and last sessions took place during the participants’ Spanish classes, there were some participants who completed both the first and last sessions, but did not participate in the experiment during the second session. To this end, these participants were categorized as the maturational control group, meaning a control group that was not exposed to the target grammatical structure, and therefore control for the potential of the participants having learned the target structure over time without exposure or instruction (Leow, 2015).

**Table 8: Overview of experimental groups.**

<table>
<thead>
<tr>
<th>Type of Group</th>
<th>Participant</th>
<th>Feedback</th>
<th>Think Aloud?</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL+EF+TA</td>
<td>Heritage Learner</td>
<td>+Explicit</td>
<td>Yes</td>
</tr>
<tr>
<td>HL-EF+TA</td>
<td>Heritage Learner</td>
<td>-Explicit</td>
<td>Yes</td>
</tr>
<tr>
<td>HL+EF-TA</td>
<td>Heritage Learner</td>
<td>+Explicit</td>
<td>No</td>
</tr>
<tr>
<td>HLC</td>
<td>Heritage Learner (Control Group)</td>
<td>No Feedback</td>
<td>No</td>
</tr>
<tr>
<td>HLMC</td>
<td>Heritage Learner (Maturational Control Group)</td>
<td>No Feedback</td>
<td>No</td>
</tr>
<tr>
<td>L2+EF+TA</td>
<td>L2 Learner</td>
<td>+Explicit</td>
<td>Yes</td>
</tr>
<tr>
<td>LE-EF+TA</td>
<td>L2 Learner</td>
<td>-Explicit</td>
<td>Yes</td>
</tr>
<tr>
<td>L2C</td>
<td>L2 Learner (Control Group)</td>
<td>No Feedback</td>
<td>No</td>
</tr>
<tr>
<td>L2MC</td>
<td>L2 Learner (Maturational Control Group)</td>
<td>No Feedback</td>
<td>No</td>
</tr>
</tbody>
</table>
Assessments

In order to establish that the target grammatical structure was learned, participants underwent a battery of assessments conducted in a pretest, immediate posttest, and delayed posttest design. The pretests also served to determine the participants’ eligibility to participate in the experiment, as well as to aid in cleansing the data. All of the tests were pen-and-paper, and did not carry a time limit.

Controlled production task (CPT): A fill-in-the-blank test with contrary-to-fact conditional sentences served as the controlled production task. All three of the production tests (pretest, immediate posttest, and delayed posttest) were comprised of the similar scenarios used in the treatment. There were 6 different scenarios in total created, and were randomly assigned to the participants throughout the three sessions to account for any testing effects. The CPT consisted of two scenarios, 40 sentences total: 20 distractors, which included the imperfect subjunctive, imperfect indicative, and preterit indicative; and 20 critical items, which consisted of the pluperfect subjunctive (see Appendix D). While the missing verb was always in the main clause, the order of the main clause was randomized throughout the scenario (when grammatically allowed).

Semi-spontaneous picture description task (SSPDT): As discussed in Chapter 1, L2 learners tend to outperform HL learners in explicit, metalinguistic tasks. To counterbalance the explicit nature of the CPT, and to descriptively view what both linguistic profiles of learners produce when creating contrary-to-fact conditional sentences in the past, a semi-spontaneous picture description task was created. The task was administered to all participants immediately following the CPT in all three sessions.
The task consisted of ten pairs of pictures (taken by the researcher), where each pair represented an ‘if, then’ clause. Below the pictures, blank lines were provided for participants to write a sentence describing what was occurring in the two pictures. In order to facilitate this, participants were also provided a ‘si’ (if) in either the beginning or middle of the lines to aid them in correctly placing the main clause. Additionally, to control the answers that could be provided as much as possible, two verbs in the infinitive were provided in parentheses under the pictures (see Figure 6). The instructions (below) included a short scenario in English to inform the participants that all of these pictures and scenarios took place in the past. Similar to the CPT, three different versions of the task were created to control for test effects.

In the following task, you will see two pictures side by side. You will be asked to write one sentence describing the two pictures you see, using the verb in the infinitive in parenthesis as a guide to create your sentence. Your sentences must include ‘si’ (either in the first or second clause). Try to be as specific and detailed as possible in your sentence.

Scenario: Melissa transferred from Florida International University (FIU) in Miami, Florida to Georgetown University in Washington, D.C. two months ago. She initially had a difficult time assimilating to her new life, and now that she has, she is reflecting on how some tough moments could have been different during her first two months.
Diploma of Spanish as a Foreign Language (DELE): In line with previous empirical research in the heritage speaker strand (e.g. Montrul, 2004, 2005; Montrul & Bowles, 2010; Montrul & Foote, 2014), the proficiency levels of L2 learners and HS affect the likelihood of finding any correlation or pattern in the study. However, as mentioned previously in this study and corroborated by Valdés (2000), categorizing HS by utilizing a proficiency assessment that taps primarily into explicit and metalinguistic information developed for L2 learners (i.e. the DELE) is difficult to do. Moreover, other assessments, such as the Oral Proficiency Interview (OPI) provided by ACTFL has shown that HS also do not necessarily fit the proficiency categories predetermined for the L2 learners (Montrul, Foote, & Perpiñan, 2008; Potowski et al., 2009). Despite the fact
that there currently does not exist a consensus as to how to measure the proficiency for this heterogeneous population, in order to efficiently compare these learners, and in order to be able to control for as many variables as possible, there needs to be a basic measure to balance the groups’ proficiency levels from the start in order to assess any correlation that may or may not emerge (Montrul et al., 2008).

As a result, all of the participants of the current study completed the DELE that has been used in the majority of previous HS empirical studies (e.g. Montrul et al., 2008; Montrul & Bowles, 2010; Montrul & Perpiñan, 2011; Montrul & Foote, 2014) (see Appendix I). However, a pretest to assess the current knowledge of the target linguistic structure, as well as a comprehensive language background questionnaire (which included a self-rating section) also accompanied this proficiency assessment to triangulate the data as much as possible (see Appendix E).

Upon completion of the proficiency exam, a Cronbach’s Alpha test was conducted to assess the reliability of the instrument. Then an independent sample t-test was conducted in order to compare whether or not the experimental groups shared similar proficiency levels (Montrul & Perpiñan, 2011).

**Questionnaires**

**Language Background Questionnaire:** An adapted version of Montrul’s5 (2012) language background questionnaire was utilized in order to take into account each participant’s particular history with the Spanish language. The adapted version used in this study included questions regarding country of birth, residency, first language(s), experiences with study and/or living abroad, self-rating in English and Spanish (in

5 [http://www.nhlrc.ucla.edu/data/questionnaires.asp](http://www.nhlrc.ucla.edu/data/questionnaires.asp)
reading, writing, listening comprehension, and speaking on a 4-point scale), and years spent studying Spanish. A different version adapted for heritage Spanish speakers was also created, which further delved more into identifying the different contexts in which Spanish was/is used and was acquired (see Appendix A and B).

**Debriefing Questionnaire:** Prior to the delayed posttests, a debriefing questionnaire was administered to the participants in order to assure as much as possible that prior knowledge, recognition of the target items, and/or other outside sources did not contribute to contamination of the data (see Appendix L). The questionnaire was adapted from Calderón (2014), and was used in Leow (1993, 1995). It detailed both target linguistic item as well as the distractor item and its uses, and asked the participants if they 1) recognized, 2) knew of, or 3) had any additional external exposure to these items during the experimental treatment. If participants indicated ‘yes’ to any of the aforementioned questions, then that participant’s data was further investigated to decide whether or not it should be excluded from the analysis. The debriefing questionnaire in the current study did not result in any participants being excluded from the analysis.

**Equipment**

The equipment used for this experiment included (1) Mac and PC computers in the computer labs with internet access, (2) the SurveyGizmo website (SurveyGizmo.com), (3) Microsoft PowerPoint (for the in-class explicit instruction), (4) Quicktime and Windows Media, in order to record the TA protocols, (5) Quicktime to transcribe the TA protocols.

**Procedure**

This study followed a pretest (week 1)–immediate posttest (week 2)–delayed
posttest (week 4) design (see Figure 7). In week 1, participants reported to their respective language classes, where they were presented with an explicit review of the contrary-to-fact conditional sentence in the present (the imperfect subjunctive distractor) by the researcher. They were then administered the language background questionnaire and the pretests (i.e. the CPT and SSPDT).

During week 2, participants were randomly assigned to one of three experimental conditions: + Explicit Feedback (+EF), –Explicit Feedback (-EF), or the control group. They were then asked to complete the think aloud warm up exercise (sans control group), after which they completed the treatment accordingly. Immediately after completing the experimental task, the two assessment tests were administered: the CPT and the SSPDT.

During the last session, approximately two weeks later, the researcher returned to the language classes to complete the last round of assessments (CPT and SSPDT), the DELE, and an exit debriefing questionnaire. Since the researcher needed to travel to Miami, Florida to collect the HL data, the HL study was conducted in October, and the L2 study was conducted in November.
Coding and Scoring

All of the assessment tasks (pretests, posttests, and delayed posttests) were scored and the concurrent verbal reports were transcribed, coded for levels of processing, and later scored. Although the language background and debriefing questionnaires were not coded per se, the former provided the researcher with data on the participants’ history with Spanish (and other Romance languages), and the latter served to confirm the presence or absence of prior knowledge of the target structure. A more detailed explanation of each of the coding processes follows.

Assessments

**Controlled production test**

One point was awarded for every correct item (both target linguistic and distractor item). As this study was not measuring accuracy, no points were deducted for
orthography or incorrect conjugation involving person (i.e. if the correct answer was *hubiera escrito*, a point was still awarded if the participant wrote *hubiera escrito*, *hubieramos escrito*, or some similar production). However, if the participant had instead answered *habria escrito*, it would not be scored as correct). For the distractor items, priority was given to whether or not participants could differentiate between the imperfect subjunctive and the other indicative distractors. In other words, if the correct answer was the preterit indicative *hice* and the participant wrote *hacía* in the imperfect indicative, no points were deducted. However, if the correct answer was *hice* in the preterit indicative, and the participant wrote the imperfect subjunctive, *hiciera* (or the pluperfect subjunctive *hubiera hecho*), then it would be scored incorrect. A maximum of 40 (20 target and 20 distractor) points were allotted to the CPT tests. As the study was mostly interested in the target items (pluperfect subjunctive), the statistical analyses were conducted using the target item scores (X out of 20); however, the overall scores (including the distractor items) were used in order to ascertain that there were no instances of overgeneralization (i.e. a participant writing the pluperfect subjunctive for all the responses, thereby getting a ‘perfect score’ on the target items, but a 0 on the distractor items).

The controlled production pretest served to confirm that a) that participants had no prior knowledge of the target linguistic structure, and were thus novice learners, and b) participants understood the contrary-to-fact conditional sentences resulting in the present/future (using the imperfect subjunctive) that was explicitly presented in the first session, as that was the distractor item, and crucial in measuring their ability to differentiate that from the pluperfect subjunctive in the experimental treatment. For L2
participants, a score of 0/20 on the target structure was required to qualify for the study. For HL participants, in order to account for their diverse experience with language, as well as prior knowledge, a score of $x < 5/20$ on the target structure was required (as per the mean scores of all of the HL participants). For both linguistic profiles, a minimum of a 14/20 on the distractor items was required in order to qualify for the study. Again, this number resulted from taking the averages of all the participants’ scores, and setting a ‘cutoff’ for any participant scoring below the average line.

**Semi-Spontaneous Picture Description Task**

Each individual clause (the main clause and the subordinate clause) was scored 1 point, for a total of 2 points per sentence (and 20 points in total). For the main clause, where the pluperfect subjunctive was expected, 0 was assigned to indicative forms, 0.5 for present or imperfect subjunctive forms, and 1 point for the target pluperfect form. This coding scheme was adapted from Baralt (2013), to be sensitive to developmental language production, and variations in HL Spanish. (see Norris & Ortega, 2009, for arguments for coding schema that attempts to address the developmental paths in L2 acquisition). For the subordinate clause, 1 point was assigned to the standard grammatical form presented (perfect conditional), and 0 points were scored for other forms.

**TA Protocols and Depth of Processing Coding**

Think-aloud protocols from both conditions were transcribed by the researcher, and were then separated into ‘episodes’. Episodes consisted of a verbalization that encompassed a complete thought, and usually coordinated with an individual question (in the experiment). However, if there was a complete stop with an elongated unfilled pause, the next verbalization would be identified as a new episode, despite there already existing
an episode in that same question. Level of processing was coded in accordance with the
criteria adapted from Leow (2015) in Table 9. Transcriptions were coded within one of
three categories: low, medium, or high level of processing. A low level of processing was
labeled as observations, which hold no potential for processing the grammatical target
form. An instance of this is seen below from Participant 1, where the participant states
that he is unsure of the grammatical item, and uses a low level of cognitive effort to
continue with the task:

- Since I think we’re going on a pattern, I’m going to with C [sería] but I’m not sure. No? So I’m going to go with A [fuera]. Ok, I have no idea.

[reads sentence in Spanish]. D? [habría sido]. No. Ok so it’s B [hubiera sido].

| Table 9: Operationalization of depth of processing (DOP) grammatical items. (Leow, 2015) |
|--------------------------------------|--------------------------------------|--------------------------------------|
| **Level of Awareness**               | **Low Depth of Processing**          | **Medium Depth of Processing**       |
|                                      | **Level 1**                          | **Level 2**                          |
| Noticing                             | Reporting                            | Comments on target item in relation to grammatical features |
| Shows no potential for processing the target form grammatically |                              | Arrives at an inaccurate, partially accurate, or fully accurate target underlying grammatical rule |
| **Descriptor**                       |                                      |                                      |
| - Reads target quickly               | - Comments on target item in relation to grammatical features |
| - Translates the phrase to English but leave the target in Spanish | - Spends more time processing target item |
| - Carefully pronounces target item    | - Makes comments that indicate some processing of target item. |
| - Repeats target item                | - Some level of cognitive effort to process target item grammatically |
| - Says s/he isn’t sure what it is    | - **Systematic metalinguistic hypothesis testing** |
| - Does not spend much time processing target item |                              |                                      |
| - Low level of cognitive             |                                      |                                      |
|                                      |                                      |                                      |
|                                      |                                      |                                      |
|                                      |                                      |                                      |

102
A verbalization identified as a medium level of processing is characterized by comments on the target item in relation to its grammatical features. Below is an example from Participant 11, where although she is spending more time and effort processing the target item, and there is the prospect of a rule formation, no hypotheses or actual rule formations are shown:

- Okay, so if Anita had not eaten the last piece of chocolate… had not eaten the last piece of chocolate, so it’s definitely had the last piece of chocolate. It should be A, comiera. Because it’s… is it right? Nope. So it’s… oh right because the other half is in the conditional, so C [comería] No. It’s still wrong. Is it B [hubiera comido]? Guess so. Weird.

A high depth of processing commentary is categorized by a participant arriving at an inaccurate, partially, or fully accurate target underlying grammatical rule, or putting forth a maximal cognitive effort. In this adapted criteria, the researcher omitted the descriptor “spends much time processing target item” found in the original operationalization (Leow, 2015), as differentiating “spends a bit more time processing target item” in the medium level from “spends much time processing target item” in the high level proved to be difficult, and no set length of processing time was found. Therefore, the paramount difference came in actually showing greater cognitive effort manifested in attempts to come to an understanding of a rule (as per levels of awareness). However, unlike in the operationalization of levels of awareness, the rule formulation does not have to be accurate. An example of a high level of processing is seen in Participant 15 below:
• If Esteban would have had patience… if he had patience… he wouldn't be as impulsive. Hm, ok, so like the first one. If he had more patience, he wouldn’t be as impulsive, because it’s just like the activity that we did in the last session. He would be more impulsive. Good, so I was right. Because the conditional tense which is expressed in the second clause shows that this action may be changed, but still is improbable. It could be in the future.

In order to answer RQ4, which addressed any potential correlation between the participants’ levels of processing and subsequent performance on the assessment tasks, scores were assigned to each instance of a verbalization within each participant’s TA protocols. A low level of processing was assigned a score of 1, a medium level was assigned a 3, and a high level was assigned a 5.

More specifically, after having assigned a numeric ‘grade’ to each verbalization, the sums of the levels of processing achieved were then divided by the total amount of instances to result in an average processing score. This tally was then placed within one of three numeric ranges that corresponded with an overall score of depth of processing. In order to maintain some distance between levels to act as a buffer, any score not falling within the categorical ranges were eliminated for this particular quantitative analysis. The score chart is outlined in Table 10 below.

<table>
<thead>
<tr>
<th>Average TA DOP score</th>
<th>Overall DOP level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00-2.00</td>
<td>Low</td>
</tr>
<tr>
<td>2.50-3.50</td>
<td>Medium</td>
</tr>
<tr>
<td>4.00-5.00</td>
<td>High</td>
</tr>
</tbody>
</table>
Changes to Leow’s (2015) DOP chart

During the pilot study, one difficulty in addressing levels of processing scores as quantitative was that usually, any episode that wasn’t considered either “medium” or “high” processing would automatically default to “low”. This caused inaccurate tallies and averages, as not all “low” processing scores were actually accurate representations of a “low” level of processing. To remedy this, this study adapted Leow’s (2015) Levels of Processing chart to include a category titled N/A, which includes instances of the participant not processing at all. For instance, Participant 77:


Unlike the previous example of a participant processing at a low level, this particular participant is not putting forth any cognitive effort worth rating. She is merely making guesses as to which answer would be correct for that question. No translating to English took place, the target item was not repeated, and ‘does not spend much time processing target item’ appears to be more closely relayed to ‘spends no time processing target item’. This is not to say that a participant performing at this level (or lack thereof) of processing cannot demonstrate low, medium, and even high levels of processing elsewhere in his/her TA. However, for the sake of trying to tabulate a quantitative approach to processing in order to find correlations between levels of processing and subsequent performance on assessment tasks, this needed to be remedied. As such, the adapted processing chart is seen below (see Table 11).
Table 11: Adapted operationalization of DOP grammatical items. (Leow, 2015)

<table>
<thead>
<tr>
<th>LEVEL OF AWARENESS</th>
<th>NO LEVEL OF DOP</th>
<th>LOW DOP</th>
<th>MEDIUM DOP</th>
<th>HIGH DOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>POINTS ALLOTTED</td>
<td>LEVEL 0</td>
<td>LEVEL 1</td>
<td>LEVEL 2</td>
<td>LEVEL 3</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>DESCRIPTION</td>
<td>Shows absolutely no possibility of processing target form</td>
<td>Shows little to no potential for processing the target form grammatically</td>
<td>Comments on target item in relation to grammatical features</td>
<td>Arrives at an inaccurate, partially accurate, or fully accurate target underlying grammatical rule</td>
</tr>
<tr>
<td>DESCRIPTOR</td>
<td>-Only identifies answers provided. -Quickly goes from one option to the next when answer is incorrect. -Spends no time processing target item -Doesn’t attempt to talk about target item, or thoughts -Says s/he isn’t sure or “I don’t know” -Speeds through answers/target/reading.</td>
<td>-Reads target quickly -Translates the phrase to English but leave the target in Spanish -Carefully pronounces target item -Repeats target item -Says s/he isn’t sure what it is (regarding the target structure) -Does not spend much time processing target item -Low level of cognitive effort to process target item grammatically</td>
<td>-Comments on target item in relation to grammatical features -Spends more time processing target item -Makes comments that indicate some processing of target item. -Some level of cognitive effort to process target item grammatically -Systematic metalinguistic hypothesis testing</td>
<td>-Arrives at an inaccurate, partially accurate, or fully accurate target underlying grammatical rule -Makes hypotheses regarding target item -Provides an inaccurate, accurate, and/or partially accurate rule -Corrects previous translation -High level of cognitive effort to process target item grammatically</td>
</tr>
</tbody>
</table>

The researcher coded the verbalizations and a colleague trained in levels of processing codification separately coded 20% of the protocols in order to establish interrater reliability (IR). Initial IR agreement calculated highly at 98%, which was
calculated by dividing the number of episodes transcribed and coded identically over the total number of items transcribed and coded by both researchers. The researcher and colleague met to discuss the cases of disagreement (4 instances), which were all resolved by acquiescing on what should be considered as a verbalization of a grammatical rule. Additionally, to further ensure the reliability of the transcriptions and the coding, 20% of the TA protocols were randomly selected two months after the original transcription. They were re-transcribed and re-coded by the original researcher. The intrarater agreement was determined via the same procedure for the IR agreement. The results showed 100% intrarater agreement.
CHAPTER 6: RESULTS

This chapter presents the results of all the qualitative and quantitative analyses necessary to answer the research questions. The chapter begins with a brief overview of the series of quantitative statistical analyses performed both within and between linguistic profiles utilized to analyze the data, as well as a rationale as to why certain tests were used. Additional statistical information, such as descriptive statistics, are included to ensure validity when applicable. The chapter then continues with the results of each research question, in numerical order. Lastly, the chapter concludes with final synopses summarizing the results.

Statistical Analyses

Descriptive Statistics

Descriptive statistics were calculated for each groups’ (linguistic profile) pretest, immediate posttest, and delayed posttests performance. Means, standard deviations, and confidence intervals were computed and included in the principal chart in order to examine the predominant tendency, variability, and distribution of scores for each assessment, based on both independent groups (i.e. +/- explicit condition and linguistic profile). The pretest mean scores were calculated in order to have a baseline as to what should be the cutoff score in the performance of the distractor item (imperfect subjunctive) in order to qualify for the study (in both L2 and HL participants), as well as to serve as a cutoff score for HL participants in how much prior knowledge of the target item should be accepted in order to qualify for the study. Additionally, to ensure comparability between participants in the four experimental groups (+Explicit /HL, -Explicit/HL, +Explicit/L2, -Explicit/L2), a one-way ANOVA was run for both
assessment tasks (CPT and SSPDT), as well as for the DELE.

**Effect sizes**

Effect sizes are “indices of practical significance, which supplement information from traditional statistical significance tests” (Plonsky & Oswald, 2014, p. 2). Recent trends in the SLA field have led to researchers critiquing the sole use of *p* values in L2 research (see Larson-Hall & Plonsky, 2015; Nassaji, 2012; Norris & Ortega, 2006; Oswald & Plonsky, 2010; Plonsky, 2009, 2011 for a full review on the controversy), and have instead advocated for the primary use of descriptive statistics, *d* value (which expresses the mean difference in terms of its standard deviation “across groups or within a reference group across two points in time” (Plonsky & Oswald, 2014, p. 3)), and Confidence Intervals, which tend to be more meaningful, and allow for even nonsignificant findings to contribute advantageous information. Additionally, the use of *d* values will allow for the data and results of this dissertation be utilized in future meta-analyses, and for it to be compared to other studies following similar suit.

**Repeated Measures Analysis of Variance (ANOVA)**

Repeated-measures ANOVA were used to primarily address the second research question on the interaction between Time (pretest, immediate posttest, and delayed posttest performance scores), Linguistic Profile (L2 learner and HL learner), and Experimental Condition (+Explicit Feedback, -Explicit Feedback, Control Group, Maturational Control Group); where Time was the within-subject factor while LP and EP were the between-subject factors. Separate ANOVAs were run in order to assess the effects of these variables on the two types of tasks (CPT and SSPDT), independently.
Correlations (r)

Correlations (r) were run in order to answer the third research question, which sought to find relationships between the levels of processing (DOP) and the subsequent performances on the immediate and delayed posttests. These correlations were run in two different ways – one comparing the levels of DOP on the test evaluations between HL and L2 (separately), and another which did the same, but separated the groups into HL + explicit, HL – explicit, L2 + explicit, L2 – explicit. Participants in the HL + explicit group that did not think aloud were not part of these correlations. Plonsky and Oswald’s (2014) recommendations were used to categorize these correlations into one of three levels: 1) small (\( r = 0.25 \)), 2) medium (\( r = 0.40 \)), and 3) large (\( r = 0.60 \)).

Programs

Programs utilized for these analyses were Microsoft Excel (version 15.291) for the Macbook Pro, the Statistical Package for the Social Sciences (SPSS), version 24, and the website Langtest.jp, which hosts an R program to run effect sizes and correlations, by Dr. Atsushi Mizumoto, an Associate Professor of Applied Linguistics at Kansai University in Osaka, Japan.

