THE EFFECTS OF TYPE OF FEEDBACK, AMOUNT OF FEEDBACK AND TASK-ESSENTIALNESS IN A L2 COMPUTER-ASSISTED STUDY

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

By

Maria Luisa Filgueras Gómez, M.S.

Washington, DC
August 26, 2016
THE EFFECTS OF TYPE OF FEEDBACK, AMOUNT OF FEEDBACK AND TASK-ESSENTIALNESS IN A L2 COMPUTER-ASSISTED STUDY

Maria Luisa Filgueras Gómez, M.S.

Thesis Advisor: Ronald P. Leow, Ph.D.

ABSTRACT

For decades SLA research has looked at learner-external factors as influencers of learning outcomes, one of them being computerized types of feedback, with non-conclusive results. Moreover, a potentially significant characteristic of feedback, amount of feedback, has often been overlooked in previous investigations (Li, 2010; Loewen, 2012). To date, only one study (Ellis, Sheen, Murakami, & Takashima, 2008) has addressed the effects of providing different amounts of feedback on the same error with results indicating no advantage for single or multiple corrections. In addition, while task-essential (TE) practice has been identified as a desirable and optimal feature when designing a task (Loschky & Bley-Vroman, 1993), and has been included in several studies, its isolated contribution has not yet been identified.

Within CALL, the present study seeks to fill these gaps by investigating the effects of type of feedback provided (i.e., “right/wrong” vs. rule explanation) during input-based practice (task-essential vs. non task-essential). Participants were randomly distributed in three groups: Rules [task-essential practice with rule explanation in the feedback], Right/Wrong [task-essential practice with right/wrong feedback], and NF [task-essential practice without feedback]. In addition, the amount of feedback was tracked in order to investigate whether there was a threshold from which participants started to require fewer feedback episodes to achieve accuracy. Finally, to determine the isolated contribution of
TE practice, a fourth group, NTE (Non Task-essential practice without feedback), was included. Results indicated that “right/wrong” feedback was more beneficial than rule explanation when interpreting temporal sentences at immediate post- and delayed posttests. The production of critical items at immediate post- and delayed posttests was similar for both feedback groups. As for amount of feedback, different types of feedback did not require different amounts, and the same amount of feedback led to L2 development regardless of feedback types. Interestingly, feedback was negatively correlated to development. Regarding the trajectory of the