Reactivity

In order to ascertain whether or not there were reactivity effects, and to determine to what extent these effects played a role, the performance of the +explicit feedback TA group was compared with the performance of the +explicit feedback – TA group on both the Controlled Production Task (CPT) and the Semi-Spontaneous Picture Description Task (SSPDT). A One-Way ANOVA was run, and it revealed no statistically significant difference between the groups in either the CPT (\( F(1, 36) = 0.052, p = 0.822 \)) and the
SSPDT ($F(1, 36) = 0.026, p = 0.953$). Therefore, the data from the –TA group were included in the data for analysis for Research Question 2. The descriptive statistics, as well as the results of the ANOVA, are presented below in Tables 12 and 13.

**Table 12: Repeated measures analysis of variance (ANOVA) results of reactivity.**

<table>
<thead>
<tr>
<th>Source</th>
<th>$Df$</th>
<th>Mean square</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT</td>
<td>1</td>
<td>0.237</td>
<td>0.052</td>
<td>0.822</td>
</tr>
<tr>
<td>SSPDT</td>
<td>1</td>
<td>0.026</td>
<td>0.003</td>
<td>0.953</td>
</tr>
</tbody>
</table>

*p<0.01

**Table 13: Descriptive statistics for results of reactivity.**

<table>
<thead>
<tr>
<th>Source</th>
<th>$N$</th>
<th>$M$ (SD)</th>
<th>[CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT</td>
<td>19</td>
<td>18.53 (1.611)</td>
<td>[17.75, 19.30]</td>
</tr>
<tr>
<td>SSPDT</td>
<td>19</td>
<td>7.10 (2.811)</td>
<td>[5.75, 8.46]</td>
</tr>
</tbody>
</table>

*M= mean, SD= standard deviation, CI= confidence interval

**Proficiency levels**

An independent sample t-test was run to compare the HL and L2 participant groups’ proficiency levels. Tables 14 and 15 below shows the results. Based on these results, no statistically significant difference was found between the proficiency levels of L2 ($M= 32.14, SD= 9.519$) and HL ($M= 32.18, SD= 9.627$) participants, as measured by the DELE ($t= -0.022, p= 0.921, CI [-2.889, 2.825]$).

**Table 14: Descriptive statistics for results of DELE.**

<table>
<thead>
<tr>
<th>LINGUISTIC PROFILE</th>
<th>N</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2</td>
<td>90</td>
<td>32.14</td>
<td>9.519</td>
</tr>
<tr>
<td>HL</td>
<td>85</td>
<td>32.18</td>
<td>9.627</td>
</tr>
</tbody>
</table>
Table 15: Independent t-test results of DELE.

<table>
<thead>
<tr>
<th>t</th>
<th>Mean difference</th>
<th>P</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DELE</td>
<td>-0.022</td>
<td>-0.032</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[-2.889, 2.825]</td>
</tr>
</tbody>
</table>

*p<0.01

Research Question 1

How do heritage language learners of Spanish process incoming data in their heritage language? Is this process similar or different to L2 learners of Spanish?

In order to determine how heritage language learners processed data in their heritage language, a qualitative approach was taken, which consisted of utilizing the Grounded Theory (Glaser & Strauss, 1967), whereby pattern models of explanations of the data emerge. Reason (1981) explains,

The information that is gathered in the field situation is used by the holist to build a model which serves both to describe and explain the system. The model is built by [quoting Diesing (1972)] “connecting themes in a network or pattern” (p. 155); the connections may be of various kinds, but they are “discovered empirically rather than inferred logically” (p. 156); the result of this is an empirical account of the whole system. This account explains the system because it describes the kinds of relations the various parts have for each other, so that the “relationship between that part and the other parts serve to explain or interpret the meaning of that part” (p. 158). This type of explanation is called a pattern model of explanation. (p. 185-186)

After transcribing the groups who were required to think aloud, and after cleansing the TA data of those participants (HL: N= 3; L2: N= 6) who did not follow instructions (e.g. they spoke too low and thus were unintelligible, they did not think aloud consistently
throughout the experiment), there were a total of 84 participants whose TA were included in this analysis (HL: \( N= 34 \); L2: \( N= 50 \)). Transcriptions were analyzed and, once patterns were found, they were classified into the following categories (see Table 16)

Table 16: Think-Aloud categories for analyzing qualitative processing research question (RQ) 1.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>SUB-CATEGORY</th>
<th>DESCRIPTORS</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>METALINGUISTIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Grammar</td>
<td>Makes mention of</td>
<td>“I think it’s C because it would be if he was not… well I don’t remember. Because uh tuvieron is preterit third person plural… not sure what tirar is. But habria is in that tense so the answer should be… comiera.” (Participant 1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>explicit grammatical</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>structures, grammar</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior Knowledge</td>
<td>Alludes to previous</td>
<td>“Oh well I knew that since it was conditional in the second half, well, it needed to be imperfect subjunctive in the beginning, so it was A.” (Participant 90)</td>
<td></td>
</tr>
<tr>
<td>(Learned)</td>
<td>grammatical instruction, either from the experimental pretest lesson, or prior formal instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison Grammar</td>
<td>Compares current</td>
<td>“So if Esteban had returned home with the cake. Oh! It’s in the past! That’s why. It’s in the past. Ok ok. So it’s preterit, and it’s pluperfect now. Ok. … so it’s double past. Just like when I learned pluperfect. Okay, that makes sense. That makes sense.” (Participant 5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>input grammar to</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>grammar of other</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>languages</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“Ok, so in this one… it’s habria tirado. Hm. I wonder if it has to match haber plus past participle in the two sentences. So not at all like what we learned with the imperfect subjunctive last time. So like if it’s habria tirado, it would need to be… hubiera comido? I guess that makes sense. Let’s try it out… yes!” (Participant 13)</td>
</tr>
</tbody>
</table>

“I wonder if this is like Italian abbia parlato, which would be… había hablado. Hm. I don’t think so.” (Participant 75)

“Starting to think this is like what I learned in Portuguese with tivesse like that super complicated form. Was it in the past? Or the future? Should’ve take Spanish instead of Portuguese.” (Participant 134)
Table 16: (Cont.)

NON-METALINGUISTIC

<table>
<thead>
<tr>
<th>Classification</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intuition</strong></td>
<td>Makes references to participant’s ‘gut instinct’ and/or feelings of the validity of the grammatical structure</td>
<td>“That just doesn’t sound right to me. I can’t put my finger on it. But it sounds really, really weird and like it just can’t be it. I don’t know why.” (Participant 172)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Don’t remember if patience should go with haber or tener, but my gut is telling me to go with um tener. But potentially, if I look at all of these different verb conjugations, it could guide me to the correct answer, too.” (Participant 19)</td>
</tr>
<tr>
<td><strong>Prior Knowledge (Acquired)</strong></td>
<td>Mentions previous instances where participant heard or was exposed to the target structure</td>
<td>“Wait… I know this. Isn’t this that super fancy word that they use in the bible? ‘Si hubiese tenido something something like that’ about love and (expletive) in the Corinthians. Ooooh! I think it IS. Dale!” (Participant 149)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Oh my God. This is like that Marco Antonio Solis song. ‘Si no te hubieras ido, sería tan feliz. No hay nada más dificil que vivir sin ti… muriendo en la espera de verte llegar…’ Gotta download that song again. So this is probably the same right?” (Participant 128)</td>
</tr>
</tbody>
</table>

Participants’ individual episodes were categorized (when applicable) into one of these classifications. However, individual instances of each category were not taken into account; rather, at least once instance of a category in a participant’s transcription sufficed to identify that transcription as possessing that category. Therefore, one transcription could be identified as representing various TA categories. For instance, Participant 172 had two episodes that were categorized as “Intuition”, one as “Prior Knowledge (Acquired)” and three as “Prior Knowledge (Learned)”. In this case, Participant 172’s transcription, therefore, would be identified as “INTUITION/ PRIOR KNOWLEDGE (ACQUIRED)/ PRIOR KNOWLEDGE (LEARNED)”, without
specifying which occurred more. Table 17 shows the total number of participants in each linguistic profile whose transcription was identified as having one of the aforementioned TA categories.

Table 17: Results per category used for analyzing qualitative processing RQ1.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>SUB-CATEGORY</th>
<th>NUMBER OF HL (AND % OUT OF 34)</th>
<th>NUMBER OF L2 (AND % OUT OF 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>METALINGUISTIC</td>
<td>Explicit Grammar</td>
<td>10 (29%)</td>
<td>48 (96%)</td>
</tr>
<tr>
<td></td>
<td>Prior Knowledge (Learned)</td>
<td>12 (35%)</td>
<td>24 (48%)</td>
</tr>
<tr>
<td></td>
<td>Comparison Grammar</td>
<td>1 (3%)</td>
<td>5 (10%)</td>
</tr>
<tr>
<td>NON-METALINGUISTIC</td>
<td>Intuition</td>
<td>30 (88%)</td>
<td>6 (12%)</td>
</tr>
<tr>
<td></td>
<td>Prior Knowledge (Acquired)</td>
<td>20 (59%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

To summarize, there is a noticeable difference between the types of TAs employed by HL and L2 participants in this case. While HL primarily employed non-metalinguistic strategies to attempt to process the target items (primarily Intuition), L2 learners based their processing more on their metalinguistic knowledge, particularly their understanding on explicit grammatical rules, and how to apply it to the current input. Aside from the Prior Knowledge (Acquired) category, which no L2 learner utilized as a processing strategy in this experiment, each category of processing strategy had at least one instance of use per linguistic profile. These results, as well as further analysis, will be further discussed in the next chapter.
Research Questions 2 and 3

Does the Speaker Profile of participants have an effect on their subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task)? If so, does the effect last for 2 weeks?

Does the type of feedback (+explicit, -explicit, or control) have an effect on participants’ subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task)? If so, does the effect last for 2 weeks?

Descriptive Statistics

Descriptive statistics for the CPT and SSPDT are listed below:

Table 18: Descriptive statistics for controlled production task (CPT) to answer RQ2.

<table>
<thead>
<tr>
<th>CPT:</th>
<th>Pretest MEAN, (SD), [CI]</th>
<th>Posttest MEAN, (SD), [CI]</th>
<th>Delayed Posttest MEAN, (SD), [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL Learner</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=38</td>
<td>+Explicit condition</td>
<td>0.37, (0.998), [0.107, 0.630]</td>
<td>18.45, (2.114), [16.929, 19.966]</td>
</tr>
<tr>
<td>N=18</td>
<td>-Explicit condition</td>
<td>0.89, (1.451), [0.509, 1.269]</td>
<td>17.61, (4.717), [15.405, 19.818]</td>
</tr>
<tr>
<td>N= 20</td>
<td>Control Group</td>
<td>0.35, (1.137), [-0.011, 0.711]</td>
<td>0.00, (0.000), [-2.093, 2.093]</td>
</tr>
<tr>
<td>N= 9</td>
<td>Mat. Control Group</td>
<td>0.44, (1.333), [-0.093, 0.982]</td>
<td>N/A</td>
</tr>
</tbody>
</table>

| L2 Learner | | | |
| N=31 | +Explicit condition | 0.00, (0.000), [-0.290, 0.290] | 14.03, (7.662), [12.351, 15.714] | 7.10, (8.765), [4.662, 9.532] |
| N=25 | -Explicit condition | 0.00, (0.000), [-0.322, 0.322] | 12.60, (7.773), [10.728, 14.472] | 10.00, (8.036), [7.289, 12.711] |
| N= 23 | Control Group | 0.00, (0.000), [-0.336, 0.336] | 0.00, (0.000), [-1.952, 1.952] | 0.00, (0.000), [-2.827, 2.827] |
| N= 11 | Mat. Control Group | 0.00, (0.000), [-0.486, 0.486] | N/A | 0.00, (0.000), [-4.088, 4.088] |

*SD = standard deviation; CI = confidence interval
Table 19: Descriptive statistics for semi-spontaneous picture description task (SSPDT) to answer RQ2.

<table>
<thead>
<tr>
<th>SSPDT</th>
<th>Pretest MEAN, (SD), [CI]</th>
<th>Posttest MEAN, (SD), [CI]</th>
<th>Delayed Posttest MEAN, (SD), [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HL Learner</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=20</td>
<td>Control Group</td>
<td>3.05, (2.206), [2.246, 3.854]</td>
<td>2.60, (2.303), [1.484, 3.716]</td>
</tr>
<tr>
<td>N=9</td>
<td>Mat. Control Group</td>
<td>2.44, (1.878), [1.247, 3.642]</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>L2 Learner</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=11</td>
<td>Mat. Control Group</td>
<td>4.27, (1.272), [3.189, 5.356]</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*SD = standard deviation; CI = confidence interval

Prior to running statistical analyses, a one-way ANOVA was run separately for each task at the pretest time to ensure that there were no statistical significant differences between the experimental conditions at the start of the experiment. Neither the CPT (p=0.075) nor the SSPDT (p=0.220) demonstrated a statistical significant different at the pretest session. Therefore, a repeated-measures 2 x 3 x 3ANOVA was run separately on the scores obtained on the CPT and the SSPDT to ascertain whether there were significant main effects for the independent variables. Speaker profile (HS vs. L2), Type of feedback (explicit vs. implicit vs. control) were entered as the between-subject factors.
while Time (pretest vs. immediate posttest vs. delayed posttest) was entered as the within-subject factor.

Table 20: ANOVA summary table for CPT RQ2 for within-subject.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>1711.595</td>
<td>73.285</td>
<td>0.000</td>
<td>0.446</td>
</tr>
<tr>
<td>Time*Speaker profile</td>
<td>1</td>
<td>28.232</td>
<td>1.209</td>
<td>0.273</td>
<td>0.014</td>
</tr>
<tr>
<td>Time*Feedback</td>
<td>3</td>
<td>671.214</td>
<td>28.739</td>
<td>0.000</td>
<td>0.289</td>
</tr>
<tr>
<td>Time<em>Speaker profile</em>Feedback</td>
<td>3</td>
<td>15.885</td>
<td>0.680</td>
<td>0.565</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Error 167 23.355

*p<0.01

Table 21: ANOVA summary table for CPT RQ2 for between-subject.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profile</td>
<td>1</td>
<td>255.606</td>
<td>7.794</td>
<td>0.006</td>
<td>0.073</td>
</tr>
<tr>
<td>Feedback</td>
<td>3</td>
<td>2954.286</td>
<td>90.085</td>
<td>0.000</td>
<td>0.557</td>
</tr>
<tr>
<td>Speaker</td>
<td>3</td>
<td>75.347</td>
<td>2.298</td>
<td>0.079</td>
<td>0.032</td>
</tr>
<tr>
<td>Speaker*Feedback</td>
<td>3</td>
<td>111.595</td>
<td>5.182</td>
<td>0.030</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Error 167 32.794

*p<0.01

The repeated-measures ANOVA revealed significant main effects for Speaker profile ($F[1, 167] = 7.794, p = 0.006$), Type of feedback ($F[3, 167] = 90.085, p = 0.000$), and Time ($F[1, 167] = 73.285, p = 0.000$), as well as a significant interaction between Time and Type of feedback ($F[3, 167] = 28.739, p = 0.000$).
**Speaker Profile**

Table 22 demonstrates the means from the immediate and delayed posttests on the CPT, whereas Table 23 displays the means from the SSPDT.

**Table 22: Means on immediate and delayed posttest for speaker profile for CPT.**

<table>
<thead>
<tr>
<th>Posttest (CPT)</th>
<th>Mean Immediate Posttest</th>
<th>Mean Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HL Participants</strong></td>
<td>18.18 $SD = 3.169$</td>
<td>11.32 $SD = 8.722$</td>
</tr>
<tr>
<td><strong>L2 Participants</strong></td>
<td>13.39 $SD = 7.674$</td>
<td>8.39 $SD = 8.497$</td>
</tr>
</tbody>
</table>

**Table 23: Independent t-test and mean results for speaker profile for SSPDT.**

<table>
<thead>
<tr>
<th>Posttest (SSPDT)</th>
<th>Mean Immediate Posttest</th>
<th>Mean Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HL Participants</strong></td>
<td>7.321 $SD = 2.543$</td>
<td>6.411 $SD = 2.876$</td>
</tr>
<tr>
<td><strong>L2 Participants</strong></td>
<td>7.188 $SD = 3.153$</td>
<td>5.536 $SD = 3.069$</td>
</tr>
</tbody>
</table>

Given the significant main effect found for Speaker profile, independent t-tests were performed on the L2 and HL participants’ mean scores obtained on both the CPT and SSPDT’s immediate and delayed posttests (Tables 24 and 25).

On the CPT, HL participants performed significantly better than L2 participants on the immediate posttest; HL ($M = 18.18, SD = 3.169$) and L2 ($M = 13.39, SD = 7.674$), $t(112) = -4.313, p = 0.000, d = 0.82$ but not on the delayed posttest: HL ($M = 11.32, SD = 8.722$) and L2 ($M = 8.39, SD = 8.496$), $t(112) = -1.800, p = 0.773, d = 0.34$). In sum, while HL participants performed significantly better than the L2 participants on the CPT immediately after the treatment, this superior performance was not maintained two weeks later.
Table 24: Independent t-test results for immediate and delayed posttest for speaker profile for CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>F</th>
<th>p</th>
<th>T</th>
<th>Df</th>
<th>Mean difference</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate Posttest</td>
<td>79.459</td>
<td>0.000</td>
<td>-4.313</td>
<td>110</td>
<td>-4.786</td>
<td>[-6.985, -2.587]</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>0.084</td>
<td>0.773</td>
<td>-1.800</td>
<td>110</td>
<td>-2.929</td>
<td>[-6.153, 0.296]</td>
</tr>
</tbody>
</table>

Table 25: Independent t-test results for immediate and delayed posttest for speaker profile for SSPDT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>F</th>
<th>p</th>
<th>T</th>
<th>Df</th>
<th>Mean difference</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate Posttest</td>
<td>5.380</td>
<td>0.022</td>
<td>-0.247</td>
<td>110</td>
<td>-0.1339</td>
<td>[-1.2067, -0.9388]</td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>0.064</td>
<td>0.800</td>
<td>-1.557</td>
<td>110</td>
<td>-0.8750</td>
<td>[-1.9899, 0.2389]</td>
</tr>
</tbody>
</table>

On the SSPDT, the results were similar. On the SSPDT, HL participants performed significantly better than L2 participants on the immediate posttest; HL ($M= 7.321, SD= 2.543$) and L2 ($M= 7.188, SD= 3.153$), $t(112)= -0.247, p = 0.022, d= 0.05$), but not on the delayed posttest: HL ($M= 6.411, SD= 2.876$) and L2 ($M= 5.536, SD= 3.068$), $t(112)= -1.557, p = 0.800, d= 0.29$). In sum, while L2 participants performed significantly better than the HL participants on the CPT immediately after the treatment, this superior performance was not maintained two weeks later.

To summarize, differential performances were noted on the two assessment tasks. On the CPT, HL participants performed statistically better than L2 participants with a large effect size. However, this performance was not maintained by the delayed posttest two weeks later. Similar performances were recorded on the SSPDT assessment task, where the HL participants again performed significantly better than the L2 participants,
albeit with a small effect size. Again, the HL participants’ superiority did not hold by the
delayed posttest two weeks later.

**Type of Feedback**

**Controlled Production Task**

Table 26 demonstrates the means and standard deviations from the immediate and
delayed posttests on the CPT.

<table>
<thead>
<tr>
<th>CPT</th>
<th>Mean Pretest</th>
<th>Mean Immediate Posttest</th>
<th>Mean Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>+Explicit Feedback</td>
<td>0.20 SD= 0.759</td>
<td>16.46 SD= 5.764</td>
<td>8.77 SD= 8.936</td>
</tr>
<tr>
<td>-Explicit Feedback</td>
<td>0.37 SD= 1.024</td>
<td>14.70 SD= 7.056</td>
<td>11.60 SD= 8.095</td>
</tr>
<tr>
<td>Control Group</td>
<td>0.16 SD= 0.785</td>
<td>0.00 SD= 0.000</td>
<td>0.00 SD= 0.000</td>
</tr>
</tbody>
</table>

Given that the ANOVA Table revealed a significant interaction between Type of
feedback and Time \((F[3, 167] = 28.739, p = 0.000)\) in addition to the significant main
effect found for Type of feedback \((F[3, 167] = 90.085, p = 0.000)\), the main effect for
Type of feedback will be interpreted in light of this interaction.

As can be seen on Figure 8, it appears that the +Explicit feedback condition
outperformed the –Explicit feedback condition on the immediate posttest. However, on
the delayed posttest two weeks later, the –Explicit feedback group performed better than
the +Explicit group by a greater margin. At the same time, both performances of the +
and –explicit feedback groups decreased from the immediate to the delayed posttest while
the control group’s performance remained consistent across all stages.
To probe deeper into any potential difference in performance between the three experimental groups, a one-way between-subject ANOVA was run on both the immediate and delayed posttests.

Table 27: ANOVA between-subject results for CPT for immediate posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>2</td>
<td>3912.582</td>
<td>136.708</td>
<td>0.000</td>
<td>0.391</td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be seen in Tables 27 and 28, there was a main effect between Type of feedback groups on both the immediate ($F[2, 154] = 136.708, p = 0.000$) and the delayed posttests ($F[2, 154] = 30.020, p = 0.000$). In order to investigate where this difference specifically
occurred, a post-hoc Scheffé analysis was run.

Table 28: ANOVA between-subject results for CPT for delayed posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>2</td>
<td>3232.102</td>
<td>30.020</td>
<td>0.000</td>
<td>0.220</td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 29: Scheffé’s results for CPT for immediate posttest.

<table>
<thead>
<tr>
<th>(I) TYPE OF FEEDBACK</th>
<th>(J) TYPE OF FEEDBACK</th>
<th>Mean Difference (I−J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>+EXPLICIT</td>
<td>−EXPLICIT</td>
<td>1.766</td>
<td>1.039</td>
<td>.239</td>
<td>−.80−4.34</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td></td>
<td>16.464*</td>
<td>1.039</td>
<td>.000</td>
<td>13.89−19.03</td>
</tr>
<tr>
<td>−EXPLICIT</td>
<td>+EXPLICIT</td>
<td>−1.766</td>
<td>1.039</td>
<td>.239</td>
<td>−4.34−.80</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td></td>
<td>14.698*</td>
<td>1.154</td>
<td>.000</td>
<td>11.85−17.55</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td>+EXPLICIT</td>
<td>−16.464*</td>
<td>1.039</td>
<td>.000</td>
<td>−19.03−13.89</td>
</tr>
<tr>
<td>−EXPLICIT</td>
<td></td>
<td>−14.698*</td>
<td>1.154</td>
<td>.000</td>
<td>−17.55−11.85</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

On both the immediate and delayed posttests, the Scheffé revealed that both the +EF and -EF groups performed significantly better than the Control group (p = .000 for all comparisons). However, there were no significant differences between the performances of the +EF and −EF experimental conditions at the immediate (p = 0.239) or delayed posttest (p = 0.142). It could be postulated, as seen in previous empirical research (e.g. Sanz & Morgan-Short, 2004), that not finding a significant difference between the performances of the +EF and −EF conditions could be attributed to feedback not necessarily facilitating learning, but rather, that the task itself sufficed to promote learning. To summarize, while both experimental feedback groups (+EF and −EF) outperformed statistically the control group on both posttests, no significant differences between type of feedback were found on either the immediate or delayed posttest.
Table 30: Scheffé’s results for CPT for delayed posttest.

<table>
<thead>
<tr>
<th>(I) TYPE OF FEEDBACK</th>
<th>(J) TYPE OF FEEDBACK</th>
<th>Mean Difference (I−J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>+EXPLICIT</td>
<td>-EXPLICIT</td>
<td>-2.837</td>
<td>1.426</td>
<td>.142</td>
<td>-6.36</td>
</tr>
<tr>
<td></td>
<td>CONTROL GROUP</td>
<td>8.768*</td>
<td>1.426</td>
<td>.000</td>
<td>5.24</td>
</tr>
<tr>
<td>-EXPLICIT</td>
<td>+EXPLICIT</td>
<td>2.837</td>
<td>1.426</td>
<td>.142</td>
<td>6.36</td>
</tr>
<tr>
<td></td>
<td>CONTROL GROUP</td>
<td>11.605*</td>
<td>1.582</td>
<td>.000</td>
<td>7.69</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td>+EXPLICIT</td>
<td>-8.768*</td>
<td>1.426</td>
<td>.000</td>
<td>-12.29</td>
</tr>
<tr>
<td></td>
<td>-EXPLICIT</td>
<td>-11.605*</td>
<td>1.582</td>
<td>.000</td>
<td>-15.52</td>
</tr>
</tbody>
</table>

*. The mean difference is significant at the 0.05 level.

To address whether the experimental groups’ behavior changed significantly over time, a repeated measures ANOVA was run for each experimental group.

**+Explicit Feedback Experimental Condition**

Table 31: ANOVA within-subject results for CPT for +explicit feedback (EF; immediate and delayed posttest).

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>+EF</td>
<td>1</td>
<td>2043.196</td>
<td>52.265</td>
<td>0.000</td>
<td>0.435</td>
</tr>
<tr>
<td>Error</td>
<td>67</td>
<td>39.093</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the +Explicit Feedback experimental condition, a significant main effect for Time $(F[1, 67] = 52.265, p = 0.000)$ was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 32: Paired t-test results for +explicit feedback group on CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest vs. Immediate Posttest</td>
<td>-16.261</td>
<td>5.736</td>
<td>-23.548</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate vs. Delayed Posttest</td>
<td>6.935</td>
<td>8.350</td>
<td>4.624</td>
<td>0.000</td>
</tr>
<tr>
<td>Pretest vs. Delayed Posttest</td>
<td>-7.097</td>
<td>8.765</td>
<td>-4.508</td>
<td>0.000</td>
</tr>
</tbody>
</table>
The paired t-tests revealed that while the +EF group significantly improved from pretest to immediate posttest \((p = 0.000)\) and retained this significant improvement at the delayed posttest \((p = 0.000)\), a significant decrease in performance was also found from the immediate posttest to the delayed posttest \((p = 0.000)\).

**-Explicit Feedback Experimental Condition**

Table 33: ANOVA within-subject results for CPT for -EF; immediate and delayed posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>(Df)</th>
<th>Mean Square</th>
<th>(F)</th>
<th>(p)</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>-EF</td>
<td>1</td>
<td>205.686</td>
<td>9.001</td>
<td>0.005</td>
<td>0.76</td>
</tr>
<tr>
<td>Error</td>
<td>42</td>
<td>22.853</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the -Explicit Feedback experimental condition, a significant main effect for Time \((F[1, 42] = 9.001, p = 0.005)\) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 34: Paired t-test results for -explicit feedback group on CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>(SD)</th>
<th>(T)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest vs. Immediate Posttest</td>
<td>-14.326</td>
<td>6.869</td>
<td>-13.677</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate vs. Delayed Posttest</td>
<td>3.093</td>
<td>6.761</td>
<td>3.000</td>
<td>0.005</td>
</tr>
<tr>
<td>Pretest vs. Delayed Posttest</td>
<td>-11.233</td>
<td>7.037</td>
<td>-9.280</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the -EF group significantly improved from pretest to immediate posttest \((p = 0.000)\) and retained this significant improvement at the
delayed posttest ($p = 0.000$), a significant decrease in performance was also found from
the immediate posttest to the delayed posttest ($p = 0.005$).

**Control Group Experimental Condition**

Table 35: ANOVA within-subject results for CPT for control group; immediate and delayed posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>$Df$</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>1</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>41</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 35 reveals the results of a repeated-measures ANOVA, which shows no effect for
Time. Therefore, no additional analyses were necessary.

**Semi-Spontaneous Picture Description Task (SSPDT)**

Given that the ANOVA Table revealed a significant interaction between Type of
feedback and Time ($F[3, 167] = 90.085, p = 0.000$) in addition to the significant main
effect found for Type of feedback ($F[3, 167] = 28.739, p = 0.000$), the main effect for
Type of feedback will be interpreted in light of this interaction.

As can be seen on Figure 9, both the +Explicit and -Explicit feedback conditions
performed almost similarly on the immediate posttest. However, on the delayed posttest
two weeks later, the –Explicit feedback group performed slightly better than the +Explicit
group by a small margin. The control group consistently decreased over time, with a
mean average of less than four points.
Figure 9: Type of feedback over time on the SSPDT.

Table 36: Mean averages for SSPDT for pretest, immediate, and delayed posttest.

<table>
<thead>
<tr>
<th>(SSPDT)</th>
<th>Mean Pretest</th>
<th>Mean Immediate Posttest</th>
<th>Mean Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>+Explicit Feedback</td>
<td>3.913</td>
<td>7.275</td>
<td>5.797</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.964$</td>
<td>$SD = 3.073$</td>
<td>$SD = 3.220$</td>
</tr>
<tr>
<td>-Explicit Feedback</td>
<td>4.198</td>
<td>7.221</td>
<td>6.256</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.776$</td>
<td>$SD = 2.491$</td>
<td>$SD = 2.599$</td>
</tr>
<tr>
<td>Control Group</td>
<td>3.895</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.859$</td>
<td>$SD = 0.785$</td>
<td>$SD = 0.000$</td>
</tr>
</tbody>
</table>

To probe deeper into any potential difference in performance between the three experimental groups, a one-way between-subject ANOVA was run on both the immediate and delayed posttests.
Table 37: ANOVA between-subject results for SSPDT for immediate posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>2</td>
<td>469.879</td>
<td>34.405</td>
<td>0.000</td>
<td>0.241</td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 38: ANOVA between-subject results for SSPDT for delayed posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>2</td>
<td>151.073</td>
<td>19.076</td>
<td>0.000</td>
<td>0.017</td>
</tr>
<tr>
<td>Total</td>
<td>154</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be seen in Tables 37 and 38, there was a main effect between Type of feedback groups on both the immediate \((F[2, 154] = 136.708, p = 0.000)\) and the delayed posttests \((F[2, 154] = 30.020, p = 0.000)\). In order to investigate where this difference specifically occurred, a post-hoc Scheffé analysis was run.

Table 39: Scheffé’s results for the immediate posttest for SSPDT.

<table>
<thead>
<tr>
<th>(I) TYPE OF FEEDBACK</th>
<th>(J) TYPE OF FEEDBACK</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval Lower Bound</th>
<th>95% Confidence Interval Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>+EXPLICIT</td>
<td>-EXPLICIT</td>
<td>.0544</td>
<td>.5221</td>
<td>.995</td>
<td>-1.236</td>
<td>1.345</td>
</tr>
<tr>
<td>+EXPLICIT</td>
<td>CONTROL GROUP</td>
<td>-4.0195*</td>
<td>.5221</td>
<td>.000</td>
<td>2.729</td>
<td>5.310</td>
</tr>
<tr>
<td>-EXPLICIT</td>
<td>+EXPLICIT</td>
<td>-.0544</td>
<td>.5221</td>
<td>.995</td>
<td>-1.345</td>
<td>1.236</td>
</tr>
<tr>
<td>-EXPLICIT</td>
<td>CONTROL GROUP</td>
<td>3.9651*</td>
<td>.5795</td>
<td>.000</td>
<td>2.532</td>
<td>5.398</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td>+EXPLICIT</td>
<td>-4.0195*</td>
<td>.5221</td>
<td>.000</td>
<td>-5.310</td>
<td>-2.729</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td>-EXPLICIT</td>
<td>-3.9651*</td>
<td>.5795</td>
<td>.000</td>
<td>-5.398</td>
<td>-2.532</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the 0.05 level.

On both the immediate and delayed posttests, the Scheffé revealed that both the +EF and -EF groups performed significantly better than the Control group \((p = .000\) for all
comparisons). However, there were no significant differences between the performances of the +EF and –EF experimental conditions at the immediate \((p = 0.995)\) or delayed posttest \((p = 0.704)\). To summarize, while both experimental feedback groups (+EF and –EF) outperformed statistically the control group on both posttests, no significant differences between type of feedback were found on either the immediate or delayed posttest.

Table 40: Scheffé’s results for the delayed posttest for SSPDT.

<table>
<thead>
<tr>
<th>*(I) TYPE OF FEEDBACK (J) TYPE OF FEEDBACK</th>
<th>Mean Difference (I–J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>+EXPLICIT</td>
<td>–EXPLICIT</td>
<td>–.4587</td>
<td>.5468</td>
<td>.704</td>
<td>–1.810</td>
<td>.893</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td></td>
<td>2.9134*</td>
<td>.5468</td>
<td>.000</td>
<td>1.562</td>
<td>4.265</td>
</tr>
<tr>
<td>–EXPLICIT</td>
<td>+EXPLICIT</td>
<td>.4587</td>
<td>.5468</td>
<td>.704</td>
<td>–.893</td>
<td>1.810</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td></td>
<td>3.3721*</td>
<td>.6069</td>
<td>.000</td>
<td>1.872</td>
<td>4.872</td>
</tr>
<tr>
<td>CONTROL GROUP</td>
<td>+EXPLICIT</td>
<td>–2.9134*</td>
<td>.5468</td>
<td>.000</td>
<td>–4.265</td>
<td>–1.562</td>
</tr>
<tr>
<td>–EXPLICIT</td>
<td></td>
<td>–3.3721*</td>
<td>.6069</td>
<td>.000</td>
<td>–4.872</td>
<td>–1.872</td>
</tr>
</tbody>
</table>

*. The mean difference is significant at the 0.05 level.

**+Explicit Feedback Experimental Condition**

Table 41: ANOVA within-subject results for SSPDT for +EF; immediate and delayed posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>(F)</th>
<th>(p)</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>+EF</td>
<td>1</td>
<td>75.391</td>
<td>20.788</td>
<td>0.000</td>
<td>0.234</td>
</tr>
<tr>
<td>Error</td>
<td>67</td>
<td>39.093</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the +Explicit Feedback experimental condition, a significant main effect for Time \((F[1, 67] = 20.788, p = 0.000)\) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.
Table 42: Paired t-test results for +explicit feedback group on SSPDT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest vs. Immediate Posttest</td>
<td>-3.362</td>
<td>3.834</td>
<td>7.657</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate vs. Delayed Posttest</td>
<td>1.478</td>
<td>2.538</td>
<td>2.758</td>
<td>0.007</td>
</tr>
<tr>
<td>Pretest vs. Delayed Posttest</td>
<td>-1.884</td>
<td>2.784</td>
<td>4.149</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the +EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.007$).

-Explicit Feedback Experimental Condition

Table 43: ANOVA within-subject results for CPT for -EF; immediate and delayed posttest.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>-EF</td>
<td>1</td>
<td>20.026</td>
<td>11.999</td>
<td>0.001</td>
<td>0.222</td>
</tr>
<tr>
<td>Error</td>
<td>42</td>
<td>11.291</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the -Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 42] = 11.999, p = 0.001$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.
The paired t-tests revealed that while the -EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.001$).

**Control Group Experimental Condition**

Table 45 reveals the results of a repeated-measures ANOVA, which shows no effect for Time. Therefore, no additional analyses were necessary.

**Main Effect for Time**

Apart from Speaker profile and Type of feedback, the ANOVA showed a significant main effect for Time. A one-way within-subject ANOVA was performed on the mean scores of the pretest, immediate posttest, and the delayed posttest for each of
the experimental conditions (i.e., +EF, -EF, and Control Group) for each of the Speaker profiles (i.e. L2 and HL). These analyses reveal whether participants’ performances significantly changed over time.

**Controlled Production Task**

**L2 participant, +EF**

Table 46: Within-subject ANOVA summary table for CPT for second language learner (L2), +EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2, +EF</td>
<td>1</td>
<td>780.645</td>
<td>20.323</td>
<td>0.000</td>
<td>0.404</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>38.412</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01

Table 45 reveals the results of a within-subject ANOVA for the L2 participants in the +explicit feedback group for the CPT, which shows a significant main effect for Time ($F[1, 30] = 20.323, p = 0.000$). Paired t-tests were then run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest in order to investigate where the significant differences took place. Table 46 below shows the results.

Table 47: Paired t-test results for L2 +explicit feedback group for CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest vs. Immediate Posttest</td>
<td>-14.032</td>
<td>7.662</td>
<td>-10.197</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate vs. Delayed Posttest</td>
<td>6.935</td>
<td>8.350</td>
<td>4.624</td>
<td>0.000</td>
</tr>
<tr>
<td>Pretest vs. Delayed Posttest</td>
<td>-7.097</td>
<td>8.765</td>
<td>-4.508</td>
<td>0.000</td>
</tr>
</tbody>
</table>
The paired t-tests revealed that while the L2 +EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.000$).

**HL participant, +EF**

Table 48: Within-subject ANOVA summary table for CPT for heritage language learners (HL), +EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>1313.895</td>
<td>30.458</td>
<td>0.000</td>
<td>0.452</td>
</tr>
<tr>
<td>Error</td>
<td>37</td>
<td>43.138</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01

For the HL +Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 37] = 30.458, p = 0.000$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 49: Paired t-test results for HL +explicit feedback group for CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>$SD$</th>
<th>$T$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest &amp; Immediate Posttest</td>
<td>-18.079</td>
<td>2.306</td>
<td>-48.327</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate &amp; Delayed Posttest</td>
<td>8.316</td>
<td>9.288</td>
<td>5.519</td>
<td>0.000</td>
</tr>
<tr>
<td>Pretest &amp; Delayed Posttest</td>
<td>-9.763</td>
<td>8.873</td>
<td>-6.783</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the HL +EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement
at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.000$).

**L2 participant, -EF**

Table 50: Within-subject ANOVA summary table for CPT for L2, -EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>205.686</td>
<td>9.001</td>
<td>0.005</td>
<td>0.176</td>
</tr>
<tr>
<td>Error</td>
<td>42</td>
<td>92.465</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01

For the L2 -Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 67] = 52.265, p = 0.000$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 51: Paired t-test results for L2 -explicit feedback group for CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest &amp; Immediate Posttest</td>
<td>-12.600</td>
<td>7.773</td>
<td>-8.105</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate &amp; Delayed Posttest</td>
<td>2.600</td>
<td>6.225</td>
<td>2.088</td>
<td>0.048</td>
</tr>
<tr>
<td>Pretest &amp; Delayed Posttest</td>
<td>-10.000</td>
<td>8.036</td>
<td>-6.222</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the L2 -EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.048$).
**HL participant, -EF**

Table 52: Within-subject ANOVA summary table for CPT for HL, -EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>128.444</td>
<td>4.479</td>
<td>0.049</td>
<td>0.209</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td>28.680</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01

For the HL -Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 17] = 4.479, p = 0.049$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 53: Paired t-test results for HL -explicit feedback group for CPT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest &amp; Immediate Posttest</td>
<td>-16.722</td>
<td>4.561</td>
<td>-15.556</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate &amp; Delayed Posttest</td>
<td>3.778</td>
<td>7.574</td>
<td>2.116</td>
<td>0.049</td>
</tr>
<tr>
<td>Pretest &amp; Delayed Posttest</td>
<td>-12.944</td>
<td>7.689</td>
<td>-7.143</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the HL -EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.049$).
### L2 participant, Control

Table 54: Within-subject ANOVA summary table for CPT for L2, control.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>$Df$</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*p<0.01

### HL participant, Control

Table 55: Within-subject ANOVA summary table for CPT for HL, control.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>$Df$</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*p<0.01

Table 54 and Table 55 reveal the results of a within-subject ANOVA, which shows no significant effects.

### Semi-Spontaneous Picture Description Task

Next, a one-way within-subject ANOVA was performed on the mean scores of the pretest, immediate posttest, as well as the delayed posttest for each of the experimental conditions (i.e., +EF, -EF, and Control Group) for each of the Speaker profiles (i.e. L2 and HL). These analyses reveal whether participants’ performances significantly changed over time.
L2 participant, +EF

Table 56: Within-subject ANOVA summary table for SSPDT for L2, +EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>$Df$</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>70.258</td>
<td>17.529</td>
<td>0.000</td>
<td>0.369</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>4.008</td>
<td></td>
<td></td>
<td>*&lt;p&lt;0.01</td>
</tr>
</tbody>
</table>

For the L2 +Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 30] = 17.529, p = 0.000$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 57: Paired t-test results for L2 +explicit feedback group for SSPDT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>$T$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest &amp; Immediate Posttest</td>
<td>-3.3226</td>
<td>4.0219</td>
<td>-4.600</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate &amp; Delayed Posttest</td>
<td>2.1290</td>
<td>2.8313</td>
<td>4.187</td>
<td>0.000</td>
</tr>
<tr>
<td>Pretest &amp; Delayed Posttest</td>
<td>-1.1935</td>
<td>3.9868</td>
<td>-1.667</td>
<td>0.106</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the +EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.106$).
**HL participant, +EF**

Table 58: Within-subject ANOVA summary table for SSPDT for HL, +EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>17.053</td>
<td>5.513</td>
<td>0.024</td>
<td>0.130</td>
</tr>
<tr>
<td>Error</td>
<td>37</td>
<td>3.093</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01

For the HL +Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 37] = 5.513, p = 0.024$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 59: Paired t-test results for HL +explicit feedback group for SSPDT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest &amp; Immediate Posttest</td>
<td>-3.3947</td>
<td>2.8762</td>
<td>-2.4494</td>
<td>0.000</td>
</tr>
<tr>
<td>Immediate &amp; Delayed Posttest</td>
<td>0.9474</td>
<td>2.4872</td>
<td>1.7649</td>
<td>0.024</td>
</tr>
<tr>
<td>Pretest &amp; Delayed Posttest</td>
<td>-2.4474</td>
<td>3.0197</td>
<td>-1.4548</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the HL +EF group significantly improved from pretest to immediate posttest ($p = 0.000$) and retained this significant improvement at the delayed posttest ($p = 0.000$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.024$).
Table 60: Within-subject ANOVA summary table for SSPDT for L2, -EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>14.045</td>
<td>8.410</td>
<td>0.008</td>
<td>0.259</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>1.670</td>
<td></td>
<td>* p&lt;0.01</td>
<td></td>
</tr>
</tbody>
</table>

* p<0.01

For the L2 -Explicit Feedback experimental condition, a significant main effect for Time ($F[1, 24] = 8.410$ $p = 0.008$) was found. Consequently, paired t-tests were run comparing the pretest to the immediate posttest, pretest to the delayed posttest, and immediate posttest to the delayed posttest.

Table 61: Paired t-test results for L2 -explicit feedback group for SSPDT.

<table>
<thead>
<tr>
<th>T-test Results</th>
<th>Mean difference</th>
<th>SD</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest &amp; Immediate Posttest</td>
<td>-2.5200</td>
<td>3.2898</td>
<td>-3.830</td>
<td>0.001</td>
</tr>
<tr>
<td>Immediate &amp; Delayed Posttest</td>
<td>1.0600</td>
<td>1.8276</td>
<td>2.900</td>
<td>0.008</td>
</tr>
<tr>
<td>Pretest &amp; Delayed Posttest</td>
<td>-1.4600</td>
<td>3.0205</td>
<td>-2.417</td>
<td>0.024</td>
</tr>
</tbody>
</table>

The paired t-tests revealed that while the L2 -EF group significantly improved from pretest to immediate posttest ($p = 0.001$) and retained this significant improvement at the delayed posttest ($p = 0.008$), a significant decrease in performance was also found from the immediate posttest to the delayed posttest ($p = 0.024$).
HL participant, -EF

Table 62: Within-subject ANOVA summary table for SSPDT for HL, -EF.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>6.250</td>
<td>3.571</td>
<td>0.076</td>
<td>0.174</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td>1.750</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.01

Table 62 reveals the results of a within-subject ANOVA, which shows an insignificant main effect for Time \((F[1, 17] = 3.571, p = 0.076)\). Therefore, no additional analysis was necessary.

L2 participant, Control

Table 63: Within-subject ANOVA summary table for SSPDT for L2, control.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*p<0.01

HL participant, Control

Table 64: Within-subject ANOVA summary table for SSPDT HL, control.

<table>
<thead>
<tr>
<th>Source (CPT)</th>
<th>Df</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*p<0.01

Tables 63 and 64 reveals the results of a within-subject ANOVA for both control groups,
which show no significant effects.

**Effect Sizes**

While the \( p \) value determines whether or not there is a statistical significance in the findings, the effect size (Cohen’s \( d \)) determines the size (i.e. small, moderate, large) of the significance. Paired \( t \)-tests were performed to compare the means between experimental conditions, linguistic type, and time. Plonsky and Oswald’s (2014) range of effects will be utilized for this analysis, which is adapted for SLA studies, originally from Cohen (1988). Whereas Cohen identified an effect size of \( d = .20 \) as small, \( d = .50 \) as moderate, and \( d = .80 \) as large, Plonsky and Oswald’s between-groups range for SLA are \( d = .40 \) for small, \( d = .70 \) for moderate, and \( d = 1.00 \) for large. “These estimates of (roughly) small, medium, and large effects were chosen based on their approximate correspondence to the 25th, 50th, and 75th percentiles, respectively, for between-group contrasts in primary and meta-analytic research” (Plonsky & Oswald, 2014, p. 12). The results of all the paired effect sizes and \( t \)-tests are listed below in Table 64. However, there were a few findings that were worth emphasizing.

**Regarding the Speaker profile on the CPT posttest**

Among the L2 and HL participants in the experimental conditions (\( N = 112 \)), there was a moderate statistically significant difference between the two linguistic profiles in the experimental –EXPLICIT feedback condition in the immediate posttest; HL (\( M = 17.61, \ SD = 4.717 \)) and L2 (\( M = 12.60, \ SD = 7.773 \)), \( t(43) = 2.621, \ p = 0.02 \). Further, Cohen’s effect size value (\( d = 0.75 \)) suggested a moderate practical significance. Moreover, there was a moderate statistically significant difference between the two linguistic profiles in the experimental +EXPLICIT feedback condition in the immediate
posttest; HL (M=17.61, SD=4.717) and L2 (M=14.03, SD=7.662), t(69)=3.117, p = 0.00. Therefore, we reject the null hypothesis that there is no difference between linguistic profiles in the +explicit experimental condition. Further, Cohen’s effect size value (d=0.82) suggested a moderate to high practical significance.

**Regarding Type of feedback (+/- explicit) on CPT posttest**

Among the HL participants in the experimental conditions (N=56), there was a small statistically significant difference between the experimental +/- EXPLICIT feedback conditions in the immediate posttest; +EF (M=18.45, SD=2.114) and –EF (M=17.61, SD=4.717), t(56)=0.722, p = 0.36. Further, Cohen’s effect size value (d=0.26) suggested a low practical significance. A similar finding is seen among the L2 participants (N=56), where there was a small statistically significant difference between the experimental +/- EXPLICIT feedback conditions in the immediate posttest; +EF (M=14.03, SD=7.662) and –EF (M=12.60, SD=7.773), t(56)=0.689, p = 0.50. Therefore, we fail to reject the null hypothesis that there is no difference between the +/- explicit experimental conditions within the heritage language learners group. Further, Cohen’s effect size value (d=0.19) suggested a low practical significance.

**Regarding Speaker profile (L2/HL) (on CPT delayed posttest)**

There was a moderate statistically significant difference between the two linguistic profiles in the experimental –EXPLICIT feedback condition (N= 43) in the delayed posttest; HL (M= 13.83, SD= 7.853) and L2 (M= 10.00, SD= 8.036), t(43)=1.562, p = 0.13. Cohen’s effect size value (d=0.48) suggested a low practical significance. In the experimental +EXPLICIT feedback condition (N= 69) in the delayed
posttest; HL (M= 10.13, SD= 8.957) and L2 (M= 7.10, SD= 8.765), t(69)= 1.414, p = 0.17. Cohen’s effect size value (d= 0.34) suggested a low practical significance.

**Regarding Type of feedback (+/- explicit) (on CPT delayed posttest)**

Among the HL participants in the experimental conditions (N=56), there was a small statistically significant difference between the experimental +/- EXPLICIT feedback conditions in the immediate posttest; +EF (M= 10.13, SD= 8.957) and –EF (M= 13.83, SD= 7.853), t(56)= 1.572, p = 0.12. Cohen’s effect size value (d= 0.43) suggested a low practical significance. Similarly, L2 participants in the experimental conditions (N=56), showed a small statistically significant difference between the experimental +/- EXPLICIT feedback conditions in the immediate posttest; +EF (M= 7.10, SD= 8.765) and –EF (M= 10.00, SD= 8.036), t(56)= 1.289, p = 0.20. Cohen’s effect size value (d= 0.34) suggested a low practical significance.

**Regarding Type of feedback and Speaker profile (on CPT immediate post- and delayed posttest)**

Among the HL participants in the + EXPLICIT experimental condition (N=38), and the L2 participants in the –EXPLICIT experimental condition (N= 25), there was a high statistically significant difference between the -EXPLICIT feedback condition in the L2 group in the immediate posttest; HL (+EF) (M= 18.45, SD= 2.114) and L2 (–EF) (M= 12.60, SD= 7.773), t(63)= 3.675, p = 0.001). Cohen’s effect size value (d= 1.14) suggested a high practical significance. Interestingly, the opposite pattern appeared between the L2 participants in the +EXPLICIT experimental conditions (N=31), and the HL participants in the –EXPLICIT condition (N= 18), which showed a moderate to
high statistically significant difference between the Type of feedback conditions and
Speaker profiles in the delayed posttest; L2 (+EF) ($M = 7.10, SD = 8.765$) and HL (-EF)

Table 65: Mean of differences and t-test results between paired samples on CPT to answer RQ3.

<table>
<thead>
<tr>
<th>CPT</th>
<th>PRETEST</th>
<th>POSTTEST</th>
<th>DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL (+EF) &amp; HL (-EF)</td>
<td>$t(1.374); p= 0.18$</td>
<td>$t(0.722); p= 0.48$</td>
<td>$t(1.572); p= 0.12$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.45; [-1.03, 0.13]$</td>
<td>$d= -0.26; [-0.31, 0.84]$</td>
<td>$d= -0.43; [-1.01, 0.15]$</td>
</tr>
<tr>
<td>HL (+EF) &amp; L2 (+EF)</td>
<td>$t(2.285); p= 0.028$</td>
<td>$t(3.117); p= 0.004$</td>
<td>$t(1.414); p= 0.16$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.50; [0.01, 0.99]$</td>
<td>$d= 0.82; [0.32, 1.33]$</td>
<td>$d= 0.34; [-0.15, 0.83]$</td>
</tr>
<tr>
<td>HL (+EF) &amp; L2 (-EF)</td>
<td>$t(2.285); p= 0.028$</td>
<td>$t(3.675); p= 0.001$</td>
<td>$t(0.06); p= 0.95$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.48; [-0.05, 1.00]$</td>
<td>$d= 1.14; [0.58, 1.69]$</td>
<td>$d= 0.02; [-0.50, 0.53]$</td>
</tr>
<tr>
<td>L2 (+EF) &amp; L2 (-EF)</td>
<td>N/A</td>
<td>$t(0.689); p= 0.50$</td>
<td>$t(1.289); p= 0.20$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$d= 0.19; [-0.35, 0.73]$</td>
<td>$d= -0.34; [-0.89, 0.20]$</td>
</tr>
<tr>
<td>L2 (+EF) &amp; HL (-EF)</td>
<td>$t(2.602); p= 0.19$</td>
<td>$t(2.024); p= 0.05$</td>
<td>$t(2.77); p= 0.001$</td>
</tr>
<tr>
<td></td>
<td>$d= 0.89; [0.168, 1.61]$</td>
<td>$d= 0.53; [-0.08, 1.14]$</td>
<td>$d= 0.80; [0.18, 1.41]$</td>
</tr>
<tr>
<td>L2 (-EF) &amp; HL (-EF)</td>
<td>$t(2.602); p= 0.19$</td>
<td>$t(2.621); p= 0.02$</td>
<td>$t(1.562); p= 0.12$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.87; [0.29, 1.61]$</td>
<td>$d= 0.75; [0.10, 1.40]$</td>
<td>$d= 0.48; [-0.15, 1.11]$</td>
</tr>
</tbody>
</table>

$t= T$-test; $p= p$-value  $d= Cohen’s$ effect size ($d$), [Confidence Interval], Size of significance

Table 66: Mean of differences and t-test results between paired samples on SSPDT to answer RQ3.

<table>
<thead>
<tr>
<th>SSPDT</th>
<th>PRETEST</th>
<th>POSTTEST</th>
<th>DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL (+EF) &amp; HL (-EF)</td>
<td>$t(0.78); p= 0.45$</td>
<td>$t(1.28); p= 0.27$</td>
<td>$t(1.172); p= 0.12$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.22; [-0.79, 0.36]$</td>
<td>$d= -0.3; [-0.82, 0.22]$</td>
<td>$d= 0.30; [-0.88, 0.27]$</td>
</tr>
<tr>
<td>HL (+EF) &amp; L2 (+EF)</td>
<td>$t(1.083); p= 0.28$</td>
<td>$t(0.574); p= 0.57$</td>
<td>$t(0.94); p= 0.35$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.26; [-0.74, 0.23]$</td>
<td>$d= -0.14; [-0.63, 0.34]$</td>
<td>$d= 0.23; [-0.26, 0.71]$</td>
</tr>
<tr>
<td>HL (+EF) &amp; L2 (-EF)</td>
<td>$t(1.21); p= 0.23$</td>
<td>$t(0.432); p= 0.67$</td>
<td>$t(0.556); p= 0.58$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.30; [-0.82, 0.22]$</td>
<td>$d= 0.11; [-0.40, 0.63]$</td>
<td>$d= 0.14; [-0.38, 0.65]$</td>
</tr>
<tr>
<td>L2 (+EF) &amp; L2 (-EF)</td>
<td>$t(0.145); p= 0.89$</td>
<td>$t(0.896); p= 0.37$</td>
<td>$t(0.407); p= 0.69$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.04; [-0.58, 0.50]$</td>
<td>$d= 0.23; [-0.31, 0.77]$</td>
<td>$d= -0.11; [-0.65, 0.43]$</td>
</tr>
<tr>
<td>L2 (+EF) &amp; HL (-EF)</td>
<td>$t(0.144); p= 0.89$</td>
<td>$t(0.387); p= 0.70$</td>
<td>$t(1.973); p= 0.05$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.04; [-0.64, 0.55]$</td>
<td>$d= 0.10; [-0.50, 0.70]$</td>
<td>$d= 0.53; [-0.08, 1.14]$</td>
</tr>
<tr>
<td>L2 (-EF) &amp; HL (-EF)</td>
<td>$t(0.267); p= 0.80$</td>
<td>$t(1.433); p= 0.16$</td>
<td>$t(1.665); p= 0.10$</td>
</tr>
<tr>
<td></td>
<td>$d= -0.08; [-0.65, 0.49]$</td>
<td>$d= 0.43; [-0.21, 1.06]$</td>
<td>$d= 0.51; [-0.13, 1.14]$</td>
</tr>
</tbody>
</table>

$t= T$-test; $p= p$-value  $d= Cohen’s$ effect size ($d$), [Confidence Interval], Size of significance
\( M = 13.83, SD = 7.853 \), \( t(49) = 2.77, p = 0.001 \). Cohen’s effect size value \( d = 0.80 \) suggested a moderate to high practical significance.

Interestingly, there were no moderate or high statistical significances between the means on the performances of the SSPDT. Speculations as to why this may be the case will be discussed in the next chapter.

To summarize, in response to RQ3, while both +EF and -EF groups outperformed significantly the control group on both the immediate and delayed posttests on both the CPT and SSPDT, both experimental groups also evidenced a significant decrease in performance from the immediate to delayed posttest on both assessment tasks. In addition, both +EF and -EF groups significantly improved from the pretest to the immediate posttest and these significant differences were also retained on the delayed posttests (CPT and SSPDT) for both groups. The control group maintained the same score from pretest to immediate and delayed posttests on the CPT, and on the SSPDT, decreased in performance from the pretest to immediate and delayed posttest, although it was not significant. The main effect for Time revealed that, on the CPT, while both the L2 and HL +EF -EF groups significantly improved from the pretest to the immediate posttest and retained this significant improvement on the delayed posttests, all the groups also evidenced a significant decrease in performance from the immediate to delayed posttest. On the SSPDT, differential performances were observed. Both the L2 –EF and HL +EF groups revealed similar performances evidenced on the CPT. However, the L2 +EF group failed to retain the significant improvement obtained on the immediate posttest by the delayed posttest while there was no significant improvement demonstrated by the HL –EF group from the pretest to the immediate or delayed posttest.
Research Question 4

Is there a relationship between the levels of processing (low, medium, high) as measured by Think Aloud protocols and subsequent performances on the Controlled Production and Semi-Spontaneous Picture Description tasks (in the posttest, and delayed posttest)?

There is a small to medium positive correlation between both linguistic profiles’ levels of processing and subsequent scores on the immediate Controlled Production Task (CPT). However, there is a large, positive correlation between both participants’ levels or depth of processing and their performance on the delayed CPT posttest. However, the task selected along with the linguistic profile may play a crucial role, since on the Semi-Spontaneous Picture Description Task (SSPDT), there was a large, positive correlation between heritage language learners’ levels of processing and their performance on the delayed posttest, while the L2 learners’ levels of processing and performance on the SSPDT delayed posttest had a small, positive correlation; despite both profiles’ DOP having a large, positive correlation on the CPT delayed posttest. Table 64 below presents these findings, and individual relationships are highlighted as well.

The first pairs looked at were linguistic profile’s levels of processing compared to the two tasks (CPT and SSPDT) on the immediate and delayed posttests.

Table 67: List of correlations between CPT and immediate and delayed posttest.

<table>
<thead>
<tr>
<th>CPT TASK</th>
<th>Correlation between DOP SCORE and POSTTEST</th>
<th>Correlation between DOP SCORE and DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>$r = 0.315; [-0.026, 0.590]; p = 0.069$</td>
<td>$r = 0.644; [0.391, 0.806]; p = 0.000$</td>
</tr>
<tr>
<td>L2</td>
<td>$r = 0.389; [0.1, 0.617]; p = 0.01$</td>
<td>$r = 0.634; [0.408, 0.786]; p = 0.000$</td>
</tr>
</tbody>
</table>

$^*r =$ Pearson’s correlation, [Confidence Interval], $p =$ p value (where $p \leq 0.05$ is considered significant)
On the immediate posttest for the CPT, there is a small-medium correlation between L2 participants’ levels of processing and subsequent scores: $r = 0.389; [0.1, 0.617]; p = 0.01$, as well as on HL participants’ levels of processing and subsequent scores: $r = 0.315; [-0.026, 0.590]; p = 0.069$. However, there were large correlations found between both linguistic profiles and their performances on the delayed posttest for the CPT. L2 participants’ levels of processing and subsequent scores on the Controlled Production Task (CPT) delayed posttest: $r = 0.634; [0.408, 0.786]; p = 0.000$. And HL participants’ levels of processing and subsequent scores on the Controlled Production Task (CPT) delayed posttest: $r = 0.644; [0.391, 0.806]; p = 0.000$.

**Table 68: List of correlations between SSPDT and immediate and delayed posttest.**

<table>
<thead>
<tr>
<th>SSPDT TASK</th>
<th>Correlation between DOP SCORE and POSTTEST</th>
<th>Correlation between DOP SCORE and DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>$r = 0.212; [-0.131, 0.509]; p = 0.222$</td>
<td>$r = 0.644; [0.391, 0.806]; p = 0.000$</td>
</tr>
<tr>
<td>L2</td>
<td>$r = 0.419; [0.132, 0.641]; p = 0.006$</td>
<td>$r = 0.289; [-0.016, 0.545]; p = 0.063$</td>
</tr>
</tbody>
</table>

* $r$ = Pearson’s correlation, [Confidence Interval], $p$ = p value (where $p \leq 0.05$ is considered significant)

On the immediate posttest for the SSPDT task, however, there is a medium correlation between L2 participants’ levels of processing and subsequent scores on the immediate posttest: $r = 0.419; [0.132, 0.641]; p = 0.006$, whereas for HL participants, there is a small correlation between levels of processing and subsequent scores on the immediate posttest: $r = 0.212; [-0.131, 0.509]; p = 0.222$. A major difference, though, are the correlations found between the levels of processing and the delayed posttest on this task. Whereas there was a small correlation between L2 participants’ levels of processing and subsequent scores on the Semi-Spontaneous Picture Description Task (SSPDT) delayed...
posttest: \( r = 0.289; [-0.016, 0.545]; p = 0.063 \), there was a large correlation found between HL participants’ levels of processing and subsequent scores on the SSPDT delayed posttest: \( r = 0.644; [0.391, 0.806]; p = 0.000 \).

**Research Question 4.1**

*If so, does the experimental condition (+/- explicit) mediate this role?*

In order to answer this research question, correlations were run between +/- explicit feedback experimental conditions and linguistic profiles’ relationships between their levels of processing and subsequent performances on the CPT and SSPDT (both on the immediate and delayed posttest). These results are reported in Tables 66 and 67.

**CPT**

As can be seen in Table 68, the answer to this research question varies. On the immediate posttest, there appears to be a stronger correlation between levels of processing and the –Explicit experimental condition as measured by the CPT, whereas in the delayed posttest, both linguistic profiles’ levels of processing demonstrated a positive, strong correlation with the +Explicit experimental condition.

**Table 69: List of correlations between CPT and immediate and delayed posttest on experimental conditions and linguistic profiles.**

<table>
<thead>
<tr>
<th>CPT TASK</th>
<th>Correlation between DOP SCORE and POSTTEST</th>
<th>Correlation between DOP SCORE and DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ EF HL</td>
<td>( r = 0.098; [-0.387, 0.54]; p = 0.700 )</td>
<td>( r = 0.741; [0.419, 0.897]; p = 0.000 )</td>
</tr>
<tr>
<td>+EF L2</td>
<td>( r = 0.562; [0.172, 0.80]; p = 0.008 )</td>
<td>( r = 0.794; [0.552, 0.913]; p = 0.000 )</td>
</tr>
<tr>
<td>-EF HL</td>
<td>( r = 0.158; [-0.305, 0.562]; p = 0.505 )</td>
<td>( r = 0.337; [-0.212, 0.724]; p = 0.22 )</td>
</tr>
<tr>
<td>-EF L2</td>
<td>( r = 0.675; [0.249, 0.882]; p = 0.006 )</td>
<td>( r = 0.450; [0.009, 0.744]; p = 0.046 )</td>
</tr>
</tbody>
</table>

\(*r = \) Pearson’s correlation, [Confidence Interval], \( p = \) p value (where \( p \leq 0.05 \) is considered significant)
On the immediate posttest, there is a small correlation between HL participants’ levels of processing and subsequent scores on the immediate posttest of the Controlled Production Task (CPT) in the +Explicit experimental condition: $r = 0.098; [-0.387, 0.54]; p = 0.7$. A similar finding was observed in the -Explicit experimental condition: $r = 0.158; [-0.305, 0.562], p = 0.505$. However, for the L2 participants, there is a medium to large correlation between their levels of processing and subsequent scores on the Controlled Production Task (CPT) immediate posttest in the +Explicit experimental condition: $r = 0.562; [0.172, 0.80]; p = 0.008$; and a large correlation in the +Explicit experimental condition: $r = 0.562; [0.172, 0.80]; p = 0.008$.

On the delayed posttest, there is a large correlation between HL participants’ levels of processing and subsequent scores of the Controlled Production Task (CPT) in the +Explicit experimental condition: $r = 0.741; [0.419, 0.897]; p = 0.0$. A similar finding was observed in the +Explicit experimental condition for the L2 participants: $r = 0.794; [0.552, 0.913], p = 0.0$. However, for the –Explicit feedback conditions, both speaker profiles show a smaller correlation between the DOP scores and their performance on the delayed posttest. For the HL participants, there is a small to medium correlation between their levels of processing and subsequent scores on the Controlled Production Task (CPT) delayed posttest in the -Explicit experimental condition: $r = 0.337; [-0.212, 0.724]; p = 0.22$; and the L2 participants show a medium correlation in the -Explicit experimental condition: $r = 0.450; [0.009, 0.744]; p = 0.046$. 
### SSPDT

Table 70: List of correlations between SSPDT and immediate and delayed posttest on experimental conditions and linguistic profiles.

<table>
<thead>
<tr>
<th>SSPDT TASK</th>
<th>Correlation between DOP SCORE and POSTTEST</th>
<th>Correlation between DOP SCORE and DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ EF HL</td>
<td>( r = 0.135; [-0.354, 0.566]; p = 0.593 )</td>
<td>( r = 0.078; [-0.404, 0.520]; p = 0.758 )</td>
</tr>
<tr>
<td>+EF L2</td>
<td>( r = 0.580; [0.198, 0.809]; p = 0.006 )</td>
<td>( r = 0.522; [0.116, 0.778]; p = 0.015 )</td>
</tr>
<tr>
<td>-EF HL</td>
<td>( r = 0.197; [-0.350, 0.644]; p = 0.482 )</td>
<td>( r = 0.158; [-0.385, 0.62]; p = 0.573 )</td>
</tr>
<tr>
<td>-EF L2</td>
<td>( r = 0.515; [0.107, 0.774]; p = 0.017 )</td>
<td>( r = 0.517; [0.109, 0.775]; p = 0.016 )</td>
</tr>
</tbody>
</table>

*\( r \) = Pearson’s correlation, [Confidence Interval], \( p \) = p value (where \( p \leq 0.05 \) is considered significant)

On the immediate posttest, there was a small correlation between HL participants’ levels of processing and subsequent scores on the Semi-spontaneous picture description task (SSPDT) delayed posttest in the +Explicit experimental condition: \( r = 0.135; [-0.354, 0.566]; p = 0.593 \). Additionally, there was a small correlation between HL participants’ levels of processing and subsequent scores on the SSPDT delayed posttest in the -Explicit experimental condition: \( r = 0.197; [-0.350, 0.644]; p = 0.482 \).

However, there was a medium to large correlation between L2 participants’ levels of processing and subsequent scores on the SSPDT delayed posttest for both the +Explicit experimental condition: \( r = 0.580; [0.198, 0.809]; p = 0.006 \) the -Explicit experimental condition: \( r = 0.515; [0.107, 0.774]; p = 0.017 \).

Similarly, on the SSPDT delayed posttest, there was a small correlation between the HL participants’ levels of processing and subsequent scores in both the +Explicit experimental condition: \( r = 0.078; [-0.404, 0.520]; p = 0.758 \) and the -EXPLICIT experimental condition: \( r = 0.158; [-0.385, 0.62]; p = 0.573 \).
Again, there was a noticeable, medium to large correlation between L2 participants’ levels of processing and subsequent scores on the SSPDT delayed posttest in both the +Explicit experimental condition: \( r = 0.522; [0.116, 0.778]; \ p = 0.015 \) the – Explicit experimental condition: \( r = 0.517; [0.109, 0.775]; \ p = 0.016 \).

**Overall Summary**

For the first research question, HL participants tend to process input in Spanish using more non-metalinguistic strategies, whereas L2 participants employ metalinguistic strategies. The second research question found a main effect for Speaker profile on the performance of the CPT, but not on the SSPDT. Between the type of experimental condition (+/- explicit feedback), there was a significant main effect on both the performance of the CPT and the SSPDT. Among both Speaker profiles in the experimental conditions, there was a moderate statistically significant difference between the two Speaker profiles in both experimental conditions (+/- explicit feedback) in the immediate posttest of the CPT. Among both profiles of participants, there was a small statistically significant difference between the experimental conditions (+/- explicit feedback) in the immediate posttest of the CPT. On the delayed posttest, between the L2 and HL participants there was a moderate statistically significant difference between the two linguistic profiles in both experimental +/-Explicit feedback conditions in the delayed posttest of the CPT. Regarding the +/- Explicit experimental conditions on the CPT posttest, there was a small statistical significance in both Speaker profiles. Lastly, the fourth research question found a small to medium positive correlation between both linguistic profiles’ levels of processing and subsequent scores on the immediate controlled production task. However, there was a large, positive correlation between both
participants’ levels of depth of processing and their performance on the delayed CPT posttest. Furthermore, on the SSPDT, there was a large, positive correlation between HL learners’ levels of processing and their performance on the delayed posttest, while the L2 learners’ levels of processing and performance on the SSPDT delayed posttest had a small, positive correlation; contrary to both profiles’ levels of DOP having a large, positive correlation on the CPT delayed posttest.

This now leads to the discussion chapter, which will attempt to qualitatively explain the results found in this chapter, and provide insights as to the unique interaction of the participants’ linguistic profiles, the experimental condition, and the levels of processing. Furthermore, it will lead to pedagogical implications of these findings.
CHAPTER 7: DISCUSSION, IMPLICATIONS, FUTURE RESEARCH, AND CONCLUSION

This chapter examines the results presented in the previous chapter, and is structured in a similar format, where the findings are discussed in relation to each research question, as are additional analyses. The chapter concludes with a discussion on the study’s limitations, implications (both methodological and pedagogical), as well as the possibility of doing exploratory analyses and studies, principally regarding heritage language learners and the concurrent processes in which these learners (as well as L2 learners) engage during tasks, in order to advance understanding of how second (and/or heritage) language learning occurs.

Research Question 1

*How do heritage language learners of Spanish process incoming data in their heritage language? Is this process similar or different to L2 learners of Spanish?*

Results from the TA protocols demonstrated that there is a noticeable difference between the processing strategies employed by Heritage and L2 learners in the current study. While HL participants primarily employed non-metalinguistic strategies to attempt to process the target items (primarily Intuition), L2 participants based their processing more on their metalinguistic knowledge, particularly their understanding on explicit grammatical rules, and how to apply it to the current input (see Tables 70 and 71). Aside from the Prior Knowledge (Acquired) category, which no L2 learner utilized as a processing strategy in this experiment, each category of processing strategy had at least one instance of use per linguistic profile.

The qualitative results of how heritage language learners and L2 learners process
incoming data in Spanish are not surprising, given what has been reported in previous research insofar as L2 learners being exposed to more metalinguistic and formal learning strategies than their HL counterparts. Although both HL and L2 participants employed both types of strategies in order to process the input (as revealed in their TAs), there was a clear divide in the preference of each Speaker Profile (see Table 16 for strategy categories). Interestingly, only one HL participant’s concurrent data made note to try to compare the target structure to another language:

Participant 134: Starting to think this is like what I learned in Portuguese with tivesse like that super complicated form. Was it in the past? Or the future? Should’ve take Spanish instead of Portuguese.

It could be postulated that the reason for not employing this strategy (as much as the L2 participants) is due to the rich prior knowledge HL participants possess in Spanish, which may serve as an easier way to process the input, compared to accessing the grammar of other languages, which may or may not serve as useful for processing and learning. It should be noted that the above participant was born in the United States, scored relatively low on the DELE (27/50, compared to other HL participants), scored himself as 2 out of 3 on the speaking, writing, and reading self-evaluation section of the LBQ, and was enrolled in 2 years of Portuguese classes. He also made comments throughout the experiment (captured by the TA), that referred to his lack of confidence in his Spanish proficiency. Future research should attempt to tease apart these variables, as one or all could be a reason as to why he employed this strategy.

Although some L2 participants used non-metalinguistic strategies to process the
Spanish input (i.e. Intuition), none of them produced anything that could be categorized as prior knowledge in this category. Although prior knowledge is defined as “an internal cognitive act in which a linguistic form is related to some bit of existing knowledge (or gap in knowledge)” (Gass, 1997, p. 4), and could therefore include the metalinguistic category of “Prior Knowledge (Learned)” (in which 48% of the L2 participants produced), it was decided to divide these two into separate categories, to be able to discern a metalinguistic statement of previously learned information to a previous experience or acquisition. The data indicated that although both sets of participants employ a version of prior knowledge, where the participant is making a connection to the target form, one is metalinguistic and learned, whereas the other is non-metalinguistic and experienced. HL participants employing prior knowledge to facilitate processing made references to various types of cultural components (e.g. music lyrics, parents and/or grandparents using the form, church sermons and/or prayers, famous quotes from telenovelas (soap operas)). L2 participants referencing previous lessons alluded to formal grammatical instruction (lectures) from current or previous semesters, homework assignments, or books. For example, below is an L2 participant’s production of prior knowledge (metalinguistic, previous lesson):

Participant 18: ‘Habría comido’ seems to be a lot like ‘había comido’, I think I remember reading that había comido is the past of the past. So maybe this is like the past of the past that is conditional?

is different than an HL participant’s production: (non-metalinguistic, prior knowledge):

Participant 196: I’m like pretty sure Yayita used this word every time she
talked about how life would be like if Castro hadn’t come to Cuba. So maybe it’s like…what could’ve been in the past, but that it’s too late to change?

Interestingly, however, whether employed metalinguistically or non-metalinguistically, the majority of participants whose think aloud protocol alluded to prior knowledge [L2: N=22 (of 24); HL: N= 19 (of 20)] obtained higher scores on both the immediate and delayed posttest. This contributes to previous studies that show the importance of the role of prior knowledge (both acquired and learned) in processing (e.g. Gass, 1997; Leow, 1998, 2015; Robinson, 1995). These results could, however, not necessarily be a result of a direct relationship between activation of prior knowledge and subsequent positive performance, but more along the lines of great depth of processing and higher levels of awareness being involved. This will be discussed further under the analysis of RQ2.

Further probing attempted to investigate whether the experimental condition (i.e. +/- explicit feedback) played a role as to which strategy was employed.
Table 71: Results per category used for analyzing qualitative processing, RQ1 for L2 participants, separated into experimental condition.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>SUB-CATEGORY</th>
<th>NUMBER OF L2 IN +EXPLICIT FEEDBACK CONDITION (AND % OUT OF 28)</th>
<th>NUMBER OF L2 IN -EXPLICIT FEEDBACK CONDITION (AND % OUT OF 22)</th>
<th>TOTAL NUMBER (AND % OUT OF 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>METALINGUISTIC</td>
<td>Explicit Grammar</td>
<td>28 (100%)</td>
<td>20 (91%)</td>
<td>48 (96%)</td>
</tr>
<tr>
<td></td>
<td>Prior Knowledge (Learned)</td>
<td>7 (25%)</td>
<td>17 (77%)</td>
<td>24 (48%)</td>
</tr>
<tr>
<td></td>
<td>Comparison Grammar</td>
<td>4 (14%)</td>
<td>1 (5%)</td>
<td>5 (10%)</td>
</tr>
<tr>
<td>NON-METALINGUISTIC</td>
<td>Intuition</td>
<td>4 (14%)</td>
<td>2 (9%)</td>
<td>6 (12%)</td>
</tr>
<tr>
<td></td>
<td>Prior Knowledge (Acquired)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Table 72: Results per category used for analyzing qualitative processing, RQ1 for HL participants, separated into experimental condition.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>SUB-CATEGORY</th>
<th>NUMBER OF HL IN +EXPLICIT FEEDBACK CONDITION (AND % OUT OF 17)</th>
<th>NUMBER OF HL IN -EXPLICIT FEEDBACK CONDITION (AND % OUT OF 17)</th>
<th>TOTAL NUMBER OF HL (AND % OUT OF 34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>METALINGUISTIC</td>
<td>Explicit Grammar</td>
<td>10 (59%)</td>
<td>0 (0%)</td>
<td>10 (29%)</td>
</tr>
<tr>
<td></td>
<td>Prior Knowledge (learned)</td>
<td>4 (24%)</td>
<td>8 (47%)</td>
<td>12 (35%)</td>
</tr>
<tr>
<td></td>
<td>Comparison Grammar</td>
<td>0 (0%)</td>
<td>1 (3%)</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>NON-METALINGUISTIC</td>
<td>Intuition</td>
<td>17 (100%)</td>
<td>13 (76%)</td>
<td>30 (88%)</td>
</tr>
<tr>
<td></td>
<td>Prior Knowledge (acquired)</td>
<td>5 (29%)</td>
<td>15 (88%)</td>
<td>20 (59%)</td>
</tr>
</tbody>
</table>
As can be seen from the above table for HL participants, whereas the +Explicit feedback group seems to promote more metalinguistic and rehearsed strategies, the -Explicit feedback promotes more non-metalinguistic, abstract strategies. Although the L2 participant protocols did not show many instances of non-metalinguistic strategies, the – Explicit feedback group did promote more Prior Knowledge (Learned) strategies when compared to the +Explicit feedback group. It appears that the –Explicit feedback condition facilitated the employment of prior knowledge strategies, in both L2 and HL participants.

Rosa and O’Neill (1999) postulated a rationalization proposing that diverse varieties of instruction differing in explicitness may lead to different input processing; where a more explicit type of instruction could promote data-driven processing, and a less explicit type could facilitate a conceptually-driven processing. Although the present study deals with +/- explicit feedback rather than instruction, it could be hypothesized that when participants are left on their own to try to learn a new linguistic structure (with nothing other than +/- explicit feedback), it may facilitate more cognitive effort to process more deeply the linguistic data, when compared to participants provided with metalinguistic, explicit grammatical information. This will be elaborated under RQ2.

Aside from the processing categories, the participants’ TA protocols revealed interesting issues, which are worth mentioning, in order to be conscientious of them when planning tasks (experimental and/or pedagogical). Some of the L2 participants’ (N=10) TA episodes manifested a feeling of cognitive overload, whereby they complained that the input was much too confusing and complex for them to comprehend, or they lost the
motivation to attempt processing at a high level. Some examples of this apparent
cognitive overload are listed below:

Participant 83:  Why is this so hard…I’m really not getting this. This is
tiring me out.

Participant 104:  Whatever. This is over my head and I can’t wrap it around
my head at this point ugh.

Both L2 and HL participants’ TA protocols also revealed instances in which they would
compare the experimental input to their own variety of Spanish:

Participant 152 (HL):  Hm, like, I think I get what this is trying to say. But I
wouldn’t say it like any of the options here. So I wonder
if I’ve been saying it wrong this whole time (laughs)

Participant 133 (HL):  This doesn’t make sense, because I’ve never conjugated
these verbs in these ways before.

Participant 41 (L2):  I remember we had to talk in this was [past subjunctive]
in my high school Spanish class. But I am certain it
wasn’t with any of these. How the hell did we do it
then? I know I haven’t used these before. What did I
use? How would I say this?

Lastly, despite reminding HL participants that the forms to which they would be exposed
in this experiment were an alternative form of expressing this structure, TAs from HL
revealed that they nevertheless doubted their own Spanish abilities due to the complexity
of the exercise (\(N=20\)).

Participant 166:  Ok. So I know that I should be saying what is going on
in my head when I choose these answers but the truth is, I’m just winging it. Like, this sounds like something my mom would say, or that I would say when I pretend to know how to speak Spanish. It just feels right. But as to why? No clue.

Participant 168: That just sounds really weird to me. Whatever. I’ll go for it. Oh, I got it right! Yes, I actually know some Spanish! I’m not as bad as I thought I was in the beginning.

Participant 190: This doesn’t make any sense to me. This is some weird past tense shit. Clearly my family was right that I don’t know how to speak Spanish.

To summarize the results of RQ1, both L2 and HL participants employed both metalinguistic and non-metalinguistic strategies in order to process input in Spanish. Whereas the majority of L2 participants utilized metalinguistic strategies (such as referencing explicit grammar or previous lessons, and/or comparing the input to other grammars), HL participants employed mostly non-metalinguistic strategies (such as grammatical judgment and/or prior knowledge). Care must be taken in extrapolating these results, as this could be a result of this specific task, the proficiency level, the target item, or of this particular group of participants.
Research Question 2

Does the Speaker Profile of participants have an effect on their subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task? If so, does the effect last for 2 weeks?

Although both L2 and HL participants began the pretest with similar scores, the HL participants significantly outperformed the L2 participants on both tasks (CPT and SSDPT) at the immediate posttest. However, this significant advantage was not maintained by the delayed posttest and both HL participants and L2 participants significantly decreased in performance. To explicate the significant decrease from immediate posttest to delayed posttest, an analysis of participants’ scores on these two posttests was conducted.

Table 73: Analysis of participants’ scores on immediate and delayed CPT posttest.

<table>
<thead>
<tr>
<th>ON THE TARGET ITEM (PLUPERFECT SUBJUNCTIVE):</th>
<th>L2 PARTICIPANTS (52 TOTAL)</th>
<th>HL PARTICIPANTS (55 TOTAL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCORED 14 OR MORE POINTS (OUT OF 20) ON THE IMMEDIATE POSTTEST</td>
<td>35</td>
<td>53</td>
</tr>
<tr>
<td>SCORED 14 OR MORE POINTS (OUT OF 20) ON THE DELAYED POSTTEST</td>
<td>22</td>
<td>31</td>
</tr>
<tr>
<td>SCORED A 20/20 ON THE DELAYED POSTTEST</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>SCORED 13 OR FEWER POINTS (OUT OF 20) ON THE DELAYED POSTTEST</td>
<td>13</td>
<td>22</td>
</tr>
<tr>
<td>SCORED A 0/20 ON THE DELAYED POSTTEST</td>
<td>10</td>
<td>18</td>
</tr>
</tbody>
</table>

As can be seen in Table 72, out of the 35 L2 participants who scored 14 or more points on the immediate posttest CPT (what would be considered a ‘passing score’ from a
pedagogical and curricular perspective), 22 (63%) also scored 14 or more points on the delayed posttest. These cutoff scores are taken from how these scores would be reflected as a passing grade in a pedagogical setting, as this study is framed in ISLA (Leow & Cerezo, 2016). Interestingly, out of those 22 participants, 7 (32%) scored a perfect score (20/20) on the delayed posttest, whereas 10 (45%) scored a 0/20, despite having performed well on the immediate posttest two weeks prior. Similarly, out of the 53 HL participants that scored 14 or more points on the immediate posttest CPT, 31 (58%) also scored 14 or more points on the delayed posttest. Following a comparable pattern to the L2 participants, out of those 31 participants, 11 (35%) scored a perfect score (20/20) on the delayed posttest, whereas 18 (58%) scored a 0/20, despite having performed well on the immediate posttest two weeks prior. Although a decrease is usually expected between the immediate and delayed posttests, it was unanticipated to see participants score a 0, similar to what they had scored in the pretest. As concurrent measures were collected during the experiment in order to measure the levels of processing for RQ4, this data was useful in order to explain this phenomenon.

First, the participants scoring a 20/20 on the delayed posttest were explored.

Table 74: Participants’ DOP scores based on 20/20 score on delayed posttest.

<table>
<thead>
<tr>
<th>Scored a perfect 20/20 on the DELAYED POSTTEST:</th>
<th>L2 Participants (7)</th>
<th>HL Participants (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High DOP</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Medium DOP</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Low DOP</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Of the 7 HL participants who scored a perfect 20/20 on the delayed posttest, 7 were categorized as having overall processed at a high level, as per the characterization DOP chart in the previous chapter, adapted from Leow (2015). Of the 11 HL participants, 9 were categorized as processing at a high level as well. These findings are collaborated by previous empirical research in levels of processing (e.g., Adrada Rafael, 2017; Hsieh et al., 2016; Rott, 2005), which found that deeper processing of the input appeared to contribute to a greater maintenance of grammatical and lexical data. However, the 2 HL participants who were categorized as processing at an overall medium level were noteworthy. When these 2 participants’ think aloud protocols were further reviewed, it was revealed that both participants relied on robust prior knowledge (in the form of cultural associations) in order to process and comprehend the target structure:

Participant 132:  
*Oh my God. This is like that Marco Antonio Solis song.*

‘Si no te hubieras ido, sería tan feliz. No hay nada más difícil que vivir sin ti… muriendo en la espera de verte llegar…’ *Gotta download that song again. So this is probably the same right?*

Participant 149:  
*Wait… I know this. Isn’t this that super fancy word that they use in the bible? ‘Si hubiese tenido something something like that’ about love and (expletive) in the Corinthians. Oooooh! I think it IS. Dale!*

Once these two participants made their strong cultural connection (at the 8 minute, 12 second mark for Participant 132, and at the 5 minute, 45 second mark for Participant 149) utilizing prior knowledge, their cognitive effort (i.e. depth of processing) quickly
decreased, therefore being identified overall as possessing a medium level of processing. This occurrence has been reported previously by Calderón (2013). Her study investigated the association between learners’ levels of proficiency, levels of depth of processing, levels of awareness, and learners’ intake of linguistic items in aural input. She discovered that once a participant reached a high level of awareness (i.e., awareness at the level of understanding), the participant’s level of processing decreased. She revealed, for instance, a participant with an intermediate proficiency in Spanish whose think aloud protocol totaled 588 words. Within the first 297 words (at which the participant reached awareness at the level of understanding), there were 11 instances of depth of processing (2 were high, 9 were low). However, after reaching the pinnacle of level of awareness, only three more levels of processing (all low), were seen. As Leow (2015) explains, “this hints that once awareness at the level of understanding is reached, high levels of depth of processing are not only unnecessary, but also infrequent” (p. 221). Although the role of prior knowledge with regards to processing has been previously discussed (cf. Leow, 2015), it is interesting that despite not putting forth much cognitive effort (in comparison to other participants), these two participants were able to successfully score a 20/20 two weeks later on the delayed posttest based on their cultural associations via their prior knowledge.

Next, the participants who scored a 0 out of 20 on the delayed posttest were further investigated.
Table 75: Participants’ DOP scores based on 0/20 score on delayed posttest.

<table>
<thead>
<tr>
<th></th>
<th>L2 Participants (10)</th>
<th>HL Participants (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High DOP</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Medium DOP</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Low DOP</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

Not surprisingly, out of the 10 L2 participants who scored a 0 out of 20, 3 were categorized as having an overall low level of processing. This is similarly seen with the 8 HL participants (out of 18), who were identified as processing at a low level. A low level of processing is usually associated with a weaker performance and considerable decrease in their scores from the immediate to the delayed posttest (cf. Hsieh et al., 2016; Rott, 2005).

Although it was unusual that the other 7 L2 participants in this grouping were labeled as having reached a medium level of processing, their delayed CPT revealed that all 7 participants had used a form of an auxiliary verb with a past participle (e.g., *había tenido, ha tenido, habría tenido*). Although these were incorrect, and were therefore not awarded points in the CPT, it could be hypothesized that the participants were unable to remember the correct form, *hubiera*, but knew that the form consisted of an auxiliary. When these participants’ pretests were compared to their delayed posttest, it revealed that none of the 7 participants had used auxiliary verbs on their pretest. This, however, was not the case with the 6 HL participants who were identified as having achieved a medium level of processing, or the 4 HL participants categorized as processing at a high level. In order to explain these three sets of participants, a *post hoc* analysis was explored that
attempted to categorize these levels of processing more specifically into levels of awareness following Leow’s (2015) model that postulates that as depth of processing rises, so too does the potential for higher levels of awareness.

The strand of awareness research is closely associated with that of depth of processing (for a comprehensive look at the strand of awareness and depth of processing, as well as the similarities and differences between the two (see Leow, 2012, 2015 for further elaboration). Just as depth of processing is separated into three distinct categories (low, medium, high), levels of awareness are also separated into three levels, and are associated with the amount of cognitive effort put forth in attempting to process the input. The lower levels of awareness (i.e. awareness at the level of noticing and awareness at the level of reporting) are associated with minimum to medium cognitive effort and a low to medium level of processing (cf. Leow, 2012, 2015), whereas the highest level of awareness – awareness at the level of understanding, consists of strong cognitive effort, hypothesis testing and rule formulation, and/or a conscious activation of prior knowledge (cf. de la Fuente, 2016; Hsieh at al., 2016; Leow, 1997, 1998, 2001, Rosa & Leow, 2004; Rosa & O’Neill, 1999; Sachs & Suh, 2007) and is associated with the highest level of processing (cf. Leow, 2012; Leow et al., 2008; Martínez-Fernández, 2008; Morgan-Short et al., 2012; Qi & Lapkin, 2001; Rott, 2005). The paramount difference between levels of awareness and depth of processing is found at the highest level of both. At the highest level of processing, a participant could display a high level of cognitive effort, spend time processing the input, formulate hypotheses and rules, and yet never reach the underlying rule of the target form. However, by just putting forth the cognitive effort, the participant would still be coded as a high level of processing. Contrastively, in order for a participant
Table 76: Results from DOP and levels of awareness coding.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Level of Processing</th>
<th>[+awareness at level of understanding]</th>
<th>[-awareness at level of understanding]</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2-1</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-2</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-3</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-4</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-5</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-6</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-7</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-1</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-2</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-3</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-4</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-5</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-6</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-7</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-8</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-9</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-10</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>
to be coded at the highest level of awareness (i.e., awareness at the level of understanding), the participant’s protocol must demonstrate an accurate and full or partial understanding of the target structure. In other words, a participant could be coded as [+awareness at the level of understanding, + high level of processing], or [-awareness at the level of understanding, + high level of processing]; although awareness at the level of understanding usually implicates a high level of processing, the opposite is not true.

The participants (N= 17) who scored a 0 out of 20 on the delayed posttest, and were categorized as either processing at a medium or high level were re-coded for either [+ awareness at the level of understanding] or [- awareness at the level of understanding], based on whether they arrived at an accurate and/or full understanding of the target structure. Unlike the coding schema utilized to analyze correlations for RQ4, the current awareness coding is not based on the amount of +/- instances of awareness. If a participant displays at least one instance of accurate or full understanding of the target structure, it suffices to be coded as [+aware]. Results of the analysis are listed below in Table 76.

Participant 29  
Okay… so it looks like… if I see two verbs in the last sentence and one of them starts with an ‘h’, then that means that the first sentence is also going to have… two words and one of them will have an ‘h’. I’m pretty sure that’s not how I should be learning a language. But I’m going for it!
[- awareness at the level of understanding; + medium level of processing]

Participant 194

These are… all in the past. Right? Yeah. Ok. So…
‘habría…’ What does habría mean? Maybe habría IS what makes it in the past, and hubiera… um… so like “if he wouldn’t have thrown the fridge… the cake wouldn’t have fallen.” Ok, so hubiera is like the ‘if’ and the habría is like what makes it in the past. I think…

[- awareness at the level of understanding; + high level of processing]

The results indicate that perhaps the level of processing, as much as it may facilitate learning, may not be sufficient to address the reasons as to why a participant may score extraordinarily high or low, especially in a delayed posttest. Therefore, it could be postulated that participants’ medium level of processing (i.e. an average amount of cognitive effort) was enough to notice the target form, but not fully comprehend or decode it to the point of learning it robustly, and retaining the structure. Furthermore, the HL participants who put forth a high level of cognitive effort were able to score highly on the immediate posttest, but by not achieving that high level of awareness, they were unable to retain the information.

Additional investigation as to what occurred in the 10 HL participants who were categorized as processing at either a high or medium level revealed an interesting pattern. Unlike the L2 participants who utilized the incorrect auxiliary verbs, the HL participants reverted to the same form they utilized in the pretest, prior to the experiment. For
example, Participant 194 substituted ‘hubiera comido’ on the pretest for the imperfect indicative, ‘comía’. On the immediate posttest, the participant would use ‘hubiera’, and finally, on the delayed posttest, the participant returned to using the imperfect indicative. This pattern was, for the most part, repeated throughout the CPT results of these 10 participants. It could be postulated that for more complex structures that are not as salient (e.g. the current study’s linguistic target structure, the pluperfect subjunctive), a sort of fossilization, or in other words, a linguistic preservation of a grammatical structure used to express a non-salient form (that may or may not be accurate) could have occurred. Although this is not normally seen in L2 participants, it seems plausible for HL participants. Whether or not the pluperfect subjunctive existed in their Spanish variety, at some point in their interactions, there must have been a need to express the contrary-to-fact conditional sentences in the past tense. If the form that was used by these participants is not a standard grammatically-accepted form (such as hubiera/hubiera, hubiese/habría, etc.), then trying to learn a brand new complex structure in one lesson may be enough to perform well on an immediate posttest, but may not be enough to be retained (or even replace the previous form) two weeks later. However, results from the current study may aid in providing tools to facilitate this process. Whereas the four participants who scored a 0 out of 20 in the delayed posttest demonstrated high levels of processing throughout the experiment, (but did not display awareness at the level of understanding), nearly all of the participants that scored a 20 out of 20 on the delayed posttest exhibited both high levels of processing and awareness at the level of understanding.
Table 77: Results from DOP and levels of awareness coding of participants who scored 20/20 on the delayed posttest.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Level of Processing</th>
<th>[+awareness at level of understanding]</th>
<th>[-awareness at level of understanding]</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2-1</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>L2-2</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>L2-3</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>L2-4</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>L2-5</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>L2-6</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>L2-7</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-1</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-2</td>
<td>MEDIUM</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-3</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-4</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-5</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-6</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-7</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-8</td>
<td>HIGH</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>HL-9</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-10</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>HL-11</td>
<td>HIGH</td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>
When comparing the results of the HL participants scoring a 20/20 and a 0/20 on the delayed posttest, it becomes clear that while some HL participants could learn (or replace) their previous form in one lesson, it is not the equivalent for all HL participants. Many variables that have not been investigated in the current study could have been responsible for this difference, including the social and language histories of the individual participants, specific (i.e. accurate) ages and percentages of rich exposure to Spanish (aside from what was reported by the participants), and the variety of Spanish spoken at home. This plausible explanation returns to the argument presented in chapter 1, that the HL population is a heterogeneous one. In two of her studies, Silva-Corvalán (1994, 2014) investigated the grammar of verb tense, aspect and mood, verb clitics, as well as subjects via two corpora consisting of recordings of sociolinguistic conversations of 1) 50 Mexican Spanish-English adult bilinguals living in California (Silva-Corvalán, 1994), and 2) a longitudinal study of the first six years of her two Spanish-English bilingual grandchildren’s lives (Silva-Corvalán, 2014). Some of her overall findings and conclusions corroborate the findings of this study, as well as the reasoning presented to explain the results. For example, in her 2014 study, she observed no differences between her two grandchildren’s English discourse and that of monolingual English children. In Spanish, however, although she found that the simple indicative mood tenses were not challenging, the tenses that were less frequent, non-salient, and complex (such as the subjunctive) were either unstable or simply not acquired by the end of the first six years. She attributed this to a reduction in exposure of Spanish after the first six years, and further added that it appeared clear to be that the reduced amount of exposure to Spanish and the scarcer opportunities for its use eventually triggered the disparities between the
speech of a second- or third- generation HL speaker and that of their parents who have been responsible for the language input. Recently, Silva-Corvalán (2017) has expanded on her two studies by connecting them, stating that there are solid links between the bilingual acquisition occurring in the early years, and some characteristics of adult heritage language speakers’ grammars (e.g. increased production of overt subject pronouns, reduced tense, aspect, and mood system). She further drew parallels between these early and adult HL speakers’ language qualities, and discussed how these connections sustain the notion that their reduced grammars are from a ‘halted process’ of their acquisition and opportunities in the earlier years, rather than language attrition or ‘unlearning’ in adulthood.

**Research Question 3**

*Does the Type of feedback (+Explicit, -Explicit, or control) have an effect on participants’ subsequent target production, as measured by (1) a controlled production task and (2) a semi-spontaneous picture description task)? If so, does the effect last for 2 weeks?*

Both +EF and –EF groups significantly improved from the pretest to the immediate posttest and these significant differences were also retained on the delayed posttests (CPT and SSPDT) two weeks later for both groups. Additionally, both experimental groups also evidenced a significant decrease in performance from the immediate to delayed posttest on both assessment tasks. Although +EF and –EF groups outperformed significantly the control group on both the immediate and delayed posttests on both tasks, post-hoc Scheffé analyses revealed that there were no significant differences between the performances of the two feedback groups. Nevertheless, as
various studies have come forth advocating for primarily utilizing descriptive statistics over solely focusing on the p-value, (see Larson-Hall & Plonsky, in press; Nassaji, 2012; Norris & Ortega, 2006; Oswald & Plonsky, 2010; Plonsky, 2009, 2011a for a full review on the controversy), it can be noted that despite not finding a significant difference between the two feedback groups, interesting patterns did arise.

For both assessment tasks, the +Explicit feedback group performed as well as, or better than the –Explicit feedback group, whereas in the delayed posttest two weeks later, the –Explicit feedback group outperformed the +Explicit feedback group. Further probing into effect sizes revealed a small statistically significant difference between the +/– Explicit feedback conditions on the immediate CPT posttest (advantage +EF). On the delayed CPT, there was a small statistically significant difference between the +/–Explicit feedback conditions (advantage –EF). A similar, although insignificant, pattern arose on the SSPDT. When investigating the descriptive statistics, the +Explicit group displayed greater mean averages on both the CPT and SSPDT on the immediate posttest, whereas the –Explicit group displayed greater mean averages on both the CPT and SSPDT on the delayed posttest.

<table>
<thead>
<tr>
<th>TASK</th>
<th>IMMEDIATE POSTTEST</th>
<th>DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+EF</td>
<td>-EF</td>
</tr>
<tr>
<td>CPT</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>SSPDT</td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

Consistent with previous investigations, providing –Explicit feedback (i.e. in this experiment, only saying whether an answer chosen was correct or incorrect) resulted in
greater mean averages (although nonsignificant) than those in the +Explicit feedback (i.e. providing an explicit grammar explanation right after correct or incorrect answer was chosen by the participant) two weeks after the treatment as per the delayed posttest. However, immediately following the treatment, the reverse occurred. As per previous studies, participants provided with metalinguistic information in the shape of feedback exhibited an immediate benefit over those participants who were only informed of the correctness of their responses (e.g. Bowles, 2008; Lado et al., 2013; Rosa & Leow, 2014; Stafford et al., 2012). This was corroborated by the descriptive statistics of the immediate posttest on both the controlled production and semi-spontaneous picture description task, whereby both groups performed similarly (with +Explicit feedback group having a slight advantage). However, after two weeks, the participants who received the explicit, metalinguistic feedback did not maintain their advantage, and both type of feedback groups displayed evidence of learning decay (although the advantage here was displayed by the –Explicit feedback group). These findings are not groundbreaking in the field of SLA, as there are published studies (e.g. Bowles, 2008; Lado et al., 2013; Stafford et al., 2012), as well as meta-analyses (e.g. Li, 2010; Mackey & Goo, 2007) that have reported results in which participants that received less explicit forms of feedback performed better (e.g., Cerezo et al., 2016), as well (e.g. Bowles, 2008), or showed prolonged improvements (by not displaying as great a loss at the delayed posttest) (e.g. Cerezo et al., 2016; Lado et al., 2013; Stafford et al., 2012) as participants who received more metalinguistic feedback.

When looking at the Type of feedback and the Speaker profile together, a pattern arises when investigating effect sizes and mean averages. On the CPT, it appears that
participants, both L2 and HL learners, performed better on the immediate posttest if they belonged to the +Explicit feedback group, rather than the –Explicit feedback group, though the difference was not significant for the HL participants (a difference of 0.84 points on average between +/- Explicit groups) and significant for the L2 participants (a 1.43 point mean difference between +/- Explicit groups). Interestingly, however, is that this pattern changed on the delayed posttest. Both the L2 and HL participants in the – Explicit feedback group outperformed the +Explicit group, and by greater margins (a difference of 3.7 points on average between -/+ Explicit HL groups, and 2.9 points on the mean difference between -/+ Explicit L2 groups). This similar pattern did not hold true for the SSPDT immediate posttest, where the HL participants in the –Explicit condition slightly outperformed those in the +Explicit condition (a difference of 0.75 points) and the L2 participants in the +Explicit condition maintained small the advantage (a difference of 0.74 points). However, both Speaker profiles in the –Explicit feedback group managed to again outperform their +Explicit feedback counterparts in the delayed posttest, albeit by a much smaller, nonsignificant margin (a difference of 0.87 points on average between +/- Explicit HL groups, and 0.34 points on the mean difference between +/- Explicit L2 groups). This information is presented in Table 79 below.

This is in line with Hsieh, Moreno, and Leow (2016), who found that participants processing the complex gustar verb in Spanish in a less explicit instructional condition performed as well as the participants who received an explicit grammatical explanation on the immediate posttest. Further, the participants in the less explicit condition performed better than the more explicit group on the delayed posttest. Hsieh et al. also discovered that their less explicit condition prompted more cognitive effort during the
experimental task and, interestingly, that this effect remained two weeks later (for the delayed posttest). The heightened amount of cognitive effort could justify why the -Explicit condition in the current study outperformed the +Explicit group on the delayed posttest.

Table 7.9: Overview of performance on immediate and delayed posttest, separated by speaker profile and type of feedback.

<table>
<thead>
<tr>
<th>TASK</th>
<th>IMMEDIATE POSTTEST</th>
<th>DELAYED POSTTEST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPEAKER PROFILE</td>
<td>+EF</td>
</tr>
<tr>
<td>CPT</td>
<td>L2</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>HL</td>
<td></td>
</tr>
<tr>
<td>SSPDT</td>
<td>L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HL</td>
<td></td>
</tr>
</tbody>
</table>

Although all groups of participants, regardless of Speaker profile and/or Type of feedback did experience a significant loss of performance score over time (i.e., HL +EF: -8.32; HL –EF: -3.78; L2 +EF: -6.93; L2 –EF: 2.6 points in the immediate posttest versus the delayed posttest on the CPT; HL +EF: -0.95; HL –EF: -0.83; L2 +EF: -2.13; L2 –EF: 1.06 points in the immediate posttest versus the delayed posttest on the SSPDT), it seems that overall, participants in the –Explicit feedback group, regardless of Speaker profile, lost fewer points on average, as per the effect sizes. Furthermore, in general, this possible negative effect of time was compensated by the fact that all participants still managed to maintain a significantly higher gain score between the pretest and the delayed posttest. Finally, both control groups (i.e., maturational and experimental) behaved as
anticipated, in other words, no significant differences were observed between the mean scores obtained in the pretest, immediate, and delayed posttest.

**Research Question 4**

*Is there a relationship between the levels of processing (low, medium, high) as measured by Think Aloud protocols and subsequent performances on the Controlled Production and Semi-Spontaneous Picture Description tasks (in the posttest, and delayed posttest)?*

a)  *If so, does the Type of feedback mediate this role?*

To answer RQ4, positive correlations were found between levels of processing and subsequent performance on the CPT and SSPDT. Small to medium positive correlations were found between both Speaker profiles’ levels of processing and subsequent scores on the immediate CPT. However, there was a large, positive correlation between both HL and L2 participants’ levels of processing and their performance on the delayed CPT posttest. On the SSPDT, there was a large, positive correlation between HL learners’ levels of processing and their performance on the delayed posttest, while the L2 learners’ levels of processing and performance on the SSPDT delayed posttest had a small, positive correlation; contrary to both Profiles’ levels of DOP having a large, positive correlation on the CPT delayed posttest.

While larger correlations were observed on the delayed posttest two weeks later for both L2 and HL participants in the CPT, the same did not hold true for the L2 learners in the SSPDT. This could be explained by the type of assessment task administered. For the SSPDT, since it is more difficult and requires more practice to produce (because of its spontaneity), one short treatment session using feedback may not have been enough for the participants to fully internalize the structure, especially two weeks later, and despite
whether they processed deeply. The results of this study may suggest that the role type of task plays may be relative to depth of processing. As the SSPDT is more cognitively complex than the CPT, it is understandable why this occurred. Although type of assessment task (e.g., Leow, 1998) was not an independent variable, it should be investigated in future studies.

The +Explicit feedback experimental condition appeared to facilitate a larger correlation with performance and levels of processing, on both the CPT and SSPDT in the delayed posttest. However, as was elaborated on the discussions of RQ1 and RQ2, the importance may not just rely on identifying the levels of processing. Rather, it relies on the quality of the processing (e.g. prior knowledge, metalinguistic vs. non-metalinguistic), as well as if the high levels of processing facilitate successfully achieving high levels of awareness (i.e. awareness at the level of understanding). In other words, it may rely on how a participant reaches a high level of processing. For example, perhaps a participant who employs a high level of processing by making a hypothesis regarding the target item may have a higher rate of success than another participant who also employs a high level of processing, but instead does so by correcting a previous translation.

Based on the data, it is found that the deeper the processing of the target items, the likelier that it was for both Speaker profiles, in either Type of feedback, to perform better on subsequent assessments for the controlled production task. Although a similar finding was discovered on the semi-spontaneous picture description task, it was on a smaller scale, correlation-wise. This supports previous findings that found empirical support for the role of depth of processing (e.g., Adrada Rafael, 2017; Calderón, 2013; Bird, 2012; Bowles, 2003; Hsieh, Moreno, & Leow, 2016; Leow, 1997, 2001b; Rosa & Leow, 2004;
Rosa & O’Neill, 1999; Sachs & Suh, 2007, see studies on Laufer and Hulstijn’s (2001) Involvement Load Hypothesis, e.g., Hulstijn & Laufer, 2001; Rott, 2005). The present findings are in line with Calderón’s unpublished dissertation (2014), where she found a positive relationship between deeper processing and the subsequent performance of the linguistic target form (i.e. the preterit third person singular in Spanish). This is also corroborated by other previous studies such as Leow (1997, 2001) and Rosa and Leow (2004). It can be argued that a level of high cognitive effort (i.e. deep processing) is minimally required to successfully promote the learning of a target form or structure, especially if the linguistic item is complex and attempting to maintain it after the immediate posttest. When referring to Leow’s (2015) model, it shows that intake of a new linguistic structure occurs at an early stage, whereas to transform that structure from intake to knowledge occurs at a later stage. Therefore, it seems necessary that to undergo that transformation, a deeper processing is activated. A higher depth of processing appears to facilitate the internalization necessary for retained knowledge of a grammatical structure for delayed production. While participants showing a low depth of processing were still able to produce the target grammatical structures at the immediate posttest, the inability to produce said structures two weeks later might be due to having internalized the structure shallowly. However, apart from the level of processing, a greater role should be given to the quality of processing. The strategies employed to process the target structure may be just as paramount, given that a language learner could put forth an intermediate level of cognitive effort (insofar as what would be categorized as such based on the DOP schema), yet still be just as successful in their subsequent performance as a learner who yielded a higher level of processing. An instance of this are
the two HL participants who were able to activate a successful prior knowledge episode (e.g. a cultural association with a song or religion, prior experience with the form from their community of practice) that immediately afforded them the opportunity to not have to continue to process deeply in order to complete the task.

Though there is a paucity of previous studies that have addressed the specific relationship between the levels of processing and its relationship with subsequent performances (e.g. Adrada Rafael, 2017; Leow et al., 2008; Morgan-Short et al., 2012), there are no studies at the moment of this dissertation addressing the particular connection between levels of processing and feedback varying in degrees of explicitness. There have been studies, however, that have addressed the relationship between levels of awareness in +/- Explicit conditions, which corroborate the findings of this study, in that a more explicit condition facilitates a correlation with levels of processing. Rosa and O’Neill (1999), for instance, found that participants who received more explicit information about the linguistic target produced more protocols categorized as awareness at the level of understanding when compared to those participants in the less explicit condition. These findings, however, were not significant. This variance in instances at awareness at the level of understanding did not convey to subsequent performance, as the rule-search condition outperformed the more explicit condition. This is similarly seen in the current study, where although the findings were not significant, the +Explicit condition produced a greater amount of participants categorized as reaching high levels of processing. Moreover, Rosa and Leow (2004b) found that those participants in the more explicit condition reported higher levels of awareness, and outperformed the less explicit conditions. Consequently, these levels of awareness at the level of understanding
were related to a higher level of accuracy and knowledge. As the current study did not deal with instruction, but rather feedback, and not all participants were coded with the levels of awareness schema, it may explain as to why the current study’s results do not align with previous research on levels of awareness and degrees of explicitness.

Furthermore, Hsieh et al. (2016) pointed out the difference in awareness coding at the level of understanding. However, participants in Rosa & Leow (2004b) and Rosa and O’Neill (1999), that coded awareness at the level of understanding (they received the rule and reiterated it), did not arrive at this coding similarly to those in Hsieh et al.’s (2016) study, who came up with the rule on their own. Therefore, the difference appears to also be external vs. internal influence, respectively.

**Implications**

This study has implications for research and pedagogy in the field of heritage language learners, as well as in (I)SLA, overall. First and foremost, this study emphasizes the significance of collecting concurrent data (e.g., verbal reports by participants while performing the task, eye-tracking) if researchers are interested in the L2 learning process in ISLA. Without concurrent data, follow-up analyses that explain participants’ levels of processing, perceptions of the structure, awareness at the level of understanding, and even whether or not a participant was following instructions (for the methodological purpose of cleansing data from inexplicable outliers) would not have been possible. Concurrent data served as a fundamental tool in explicating the findings of this study in relation to the independent variables.

An interesting (albeit surprising) finding to consider based from this study is the way students are evaluated and accepted to register for HL classes. Some of the HL
participants were excluded from the current study because they were not heritage speakers (e.g. two participants had completed college in a Hispanic country and spoke little to no English, one participant did his think aloud in Haitian Creole, because he had never previously taken a Spanish course). It is difficult to monitor student enrollment in these courses, particularly when instructors are fearful of their courses being cancelled due to insufficient registration. Heritage language courses already experience enough discrepancy in proficiency levels of this heterogeneous population, without the addition of having completely novice and/or (almost) monolingual Spanish speakers in the same classroom. Care should be taken not only in offering various levels of heritage language courses for different learners, but also in the process of evaluation that results in students being allowed to register for these classes. Motivation and self-esteem of heritage language learners needing these classes, among other variables, could be affected by not controlling the types of students in these classes.

An additional conclusion to contemplate is that of choosing appropriate pedagogical approaches for HL learners in the classroom, regardless of its status as a mixed L2/HL or single HL classroom. This study suggests that they are capable of performing similarly to L2 learners with respect to complex structures, but it seems that they differ slightly to L2 learners on the type of knowledge to which they arrive. Furthermore, as it was discussed in RQ2, it may take more than the standard lesson that an L2/HL learner may require to learn a new structure, or for that it may take more time, trials, lessons, and/or assignments for a new or complex structure to ‘break through’ an already-existing form. For both Speaker profiles, repetition and reinforcement of the material, particularly of non-salient, complex material, may be crucial for some learners.
For both learners, reactivation of possible prior knowledge (e.g., for HL learners: prior knowledge of the same/previous lesson, or of cultural connections and/or associations; for L2 learners: connections to previous explicit grammatical instruction) may prove to be extremely advantageous in the classroom setting, as they facilitated deeper processing, and higher levels of awareness (i.e. at the level of understanding). Lastly, perhaps developing tasks that involve separate and/or combined Speaker profiles (in order to target mixed and language classrooms, depending on the school’s resources) in such a way as to facilitate higher levels of processing and awareness would result in greater retention and performance. This could be done by introducing guided induction, E-tutors, Task-Based Language Teaching (TBLT), or other forms of scaffolding and differentiated instruction within the classroom, which afford the instructor the opportunity to adapt lessons to promote deeper processing. Of course, this is limited to the conditions of this study, so the field of HL teaching needs more research that will shape how instructors can provide an optimal learning environment for HL learners. Most importantly, however, the understanding of HL learners’ linguistic profiles and experiences are paramount to design effective curriculum and instruction given the heterogeneity of the population.

Limitations

As is the case with any empirical investigation, this dissertation has some limitations, which the researcher anticipates will drive future investigations within this relatively novel subset of heritage language research.

Firstly, as has been stated previously, the DELE is not an accurate manner in which to measure proficiency, as it has a tendency to be too complex for both L2 and HL learners alike. Additionally, as is widely known in heritage language studies, it is
extremely difficult to try to perfectly control for all of the diverse variables (e.g. amount of language exposure, variety of Spanish spoken at home/country of origin, motivations, length of residency, age of arrival) that constitute the heterogeneous heritage speaker population. These factors (individually and together) could have had an effect on the participants’ proficiency levels, the type or level of processing they employed, and so forth. Future studies should perhaps replicate this study at a much grander scale, in order to be able to tease apart these variables, and cleanse the data completely from these variables. Furthermore, one variable not investigated in this study (which could result in interesting future sociolinguistic research) is the idea of language and identity, and how it relates to the proficiency of heritage language learners (i.e. do they identify more with English or Spanish?).

An additional limitation of this study is the setting of the pretest and delayed posttest. In the pilot study, all three sessions were conducted in the computer lab. This therefore afforded different opportunities for the participants in the pilot study than those in the main study. For instance, participants in the pilot study were able to complete the imperfect subjunctive PowerPoint lesson individually and at their own pace, and although time on task was recorded for the assessments, there was never a time constraint (or the perception of one). Since the researcher needed to travel to Miami in order to collect the heritage learner data, and was therefore bound by time and resources, the committee felt that there would be a better chance of recruiting more participants if the study (the pretest and delayed posttest sessions) were to be conducted in the classroom. Although all of the participants were able to complete the assessments, there is no way of ensuring that they did not feel rushed or felt a time constraint. Furthermore, due to the sessions being held in...
the classroom, it was impossible to collect concurrent data during the assessments, which would have provided richer data as to how participants were processing and performing during the tasks. Moreover, although the pretest served to evaluate whether the participants had knowledge of the distractor structure (i.e. the imperfect subjunctive) prior to the experiment, since the PowerPoint lesson was presented by the researcher (instead of self-paced), it might have affected how well the participants understood the structure in the long run.

**Future Research**

Future research should investigate the role of DOP and feedback on different levels of heritage language learners of Spanish (as well as of other languages), differing degree of complexity (e.g. imperfect/preterit distinction, por/para), and different target items (e.g. different grammatical structures, lexical items). Different levels of proficiency could employ varied levels of processing and awareness (i.e. lower levels of proficiency in a similar task could yield little to no high levels of processing, despite experimental conditions and linguistic profiles).

Additionally, another delayed posttest several weeks after the initial delayed posttest would aid in examining the extent of learning that took place, particularly in correlation to the levels of processing that occurred. Another variable that should be examined in future studies is working memory, as it may influence the outcome of the linguistic type and the processing and/or learning that could occur. Participants’ learning styles, and mode of instruction should also be further investigated, as these variables could interact with the results produced by the treatment.

Only one modality was explored in this dissertation. It would be interesting to see
how other modalities, such as oral, aural, and written would play a role in learning, feedback, and processing. Furthermore, it would be intriguing to see if the mode of the experiment (i.e. interaction), through facets such as video Skype, written (e.g. iChat), or Face-to-Face would play a role. This is especially crucial considering the rising demand for online language courses and computerized modes to promote and teach foreign and heritage languages.

Additional investigations are also necessary that examine how different tasks and assessments facilitate opportunities for the promotion of higher levels of processing and awareness. Presently, only a paucity of studies has explored this area, and none possesses concurrent or qualitative data in addition to quantitative data. In addition, as mentioned in the beginning of this study, although the field of HL learners is not new, HL learners are just beginning to be explored in the field of psycholinguistics, particularly within the areas of awareness, processing, and concurrent data. Future research specifically targeting levels of processing and awareness need to be conducted, particularly for purposes of generalization and comparability to previous research in these areas with L2 learners.

Lastly, as mentioned in the limitation section, future studies should attempt to hold all three sessions in a computer lab, to a) collect concurrent data during the assessments and the initial distractor presentation, b) allow participants to complete the presentation at their own pace, and c) for the participants to not have a time constraint in completing the assessments.

Conclusions

The present study sought to investigate 1) how Heritage Language learners
process input in their heritage language, and how it was similar and/or different to their L2 counterparts, 2) whether Speaker profile had an effect on the performance of assessment tasks, and if so, whether it would be maintained after two weeks, 3) whether the explicitness of feedback played a role in the performance of subsequent tasks, and if so, whether it would be maintained after two weeks and 4) whether there was a correlation between levels of processing and performances on the immediate and delayed posttests.

The first research question showed that HL and L2 learners employ different approaches to process L2 information. Whereas L2 learners utilize more metalinguistic strategies involving previous lessons and explicit grammatical knowledge, HL learners employed primarily non-metalinguistic strategies to include prior knowledge (e.g., cultural connections, prior experience with the language) and grammatical judgment. The role of prior knowledge was highlighted in the current study, as both Speaker profiles found ways in which to utilize it (i.e. prior knowledge regarding previous instruction, prior knowledge regarding experience), and were able to make great gains through its use.

With regards to Speaker profile, HL participants outperformed L2 participants on both the immediate and delayed posttests, on both tasks. However, although the difference was significant on both tasks in the immediate posttest, the significance did not hold on the delayed posttest two weeks later. It appears that although HL participants were able to successfully grasp the new structure for the immediate posttest, by the time two weeks had passed, many reverted to the original linguistic structure they utilized to express contrary to fact conditional sentences in the past, thereby omitting the use of the
pluperfect subjunctive structure they had learned two weeks prior. This is a crucial observance that should be taken into account when teaching HL learners, that reactivation, reinforcement, and repetition of the material and input are just as important for them as they are for L2 learners, particularly concerning either complex and/or non-salient structures.

With respect to the role of Type of feedback (+/-), no difference was found between these two types of feedback on learners’ performances on both the immediate and delayed posttests of the CPT and SSDPT tasks. Both feedback conditions significantly improved their performances from the pretest to the immediate posttest on both CPT and SSDPT; however, there was a significant decrease in performances between the immediate and delayed posttests. These findings provide new evidence that questions the effectiveness of previous studies that report the effectiveness of more explicit type of instruction over less explicit (i.e. implicit) (e.g. Norris & Ortega, 2000; Spada & Tomita, 2010). One plausible explanation may lie in the amount of cognitive effort required to process and comprehend the target form in the less explicit condition, compared to the more explicit condition, which provides metalinguistic information (i.e. the rule).

Additionally, the current study found evidence to suggest that although the more explicit feedback condition yielded more instances of deeper processing, it was not the sole indicator of successful performance in subsequent assessments. The strategies employed in order to reach the higher level of processing (e.g. prior knowledge), which may be reached at an intermediate level of processing, are just as crucial as the level of processing itself. In other words, the quality of higher levels of processing (i.e. the
specific strategies employed) should be further investigated, rather than just identifying the levels of processing (i.e. just labeling low, medium, high). In addition, the role of awareness at the level of understanding assisted in interpreting participants’ superior performances regarding the retention of targeted structure. Deeper processing was positively correlated with successful performance on both tasks at the delayed posttest mark two weeks later, primarily on the CPT, although not significantly at the immediate posttest level. These results support previous research that have also found positive correlations between higher levels of processing and performance (e.g., Adrada Rafael, 2017; Calderón, 2014).

In conclusion, the current study is arguably the first study to investigate depth of processing with Heritage Language learners (in comparison with their L2 counterparts), coupled with the additional variable of +/- explicit feedback. Speaker profile, and Type of feedback. This strand of research may afford instructors and researchers alike the opportunity to investigate innovative pedagogical interventions that stand to benefit all language learners (i.e. foreign or heritage learners) regardless of whether they are placed in combined or separated classrooms. Although Type of feedback and Speaker profile may play a role, to a degree, the principal argument resulting from this dissertation is that yes, the secret is indeed in the processing.
APPENDIX A: LANGUAGE BACKGROUND QUESTIONNAIRE

(L2 PARTICIPANTS)

Language Background Questionnaire (L2 Learners)

Participant # ____ / Initials ____

Gender (check one): Female _____   Male ________   Other _____

Age: __________

I. Personal Data

1. What is your highest level of education completed? Please select one:

Some college (year: ________)   Graduate School (year/program): _________

2. Country of origin: ________________________________

If you were not born in the U.S., during what ages did you live in your country of origin?

____________________________________________________________________________________

If you were not born in the U.S., for how long have you lived in the U.S.?

____________________________________________________________________________________

3. What is your native language?

   English ____   Spanish ____   Other ________________________________

4. Where are your parents/caregivers from?

   Mother: _________________   Father: _________________

   Other: ________________________________

5. What languages do your parents/caregivers speak?

   Mother: _________________   Father: _________________

   Other: ________________________________
6. What other languages do you speak (list language, age that you started learning it, and how many years have you spoken it)?

<table>
<thead>
<tr>
<th>Language</th>
<th>Age (Started Learning)</th>
<th>Years (Speaking)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. In the boxes below, rate your language ability in each of the languages that you know. Use the following ratings: 0: Poor, 1: Good, 2: Very good, 3: Native/nativelike

<table>
<thead>
<tr>
<th>Language</th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. Have you ever been to a Spanish-speaking region for the purpose of studying

   *Spanish? YES _____ NO _____

   a. If yes, when? Where? __________________________

   b. For how long? __________________________
APPENDIX B: LANGUAGE BACKGROUND QUESTIONNAIRE

(ITALIAN PARTICIPANTS)

Language Background Questionnaire (Heritage Speakers)

Participant # ___ / Initials ___ /

Gender (check one): Female ______  Male ________  Other ______

Age: __________

I. Personal Data

1. What is your highest level of education completed? Please select one:

   Some college (year: _________)  Graduate School (year/program): __________

2. Country of origin: ____________________________

   If you were not born in the U.S., during what ages did you live in your country of origin?

   __________________________________________________

   If you were not born in the U.S., for how long have you lived in the U.S.?

   ____________________________________________________________________________________

4. What is your native language?

   English ___  Spanish ___  Other ________________________________

4. Where are your parents/caregivers from?

   Mother: _______________  Father: _______________

   Other: ________________________________

5. What languages do your parents/caregivers speak?

   Mother: _______________  Father: _______________

   Other: _______________________________________________________________________________
6. What other languages do you speak (list language, age that you started learning it, and how many years have you spoken it)?

<table>
<thead>
<tr>
<th>Language</th>
<th>Age (Started Learning)</th>
<th>Years (Speaking)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. In the boxes below, rate your language ability in each of the languages that you know. Use the following ratings: 0: Poor, 1: Good, 2: Very good, 3: Native/nativelike

<table>
<thead>
<tr>
<th>Language</th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Have you ever been to a Spanish-speaking region for the purpose of studying

**Spanish?** YES ______ NO ______

c. If yes, when? Where? __________________________

d. For how long? __________________________

******************************************************************************

**III. Your Linguistic History**

1. At what age did you first begin to learn English?
2. At what age did you first begin to learn Spanish?

3. Did you begin to speak both English and Spanish before age 5? (circle one)
   
   Yes  No

4. What languages did you hear in your home between the ages of birth-5 years? (circle all those that apply)
   
   Spanish    English    Mixed    Other (specify) ______

5. What languages did your parents/caregivers use mostly when speaking to you?
   
   Spanish    English    Mixed    Both    Other

6. What languages did you use mostly when speaking to your parents/caregivers?
   
   Spanish    English    Mixed    Both    Other

7. Do you have siblings?
   
   Yes  No  How many? __________________________

   Are they older or younger? __________________________

8. What language/s did you use when speaking with your siblings?
   
   Spanish    English    Mixed    Both    Other

9. What language/s did your siblings use when speaking with you?
   
   Spanish    English    Mixed    Both    Other

10. Did grandparents live at home?
    
    Yes  No

11. What language/s did your grandparents use when speaking to you?
    
    Spanish    English    Mixed    Both    Other

12. What language/s did you use when speaking with your grandparents?
    
    Spanish    English    Mixed    Both    Other
13. Were there other caregivers in the house (baby-sitter/ other family member)?
   Yes  No  Who?

14. What language/s did your other caregiver use when speaking to you?
   Spanish  English  Mixed  Both  Other

15. What language/s did you use when speaking to your other caregiver?
   Spanish  English  Mixed  Both  Other

16. Did you play with other Spanish-speaking children?
   Yes  No

17. What languages did you use with other children/siblings?
   Spanish  English  Mixed  Both  Other

18. Did you watch TV in Spanish?
   Yes  No

19. Did your parents encourage you to speak Spanish as much as possible in the house?
   Yes  No

20. Did your parents read stories in Spanish to you?
   Yes  No
Contrary to Fact Conditional Sentences in Spanish

Conditional Sentences

* Although the rules of grammar for conditional sentences — usually those using the word *si* ("if") — can get fairly complex, in the vast majority of cases, the decision of which verb tense to use after *si* is easy to remember.

* The first thing is to remember that except in very rare cases, *si* is never followed by a verb in the present-tense subjunctive mood.
Ejemplos

* Si Francisco está encerrado en su cuarto, está viendo Netflix.
* If he is locked in his room [he may actually be in his room], he is watching Netflix.

* Si nieva mucho, podré esquiar.
* If it snows a lot [It may really snow], I can ski.

“Real vs. Unreal” Conditions

Real

* A real condition is one which may actually come about or at least is viewed as a possibility.

* In conditions where possibility is reasonably likely, si is followed by the present indicative tense.

Unreal

* An unreal or contrary-to-fact condition is one which will not come about or is viewed as being completely hypothetical.

* If the condition is unlikely or false, a past subjunctive (usually the imperfect subjunctive) is used. This is the case even when the condition is something that refers to the present.
There are two types of sentences:

PRESENT OR FUTURE TIME

AND

PAST TIME SITUATIONS

This is the one we will focus on today!

Sentences in which the condition is contrary to fact or is unlikely.

(1) “Si lloviera, me iría para el cine en vez del parque.”

For example, the subordinate clause (the first part of the sentence) si lloviera can be translated as "if it were to rain". It’s in the imperfect subjunctive (the past subjunctive) because it expresses uncertainty.

*Note the difference in meaning from this example:

“Si llueve, me voy para el cine.”

In this case, the rain is very possible. However, in example (1), while rain is a possibility, it is seen as unlikely.

*The main clause (the second clause) will be in the conditional because you are speculating about something which would happen only if the unlikely event in the subordinate clause (the first clause) happens first.
* Si yo fuera rico compraría un coche.
(If I were rich [I am not rich], I would buy a car).

* ¿Qué harías si fueras presidente
(What would you do if you were the president [you’re not the president]?)
Wait, how do you form the imperfect/past subjunctive again?

• Start with the third-person plural [ellos] form of the preterit.
  • EXAMPLE:
    COMER → ELLOS COMIERON
    HABLAR → ELLOS HABLARON
    RIER → ELLOS RIERON
• Drop the -on ending to get the verb’s imperfect/past subjunctive base.
  COMIERON → COMIER-
  HABLARON → HABLAR-
  RIERON → RIER-

<table>
<thead>
<tr>
<th></th>
<th>-a (comiera)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yo</td>
<td>-a (comiera)</td>
</tr>
<tr>
<td>Tú</td>
<td>-as (comieras)</td>
</tr>
<tr>
<td>Él/Ella/ello/uno</td>
<td>-a (comiera)</td>
</tr>
<tr>
<td>Nosotros/nosotros</td>
<td>-amos (comiéramos)</td>
</tr>
<tr>
<td>Ellos/Ellas</td>
<td>-an (comieran)</td>
</tr>
<tr>
<td>Usted</td>
<td>-a (comiera)</td>
</tr>
<tr>
<td>Ustedes</td>
<td>-an (comieran)</td>
</tr>
</tbody>
</table>

Ok, so now YOU try!
CLICK on the correct answer!

* Si Juana (estar) aquí, ¿le dirías la verdad?
  (If Juana were here [she isn’t here], would you tell her the truth?)

* Is it:
  * ESTARÍA
  * ESTUVIERA
  * HUBIERA ESTADO
  * ESTABA

Imperfect/past Subjunctive because this is unlikely to happen
Conditional- because this clause is talking about the situation in the present or future
Yay! Good job!

Remember: when it’s an unlikely scenario with the main (second) clause resulting in the present or future, you use the IMPERFECT/PAST SUBJUNCTIVE in the subordinate (first) clause!

Press HERE to continue

Oops! Try again!

CLICK HERE
Yo lloraría si Sam (ganar) la lotería,
(If Sam were to win the lottery [he won't], I would cry.)

Is it:
* GANARA
* GANARÍA
* GANABA
* GANARÁ

Here’s another! CLICK on the correct answer!

Conditional because this clause is talking about the situation in the present or future

Imperfect/past Subjunctive because this is unlikely to happen

Yay! Good job again!

Remember: when it’s an unlikely scenario with the main (second) clause resulting in the present or future, you use the IMPERFECT/PAST SUBJUNCTIVE in the subordinate (first) clause!

Press HERERE to continue
REMEmBER:

<table>
<thead>
<tr>
<th>“if” clause (SUBORDINATE CLAUSE)</th>
<th>main clause</th>
<th>time aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>si + imperfect subjunctive</td>
<td>conditional</td>
<td>Present / future time actions</td>
</tr>
</tbody>
</table>
| Si hablaras,  
If you spoke. | te creerán.  
they would believe.  
you. | (but expressed by the past tense in both English and Spanish) |
Si el libro **costar** mucho, no lo compraría.

(If the book **were expensive** [it isn't], I **would not** buy it.)

**Is it:**

* COSTARÁ
* COSTABA
* COSTARÍA
* COSTARA

**Imperfect/past Subjunctive because this is unlikely to happen**

**Conditional**- because this clause is talking about the situation in the present or future

Yay! Good job again!

Remember: when it’s an unlikely scenario with the main (second) clause resulting in the present or future, you use the IMPERFECT/PAST SUBJUNCTIVE in the subordinate (first) clause!

Press [HERE](#) to continue
**TO RECAP:**

- There are different types of conditional sentences.
- Those that feature situations (in the subordinate/first clause) are very unlikely to happen (followed by ‘si’ (if)) are called contrary to fact conditional sentences.
- If the main (second) clause results in the situation taking place in the present or future, then the verb in the subordinate (first) clause is in the **IMPERFECT/PAST SUBJUNCTIVE**.

* The **IMPERFECT/PAST SUBJUNCTIVE** is conjugated by:

  - Starting with the third-person plural [ellos] form of the preterit. (e.g. comer \(\rightarrow\) comieron)
  - Drop the -on ending to get the verb’s imperfect subjunctive base. (comieron \(\rightarrow\) comier-)
  - Add –a (or -an, -amos, depending on the conjugation). (comier- \(\rightarrow\) comiera)
¡Gracias!
Please tell the researcher you are done.

You will now take a brief assessment.
APPENDIX D: CONTROLLERD PRODUCTION TASK,

ALL SCENARIOS

**Controlled Production Task:** Inside of this envelope, you will find 2 short stories separated into two sentences per page. Read each scenario, then complete the following sentences by conjugating the verb in the parenthesis to best fit the meaning and timeline of this current scenario. After you finish each page, please flip it over (writing face down) and place it in a stack next to you. **Do not go back to any of the previous pages.**

**Scenario 1:** Ángela, a University of Miami student, forgot that she had a major test in her Spanish class last Monday. Now she realizes the consequences of forgetting about her exam, and of not studying for it.

Si Ángela ____________________ (escribir) la fecha del examen en su agenda, se habría recordado que tenía que estudiar en vez de ir a una fiesta. Su compañera de cuarto le habría recordado de su examen si Ángela le ________________ (decir) a su compañera de cuarto que iba a salir a una fiesta. Pero, cómo no le _______ (comentar) nada en esta ocasión, su compañera de cuarto no se lo recordó. De cualquier modo, habría podido estudiar si no ____________ (beber) tanto durante las fiestas. La última vez que Ángela ______________ (tomar) mucho, tuvo un gran dolor de cabeza, y durmió mucho. Si Ángela ________________ (estudiar) para su examen durante el fin de semana, no habría sacado una mala nota. Si ________________ (hacer) la guía de estudio antes de ir a la fiesta, tal vez habría sacado una mejor nota. Si ________________ (poner) la alarma la noche anterior, no se habría despertado tarde el día del examen. Se habría despertado más temprano sin su alarma si no ________________ (cubrir) la ventana con una sábana. Si ________________ (vivir) en un dormitorio más cerca a la universidad, habría
llegado a tiempo para el examen. Si no _______________ (olvidarse) de sus responsabilidades escolares, tal vez sería aceptada a un programa de estudiar el extranjero. Si ella _______________ (querer) estudiar en Argentina, necesitaría una aplicación con buenas notas y una buena carta de recomendación. Si Ángela _______________ (producir) una buena nota en la clase, su profesora le escribiría una buena carta de recomendación para su aplicación de estudiar en el extranjero. Pero, cómo no _______________ (estudiar) lo suficiente, la profesora no se la escribió. Si _______________ (asistir) a las sesiones con el tutor de español, la habría ayudado a sacar mejores notas en la clase. El comité la aceptaría al programa si su aplicación para el programa de estudiar en el extranjero _______________ (ser) mejor. Le iría mejor en la clase si _______________ (prestar) más atención a los anuncios de su profesora de español. Además, si no _______________ (sacar) una mala nota en su clase, no habría perdido la beca. Si _______________ (continuar) con su beca, Ángela no tendría que trabajar para pagar la matrícula. Este examen le _______________ (enseñar) a Ángela sobre la importancia de no salir a fiestas cuando tienes que estudiar.

**Scenario 2**: José is a Georgetown senior who woke up the day after Georgetown Day in a jail in Key West, Florida. His roommates are trying to remind him of the bad decisions he made which led him to be in this predicament.

José, en verdad (nosotros) no te _______________ (creer) cuándo nos dijiste que estabas en la cárcel (jail). Si _______________ (acostumbrar) a desayunar antes de "celebrar", no te emborracharías tan rápido como costumbre. No le habrías pedido a esa chica que te diera su agua si _______________ (traer) una botella de agua a Saxby’s. Pero, si no le _______________ (quitar) su agua, ella no se habría
reído de ti y no habría comenzado a hablar contigo. Si una chica ________________
(sonreír) después de eso, ¡definitivamente estaría interesada en ti! ¿Te recuerdas de
Patricia? Ella por seguro no ________________ (gustar) de ti, y lo supimos porque
nunca te sonrió, aunque le compraste un café.

Bueno, regresando a esta chica. Como un tonto (dummy), después que ella te dio
su número de teléfono, lo escribiste en el recibo (receipt) del café. Todavía habrías
tenido el número de teléfono de la chica si no le ________________ (devolver) el
recibo firmado al cajero. Si no ________________ (perder) el número de la chica
con quien estabas hablando, no habrías corrido atrás de ella en el medio del puente
de Francis Scott Key.

Si ________________ (correr) más durante la semana, no tendrías que pararte
tan frecuentemente para respirar en situaciones como ésta. ¿No te
_______________ (invitar) la semana pasada para que hicieras ejercicio con
nosotros? Si no ________________ (parar) para respirar en el medio de la calle, la
bicicleta no te habría pegado. Si te ________________ (golpear) la bicicleta un poco
más fuerte, te habrías caído del puente. ¡Menos mal que no pasó! Si no te
_______________ (rescatar) Kim Kardashian del medio de la calle, te habría
atropellado (run over) un carro por seguro. Todos los del alrededor
_______________ (quedarse) asombrados en verlos juntos. Pero recuerda, Kim
Kardashian jamás te prestaría atención si no te ________________ (parecer) tanto
to a Kris Humphries.

Si Kim Kardashian no ________________ (creer) que en verdad eras Kris
Humphries, no te habría llevado en su avión privado. Si (pretendiendo ser Kris
Humphries, por supuesto) no le ________________ (volver) a pedirle que se casara contigo, no te habría tirado en la cárcel por intoxicación pública. ¡Pero menos mal! Porque si el equipo de baloncesto de Miami, Los Miami Heat, ________________ (descubrir) que “Kris Humphries” de los Atlanta Hawks estuviera en Miami, te habrían matado. Por cierto, Los Heat ________________ (ganar) el juego de anoche. ¡Ahora van a los semi-finals!

Bueno José, si ________________ (saber) antes de Georgetown Day que así es como terminaría tu cuento, ¿cambiarias algo? ...¡Nunca! Best Georgetown Day EVER!

**Scenario 3:** Jude is a university student that lives with his sister, Allison, and her son, Logan. Jude has a very short temper. It causes him to act out impulsively out of anger when something doesn't go his way, and consequently he gets himself into a lot of trouble.

Si Jude ________________ (tener) más paciencia, no sería tan impulsivo. Si Allison no se ________________ (comer) el último pedazo de chocolate, Jude no habría tirado el refrigerador al piso. Si no ________________ (ser) el chocolate que se le iba a regalar a su novia, a Jude no le habría importado tanto. Jude ________________ (pasar) mucho tiempo la semana pasada específicamente buscando ese chocolate.

Si Jude ________________ (abrir) la puerta del refrigerador antes de tirarlo al piso, habría visto la torta de helado que compró Allison para el cumpleaños de Logan. Como el año pasado a Logan no le gustó la torta regular, Allison ________________ (escoger) una torta de helado este año. Si Jude no ________________ (romper) el refrigerador al tirarlo, la torta no se habría
derretido. Si Logan___________ (ver) su torta de helado en el piso, habría
lorrado. Si no ______________ (estar) tan lejos el supermercado de su casa, Jude
iría corriendo para comprarle una nueva torta de helado a su sobrino. Jude
compraría la torta de helado en el 7-11 al lado de su casa si se
_____________ (encontrar) ahí. Si Jude _______________ (comprar) otra torta
de 7-11 sin helado, Logan se daría cuenta inmediatamente. Si Jude
_____________ (manejar), podría ir al supermercado en carro. Si él
_____________ (llamar) un taxi más temprano, habría podido ir al
supermercado antes de que su novia llegara a su casa.
Si la novia de Jude no _______________ (tener) un coche, él no habría podido ir a
buscarle una nueva torta a su sobrino. Si Jude y su novia no _______________ (ir) al
supermercado para comprar la torta, no habrían tenido que cancelar su cita
romántica. ¡Menos mal que el supermercado no _______________ (cerrar) hasta
después de que compraron la nueva torta! La fiesta de Logan habría sido arruinada
si Jude _______________ (volver) a casa sin la torta nueva. Si Jude no
_____________ (resolver) esto antes de que se despertara Logan, su sobrino se
habría enterado de lo que hizo su tío. Si él _______________ (poder) pedirle
perdón a todos quienes fueron afectados por su mal carácter, lo haría en un instante.
Pero al final del día, no lo _______________ (hacer) porque él tuvo mucha
vergüenza (embarrassment) de lo que pasó.

Scenario 4: Sandra’s boyfriend Taylor didn’t come to her quinceañera (15th
birthday party). She wrote him a Facebook message to express how upset she was at
his absence.

Hola Taylor,
En verdad me _______________ (sorprender) cuando vi que no fuiste a mi fiesta de cumpleaños. Me habría gustado si _______________ (venir), porque _______________ (ser) una fiesta muy divertida, según todos nuestros amigos. Sé que _______________ (tener) que trabajar hasta muy tarde, pero no me habría enfadado tanto si _______________ (llegar) tarde. Y honestamente, te habría buscado del trabajo si me lo _______________ (pedir) con anticipación. Pero así eres... si _______________ (ser) más considerado, no serías tú. Aunque sea si _______________ (estar) enfermo, entendería mejor la razón por la cual no viniste. En verdad no lo entiendo. Por cierto, sí _______________ (recibir) tu regalo. No sé por qué dices que no me gustó. ¡Me encantó! Era exactamente lo que quería. Si no me _______________ (gustar) el regalo, lo habría cambiado por otra cosa. ¿Me habrías comprado ese regalo tan caro si _______________ (ir) a mi fiesta? Apreciaría el regalo más si (tú) _______________ (compartir) más tiempo contigo por lo general.

También me habría gustado si me _______________ (escribir) un mensaje para mi cumpleaños. Es cierto que tal vez si mi Papá _______________ (encontrar) esa carta, él se habría disgustado. Si el _______________ (saber) que tengo novio, no me dejaría salir más los fines de semana. Pero él tiene que entender- ya _______________ (cumplir) 15 años. ¡Soy una adulta ahora!

Si no te _______________ (conocer), nunca me habría imaginado que podría ser tan feliz. Pero sería más feliz si me _______________ (informar) de tus decisiones con anticipación. Si me _______________ (mandar) un texto, me habría dado
cuenta que no ibas a venir... habría parado de buscarme si sólo me lo ____________________ (advertir).

Creo que esta situación me ha dejado muy molesta, y creo que necesitamos separarnos.

Hasta luego, Sandra

**Scenario 5**: Lauren and Rachael are best friends, and are texting each other the morning after going to a Pitbull concert.

Lauren: ¡Qué increíble ________________ (estar) el concierto de Pitbull!

Rachael: Honestamente, no me ________________ (gustar) que tuvimos que pagar tanto dinero por esas entradas.

Lauren: Ah bueno. Pero si las ________________ (comprar) con anticipación, no habríamos pagado tanto por ellas.

Rachael: Sí, pero recuerda que si las ________________ (conseguir) temprano, tendrías que hacer una fila larga.

Lauren: Bueno, no harías fila para los conciertos si ________________ (entrar) en los concursos para ganarte las entradas.

Rachael: No sé... si ________________ (ganar) las entradas, se las habría donado a una obra de caridad.

Lauren: Ese es tu problema... ¡si no ________________ (ser) tan caritativa, podrías ahorrarte dinero y tiempo!

Rachael: Tienes razón. Pero no me habría sentido bien quedándome con las entradas si ________________ (ganar).
Lauren: Bueno, cambiando el tema. ¿Por qué no nos ___________ (encontrar) en el restaurante después del concierto?

R: Si ___________ (saber) que ibas a un restaurante después del concierto, habría ido contigo.

L: Pero ¿cómo vas a decir eso si yo te lo ___________ (decir) cuando entramos?

R: No te escuché. Me habría encantado acompañarte si me ___________ (invitar) con más anticipación.

L: Si me ___________ (mandar) un texto después del concierto, te habría informado de dónde estaba.

R: No hay sentido de pelear sobre eso ahora. Pero ahora en adelante, preferiría si me ___________ (dejar) saber algunos días antes.

L: Está bien. Oye, ¿te enteraste que Daddy Yankee va a tener un concierto pronto?

R: Sí, lo sé. No quiero ir.

L: ¡Pero ya te compré tu entrada como sorpresa! Si ___________ (saber) que no te gustaba Daddy Yankee, ni lo mencionaría.

R: Oh, ¡gracias! No es eso, es que quiero pasar más tiempo estudiando.

L: Bueno, si ___________ (tomar) esa decisión hace una semana, no habría gastado el dinero.

R: Perdón. ¿pero no habría sido mejor si se lo ___________ (obtener) para tu enamorado?

L: Sí pero ___________ (querer) llevar a mi mejor amiga. No hay problema.

Llevaré a otra persona.
R: Si no lo ________________ (hacer) como sorpresa, te lo habría dicho...

L: Está bien. No estoy brava.

R: Wow, Daddy Yankee... ¡habría preferido mil veces ir al concierto de él si ________________ (tener) una opción entre él y Pitbull!
APPENDIX E: SEMI-SPONTANEOUS PICTURE DESCRIPTION TASK

Semi-Spontaneous Picture Description Task

In the following task, you will see two pictures side by side. You will be asked to write one sentence describing the two pictures you see, using the verb in the infinitive in parenthesis as a guide to create your sentence. Your sentences must include ‘si’ (either in the first or second clause). Try to be as specific and detailed as possible in your sentence.

Scenario: Melissa transferred from Florida International University (FIU) in Miami, Florida to Georgetown University in Washington, D.C. two months ago. She initially had a difficult time assimilating to her new life, and now that she has, she is reflecting on how some tough moments could have been different during her first two months.

1. Si no

(estar distraída) (poner)
2. **Si**

(comprar)  
(tener suficiente)
3. **Si no**

__________________________________________

__________________________________________

__________________________________________
4. Si no
5. **Si**

________________________________________________________________________________________

________________________________________________________________________________________

________________________________________________________________________________________
6. **Si**

   ____________________________________________________________

   ____________________________________________________________

   ____________________________________________________________
(tener dolor de cabeza)  

(beber vino)  

7.  

si
8. **Si no**
9. Sí

________________________________________________________________________________________

________________________________________________________________________________________

________________________________________________________________________________________
(tener frío)  

10. ________________________________________________________________________________

____________________________________________________________________________________

(ponerse)  

si

____________________________________________________________________________________
Semi-Spontaneous Picture Description Task

In the following task, you will see two pictures side by side. You will be asked to write one sentence describing the two pictures you see, using the verb in the infinitive tense in parenthesis as a guide to create your sentence. Your sentences must include ‘si’ (either in the first or second clause). Try to be as specific and detailed as possible in your sentence.

Scenario: Melissa transferred from Florida International University (FIU) in Miami, Florida to Georgetown University in Washington, D.C. two months ago. She initially had a difficult time assimilating to her new life, and now that she has, she is reflecting on how some tough moments could have been different during her first two months.

1. ________________________________________________________________________________________________
_________________________________________________________________________________________________

si
2. Si no

_____________________________________________________________________________________________
_____________________________________________________________________________________________
_____________________________________________________________________________________________
_____________________________________________________________________________________________
3. **Si no**

_________________________________________________________________________________________

_________________________________________________________________________________________

_________________________________________________________________________________________
(arruinar)  (encarcelar)

4. **Si no**

_____________________________________________________________________________________________
_________________________________________________________________________________________________
_____________________________________________________________________________________________
_________________________________________________________________________________________________
5. Si

_______________________________

_______________________________

_______________________________

_______________________________
6. Si no

_____________________________________________________________________________________________

_____________________________________________________________________________________________

_____________________________________________________________________________________________

_____________________________________________________________________________________________
7. Si

____________________________________________________________________________________________

____________________________________________________________________________________________

____________________________________________________________________________________________

____________________________________________________________________________________________
8. Si no
_____________________________________________________________________________________________
_____________________________________________________________________________________________
_____________________________________________________________________________________________
_____________________________________________________________________________________________
9. ____________________________
   ____________________________
   ______________________________________________________
   ____________________________
   si
   ______________________________________________________
10. ________________________________________________________________________________________________

si no ________________________________________________________________________________________________
APPENDIX F: THINK-ALOUD PRACTICE

For Miami Participants

In this experiment, I want to hear your thoughts, so I want you to think your thoughts aloud throughout the experiment into the microphone. You need to think aloud starting with the Think-Aloud practice task and then throughout the experiment. That means during the task and during the questions about the task, until the end of the experiment. In other words, once you start, don’t stop thinking aloud. Don’t plan what to say; just let your thoughts speak for themselves.

I would like you to talk CONSTANTLY, CLEARLY, and LOUDLY. Please take your time. Think aloud in the language you feel most comfortable (i.e. English or Spanish).

Please practice thinking aloud with the following task.

Think aloud, that is, verbalize aloud or say whatever comes to your mind while you perform this task. You don’t need to explain what you are saying just talk aloud as you do the problem solving task below:

You invite your friend to go to El Rey de las Fritas after class. You order a hamburger with a side of fries. Your friend orders a chicken sandwich and an order of onion rings. For dessert, you each have coffee and split one slice of pie at the end. With 10% tax and 20% tip, what was the total cost of the meal?

OK, START THINKING ALOUD NOW AS YOU WORK OUT THE COST BELOW:
For Washington, D.C. participants

In this experiment, I want to hear your thoughts, so I want you to think your thoughts aloud throughout the experiment into the microphone. You need to think aloud starting with the Think-Aloud practice task and then throughout the experiment. That means during the task and during the questions about the task, until the end of the experiment. In other words, once you start, don’t stop thinking aloud. Don’t plan what to say; just let your thoughts speak for themselves.

I would like you to talk CONSTANTLY, CLEARLY, and LOUDLY. Please take your time. Think aloud in the language you feel most comfortable (i.e. English or Spanish).

Please practice thinking aloud with the following task.

Think aloud, that is, verbalize aloud or say whatever comes to your mind while you perform this task. You don’t need to explain what you are saying just talk aloud as you do the problem solving task below:

You invite your friend to go to Leo’s after class on Georgetown Day. You order a hamburger with a side of fries. Your friend orders a chicken sandwich and an order of onion rings. For dessert, you each have coffee and split one slice of pie at the end. With 10% tax and 20% tip, what was the total cost of the meal?

OK, START THINKING ALOUD NOW AS YOU WORK OUT THE COST BELOW:

<table>
<thead>
<tr>
<th>#</th>
<th>Item(s)</th>
<th>Item Cost</th>
<th>Total Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hamburger</td>
<td>$4.75</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Chicken Sandwich</td>
<td>$4.75</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Caesar Salad</td>
<td>$5.50</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Cheese Crisp</td>
<td>$5.25</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Soup</td>
<td>$3.10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>French Fries</td>
<td>$1.90</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Onion Rings</td>
<td>$2.25</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Fruit Plate</td>
<td>$2.50</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Fresh Vegetables</td>
<td>$2.50</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Soda</td>
<td>$1.50</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Iced Tea</td>
<td>$1.50</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Coffee</td>
<td>$5.75</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Latte</td>
<td>$3.80</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Milk Shake</td>
<td>$3.25</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Sundaes</td>
<td>$4.25</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Pie</td>
<td>$3.75</td>
<td></td>
</tr>
</tbody>
</table>

Sub-Total: 

Tax: 

Tip: 

TOTAL: 

238
APPENDIX G: EXPERIMENT WITH ANSWERS

Scenario 1: Jude is a university student that lives with his sister, Allison, and her son, Logan. Jude has a very short temper. It causes him to act out impulsively out of anger when something doesn’t go his way, and consequently he gets himself into a lot of trouble.

Si Jude **TUVIERA** (tener) más paciencia, no sería tan impulsivo. Si Allison no se **HUBIERA COMIDO** (comer) el último pedazo de chocolate, Jude no habría tirado el refrigerador al piso. Si no **HUBIERA SIDO** (ser) el chocolate que se le iba a regalar a su novia, a Jude no le habría importado tanto. Jude **PASÓ** (pasar) mucho tiempo la semana pasada específicamente buscando ese chocolate.

Si Jude **HUBIERA ABIERTO** (abrir) la puerta del refrigerador antes de tirarlo al piso, habría visto la torta de helado que compró Allison para el cumpleaños de Logan. Como el año pasado a Logan no le gustó la torta regular, Allison **ESCOGIÓ** (escoger) una torta de helado este año. Si Jude no **HUBIERA ROTO** (romper) el refrigerador al tirarlo, la torta no se habría derretido. Si Logan **HUBIERA VISTO** (ver) su torta de helado en el piso, habría llorado. Si no **ESTUVIERA** (estar) tan lejos el supermercado de su casa, Jude iría corriendo para comprarle una nueva torta de helado a su sobrino. Jude compraría la torta de helado en el 7-11 al lado de su casa si se **ENCONTRARA** (encontrar) ahí. Si Jude **COMPRARA** (comprar) otra torta de 7-11 sin helado, Logan se daría cuenta inmediatamente. Si Jude **MANEJARA** (manejar), podría ir al supermercado en carro. Si él **HUBIERA LLAMADO** (llamar) un taxi más temprano, habría podido ir al supermercado antes de que su novia llegara a su casa. Si la novia de Jude no **HUBIERA TENIDO** (tener) un coche, él no habría podido ir a buscarle una nueva torta a su sobrino. Si Jude y su novia no **HUBIERAN IDO** (ir) al supermercado para comprar la torta, no habrían tenido que cancelar su cita.
romántica. ¡Menos mal que el supermercado no cerró (cerrar) hasta después de que compraron la nueva torta! La fiesta de Logan habría sido arruinada si Jude hubiera vuelto (volver) a casa sin la torta nueva. Si Jude no hubiera resuelto (resolver) esto antes de que se despertara Logan, su sobrino se habría enterado de lo que hizo su tío. Si él pudiera (poder) pedirle perdón a todos quienes fueron afectados por su mal carácter, lo haría en un instante. Pero al final del día, no lo hizo (hacer) porque él tuvo mucha vergüenza (embarrassment) de lo que pasó.

**Scenario 2:** Ángela, a University of Miami student, forgot that she had a major test in her Spanish class last Monday. Now she realizes the consequences of forgetting about her exam, and of not studying for it.

Si Ángela hubiera escrito (escribir) la fecha del examen en su agenda, se habría recordado que tenía que estudiar en vez de ir a una fiesta. Su compañera de cuarto le habría recordado de su examen si Ángela le hubiera dicho (decir) a su compañera de cuarto que iba a salir a una fiesta. Pero, cómo no le comentó (comentar) nada en esta ocasión, su compañera de cuarto no se lo recordó. De cualquier modo, habría podido estudiar si no hubiera bebido (beber) tanto durante las fiestas. La última vez que Ángela tomó (tomar) mucho, tuvo un gran dolor de cabeza, y durmió mucho. Si Ángela hubiera estudiado (estudiar) para su examen durante el fin de semana, no habría sacado una mala nota. Si hubiera hecho (hacer) la guía de estudio antes de ir a la fiesta, tal vez habría sacado una mejor nota. Si hubiera puesto (poner) la alarma la noche anterior, no se habría despertado tarde el día del examen. Se habría despertado más temprano sin su alarma si no hubiera cubierto (cubrir) la ventana con una sábana. Si hubiera vivido (vivir) en un dormitorio más cerca a la universidad, habría llegado a tiempo...
para el examen. Si no **SE OLVIDARA** (olvidarse) de sus responsabilidades escolares, tal vez sería aceptada a un programa de estudiar el extranjero. Si ella **QUISIERA** (querer) estudiar en Argentina, necesitaría una aplicación con buenas notas y una buena carta de recomendación. Si Ángela **PRODUJERA** (producir) una buena nota en la clase, su profesora le escribiría una buena carta de recomendación para su aplicación de estudiar en el extranjero. Pero, cómo no **ESTUDIÓ** (estudiar) lo suficiente, la profesora no se la escribió. Si **Ángela PRODUJERA** (producir) una buena nota en la clase, su profesora le escribiría una buena carta de recomendación para su aplicación de estudiar en el extranjero. Pero, cómo no **ESTUDIÓ** (estudiar) lo suficiente, la profesora no se la escribió. Si **HUBIERA ASISTIDO** (asistir) a las sesiones con el tutor de español, la habría ayudado a sacar mejores notas en la clase. El comité la aceptaría al programa si su aplicación para el programa de estudiar en el extranjero **FUERA** (ser) mejor. Le iría mejor en la clase si **PRESTARA** (prestar) más atención a los anuncios de su profesora de español. Además, si no **HUBIERA SACADO** (sacar) una mala nota en su clase, no habría perdido la beca. Si **CONTINUARA** (continuar) con su beca, Ángela no tendría que trabajar para pagar la matrícula. Este examen le **ENSEÑÓ** (enseñar) a Ángela sobre la importancia de no salir a fiestas cuando tienes que estudiar.
APPENDIX H: CONTROL GROUP COMPREHENSION TASK

**Scenario 1:** Ángela, a University of Miami student, forgot that she had a major test in her Spanish class last Monday. Now she realizes the consequences of forgetting about her exam, and of not studying for it.

Si Ángela hubiera escrito la fecha del examen en su agenda, se habría recordado que tenía que estudiar en vez de ir a una fiesta. Su compañera de cuarto le habría recordado de su examen si Ángela le hubiera dicho a su compañera de cuarto que iba a salir a una fiesta. Pero, cómo no le comentó nada en esta ocasión, su compañera de cuarto no se lo recordó. De cualquier modo, habría podido estudiar si no hubiera bebido tanto durante las fiestas. La última vez que Ángela tomó mucho, tuvo un gran dolor de cabeza, y durmió mucho. Si Ángela hubiera estudiado para su examen durante el fin de semana, no habría sacado una mala nota. Si hubiera hecho la guía de estudio antes de ir a la fiesta, tal vez habría sacado una mejor nota. Si hubiera puesto la alarma la noche anterior, no se habría despertado tarde el día del examen. Se habría despertado más temprano sin su alarma si no hubiera cubierto la ventana con una sábana. Si hubiera vivido en un dormitorio más cerca a la universidad, habría llegado a tiempo para el examen. Si no se olvidaría de sus responsabilidades escolares, tal vez sería aceptada a un programa de estudiar el extranjero. Si ella quisiera estudiar en Argentina, necesitaría una aplicación con buenas notas y una buena carta de recomendación. Si Ángela produjera una buena nota en la clase, su profesora le escribiría una buena carta de recomendación para su aplicación de estudiar en el extranjero. Pero, cómo no estudió lo suficiente, la profesora no se la escribió. Si hubiera asistido a las sesiones con el tutor de español,
la habría ayudado a sacar mejores notas en la clase. El comité la aceptaría al programa si su aplicación para el programa de estudiar en el extranjero fuera mejor. Le iría mejor en la clase si prestara más atención a los anuncios de su profesora de español. Ademá, si no hubiera sacado una mala nota en su clase, no habría perdido la beca. Si continuara con su beca, Ángela no tendría que trabajar para pagar la matrícula. Este examen le enseñó a Ángela sobre la importancia de no salir a fiestas cuando tienes que estudiar.

Preguntas de comprensión:

1) ¿Por qué es importante que Ángela reciba buenas notas en su clase de español?
2) ¿Qué papel tuvo la compañera de cuarto de Ángela en este cuento?
3) ¿Cuál fue el resultado de Ángela perdiendo su beca?
4) ¿Cuál crees que fue el evento más importante que pasó/no pasó?
5) ¿Cómo cambiaría ese evento (en la pregunta #4) el resultado del cuento?
**Scenario 2:** Jude is a university student that lives with his sister, Allison, and her son, Logan. Jude has a very short temper. It causes him to act out impulsively out of anger when something doesn’t go his way, and consequently he gets himself into a lot of trouble.

Si Jude tuviera más paciencia, no sería tan impulsivo. Si Allison no se hubiera comido el último pedazo de chocolate, Jude no habría tirado el refrigerador al piso. Si no hubiera sido el chocolate que se le iba a regalar a su novia, a Jude no le habría importado tanto. Jude pasó mucho tiempo la semana pasada específicamente buscando ese chocolate.

Si Jude hubiera abierto la puerta del refrigerador antes de tirarlo al piso, habría visto la torta de helado que compró Allison para el cumpleaños de Logan. Como el año pasado a Logan no le gustó la torta regular, Allison escogió una torta de helado este año. Si Jude no hubiera roto el refrigerador al tirarlo, la torta no se habría derretido. Si Logan hubiera visto su torta de helado en el piso, habría llorado.

Si no estuviera tan lejos el supermercado de su casa, Jude iría corriendo para comprarle una nueva torta de helado a su sobrino. Jude compraría la torta de helado en el 7-11 al lado de su casa si se encontrara ahí. Si Jude comprara otra torta de 7-11 sin helado, Logan se daría cuenta inmediatamente. Si Jude manejara podría ir al supermercado en carro. Si él hubiera llamado un taxi más temprano, habría podido ir al supermercado antes de que su novia llegara a su casa.

Si la novia de Jude no hubiera tenido un coche, él no habría podido ir a buscarle una nueva torta a su sobrino. Si Jude y su novia no hubieran ido al supermercado para comprar la torta, no habrían tenido que cancelar su cita romántica. ¡Menos mal que
el supermercado no cerró hasta después de que compraron la nueva torta! La fiesta de Logan habría sido arruinada si Jude hubiera vuelto a casa sin la torta nueva. Si Jude no hubiera resuelto esto antes de que se despertara Logan, su sobrino se habría enterado de lo que hizo su tío. Si él pudiera pedirle perdón a todos quienes fueron afectados por su mal carácter, lo haría en un instante. Pero al final del día, no lo hizo porque él tuvo mucha vergüenza (*embarrassment*) de lo que pasó.

Preguntas de comprensión:

1) ¿Por qué fue importante que Jude comprara una nueva torta?

2) ¿Qué papel tuvo Allison en este cuento?

3) ¿Cuál fue el resultado de la novia de Jude teniendo un coche?

4) ¿Cuál crees que fue el evento más importante que pasó/no pasó?

5) ¿Cómo cambiaría ese evento (en la pregunta #4) el resultado del cuento?
APPENDIX I: DELE TEST FOR PROFICIENCY

Name ______________________________

**Multiple Choice Test**

Each of the following sentences contains a blank indicating that a word or phrase has been omitted. Select the choice that best completes the sentence.

1. Al oír del accidente de su buen amigo, Paco se puso ________.
   a. alegre  b. fatigado  c. hambriento  d. desconchado

2. No puedo comprarlo porque me__________.
   a. falta  b. dan  c. presta  d. regalan

3. Tuvo que guardar cama por estar ____________.
   a. enfermo  b. vestido  c. ocupado  d. parado

4. Aquí está tu café, Juanito. No te quemes, que está muy ____________.
   a. dulce  b. amargo  c. agrio  d. caliente

5. Al romper los anteojos, Juan se asustó porque no podía ________ sin ellos.
   a. discurrir  b. oír  c. ver  d. entender

6. ¡Pobrecita! Está resfriada y no puede ____________.
   a. salir de casa  b. recibir cartas  c. respirar con pena  d. leer las noticias

7. Era una noche oscura sin ____________.
   a. estrellas  b. camas  c. lágrimas  d. nubes

8. Cuando don Carlos salió de su casa, saludó a un amigo suyo: -Buenos días,_
   a. ¿Qué va?  b. ¿Cómo es?  c. ¿Quién es?  d. ¿Qué tal?

9. ¡Qué ruido había con los gritos de los niños y el ____________ de los perros!
   a. olor  b. sueño  c. hambre  d. ladrar
10. Para saber la hora, don Juan miró el _______.
   a. calendario    b. bolsillo    c. estante    d. despertador
11. Yo, que comprendo poco de mecánica, sé que el auto no puede funcionar sin _______.
   a. permiso       b. comer       c. aceite     d. bocina
12. Nos dijo mamá que era hora de comer y por eso _______.
   a. fuimos a nadar   b. tomamos asiento  c. comenzamos a fumar  
     d. nos acostamos pronto
13. ¡Cuidado con ese cuchillo o vas a ______ el dedo!
   a. cortarte       b. torcerte     c. comerte      d. quemarte
14. Tuvo tanto miedo de caerse que se negó a _______ con nosotros.
   a. almorzar      b. charlar       c. cantar       d. patinar
15. Abrió la ventana y miró: en efecto, grandes lenguas de _______ salían
     llameando de las casas.
   a. zorros       b. serpientes   c. cuero       d. fuego
16. Compró ejemplares de todos los diarios pero en vano. No halló _______.
   a. los diez centavos b. el periódico perdido   c. la noticia que deseaba  
     d. los ejemplos
17. Por varias semanas acudieron colegas del difunto profesor a _______ el dolor de la viuda.
   a. aliviar       b. dulcificar   c. embromar     d. estorbar
18. Sus amigos pudieron haberlo salvado pero lo dejaron _____________.
   a. ganar        b. parecer      c. perecer      d. acabar
19. Al salir de la misa me sentía tan caritativo que no pude menos que _______ a un pobre mendigo que había allí sentado.
   a. pegarle      b. darle una limosna  c. echar una mirada  d. maldecir
20. Al lado de la Plaza de Armas había dos limosneros pidiendo ___________.
   a. pedazos   b. paz   c. monedas   d. escopetas

21. Siempre maltratado por los niños, el perro no podía acostumbrarse a ________ de sus nuevos amos.
   a. las caricias   b. los engaños   c. las locuras   d. los golpes

22. ¿Dónde estará mi cartera? La dejé aquí mismo hace poco y parece que el necio de mi hermano ha vuelto a ________.
   a. dejármela   b. deshacérmela   c. escondérmela   d. acabármela

23. Permaneció un gran rato abstraído, los ojos clavados en el fogón y el pensamiento ____________.
   a. en el bolsillo   b. en el fuego   c. lleno de alboroto   d. Dios sabe dónde

24. En vez de dirigir el tráfico estabas charlando, así que tú mismo ________ del choque.
   a. sabes la gravedad   b. eres testigo   c. tuviste la culpa   d. conociste a las víctimas

25. Posee esta tierra un clima tan propio para la agricultura como para ________ ________.
   a. la construcción de trampas   b. el fomento de motines   c. el costo de vida   d. la cría de reses

26. Aficionado leal de obras teatrales, Juan se entristeció al saber _____________ del gran actor.
   a. del fallecimiento   b. del éxito   c. de la buena suerte   d. de la alabanza

27. Se reunieron a menudo para efectuar un tratado pero no pudieron ________ ________.
   a. desavenirse   b. echarlo a un lado   c. rechazarlo   d. llevarlo a cabo

28. Se negaron a embarcarse porque tenían miedo de ________.
   a. los peces   b. los naufragios   c. los faros   d. las playas
29. La mujer no aprobó el cambio de domicilio pues no le gustaba _______.
   a. el callejero  b. el puente  c. esa estación  d. aquel barrio

30. Era el único que tenía algo que comer pero se negó a _____________.
   a. hojearlo  b. ponérselo  c. conservarlo  d. repartirlo

**Cloze Test**

In the following text, some of the words have been replaced by blanks numbered 1 through 20. First, read the complete text in order to understand it. Then reread it and choose the correct word to fill each blank from the answer sheet. Mark your answers by circling your choice on the answer sheet, not by filling in the blanks in the text.

**El sueño de Joan Miró**

Hoy se inaugura en Palma de Mallorca la Fundación y Joan Miró, en el mismo lugar en donde el artista vivió sus últimos treinta y cinco años. El sueño de Joan Miró se ha (1). Los fondos donados a la ciudad por el pintor y su esposa en 1981 permitieron que el sueño se (2); más tarde, en 1986, el Ayuntamiento de Palma de Mallorca decidió (3) al arquitecto Rafael Moneo un edificio que (4) a la vez como sede de la entidad y como museo moderno. El proyecto ha tenido que (5) múltiples obstáculos de carácter administrativo. Miró, coincidiendo (6) los deseos de toda su familia, quiso que su obra no quedara expuesta en ampulosos panteones de arte o en (7) de coleccionistas acaudalados; por ello, en 1981, creó la fundación mallorquina. Y cuando estaba (8) punto de morir, donó terrenos y edificios, así como las obras de arte que en ellos (9)

El edificio que ha construido Rafael Moneo se enmarca en (10) se denomina “Territorio Miró”, espacio en el que se han (11) de situar los distintos edificios que constituyen la herencia del pintor.

El acceso a los mismos quedará (12) para evitar el deterioro de las obras. Por otra parte, se (13), en los talleres de grabado y litografía, cursos (14) las distintas técnicas de estampación. Estos talleres también se cederán periódicamente a distintos artistas contemporáneos, (15) se busca que el “Territorio Miró” (16) un centro vivo de creación y difusión del arte a todos los niveles.

La entrada costará 500 pesetas y las previsiones dadas a conocer ayer aspiran (17) que el centro acoja a unos 150.000 visitantes al año. Los responsables esperan que la institución funcione a (18) rendimiento a
principios _______ (19) la __________ (20) semana, si bien el catálogo completo de las obras de la Fundación Pilar y Joan Miró no estará listo hasta dentro de dos años.

**Cloze Test Answer Sheet**

1. a. cumplido b. completado c. terminado
2. a. inició b. iniciara c. iniciaba
3. a. encargar b. pedir c. mandar
4. a. hubiera servido b. haya servido c. sirviera
5. a. superar b. enfrentarse c. acabar
6. a. por b. en c. con
7. a. voluntad b. poder c. favor
8. a. al b. en c. a
9. a. habría b. había c. hubo
10. a. que b. el que c. lo que
11. a. pretendido b. tratado c. intentado
12. a. disminuido b. escaso c. restringido
13. a. darán b. enseñarán c. dirán
14. a. sobre b. en c. para
15. a. ya b. así c. para
16. a. será b. sea c. es
17. a. casos b. aspectos c. niveles
18. a. a b. de c. para
19. a. total b. pleno c. entero
20. a. siguiente b. próxima c. pasada
APPENDIX J: EXPERIMENT + EXPLICIT SCREENSHOTS

Treatment 1

Oops, wrong! Try again. Think: does this sentence result in a present or future situation, or in a past one?

Please press the BACK button.

Treatment 1

¡Muy bien! The condition expressed in this sentence refers to the present or future. In sentences like these, you use the imperfect subjunctive— which is formed by taking the 3rd person plural [the ellos form] preterit [the past tense] form of the verb, and replacing the -ron with -ra.

Please press the NEXT button.
APPENDIX K: EXPERIMENT – EXPLICIT SCREENSHOTS

IF

Ooops, wrong! Try again.
Please press the BACK button.

[Progress bar at 3%]

IF

Muy bien!
Please press the NEXT button.

[Progress bar at 6%]
Debriefing Questionnaire

Participant # ____ / Initials: _____

You have just participated in a study of a grammatical form in Spanish. The target grammatical forms were the imperfect and pluperfect subjunctive used in contrary-to-fact conditional sentences.

The **imperfect subjunctive** is used when the sentence is *improbable but not impossible*. Most importantly, the main clause of the sentence (the second part), results in the present or future tense.

For example: Si comiera más vegetales, tal vez tendría el colesterol más bajo. (If I would have eaten more vegetables, maybe my cholesterol would have been lower.)

The **pluperfect subjunctive** is used when the sentence is *impossible*, because it is essentially talking about a hypothetical situation in the past, which *has already happened, and therefore cannot be changed*. The main clause of the sentence (the second part) results in the past.

For example: Si no **hubiera ido** al cine a las 11 anoche, no me habría despertado tan tarde esta mañana. (If I hadn’t gone to the movies at 11 last night, I wouldn't have woken up so late this morning.)

1. Please indicate your knowledge/recognition of this form before participating in this study by marking the appropriate answer below:

   _____ Yes, I recognized this form before doing the exercises.
   _____ Yes, I knew this form before doing the exercises.

   OR _____ No, I did not recognize this form before doing the exercises.

   OR _____ No, I did not know this form before doing the exercises.

2. During ANY of the sessions (between sessions 1 and 2, or between sessions 2 and 3), did you look up either of these structures in the textbook or online, or did you ask someone other than the researcher to explain them to you?

   _____ Yes, I looked up how to use these forms during the experimental sessions.

   OR _____ No, I did not look up these forms during the experimental sessions.

Thank you for your participation!
REFERENCES


Calderon, A. M. (2014). Level of intake, depth of processing, and type of linguistic item


256


Domínguez, L. (2009). Charting the route of bilingual development: Contributions from


through a computerized task. Unpublished doctoral dissertation, Georgetown University, Washington, DC.


*Studies in second language acquisition*, 22(04), 557-584.

Leow, R. P., Egi, T., Nuevo, A. M., Tsai, Y-C. (2003). The roles of textual enhancement And type of linguistic item in adult L2 learners’ comprehension and intake.

*Applied Language Learning*, 13, 1-16.


synthesis.


Mira, M. I. M. (2009). Spanish heritage speakers in the Southwest Factors contributing to the maintenance of the subjunctive in concessive clauses. *Spanish in context, 6*(1),


experience facilitates the processing of gender agreement in Spanish heritage speakers. *Bilingualism: Language and Cognition, 17*(01), 118-138.


Roca, A., & Colombi, M. C. (2003). Insights from research and practice in Spanish as a heritage Language. In A. Roca & M. C. Colombi (Eds.), *Mi lengua: Spanish as a*


Rothman, J., & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Applied linguistics, 35*(1), 93-98.


Silva-Corvalán, C. (2014). *Bilingual language acquisition: Spanish and English in the*
first six years. Cambridge University Press.


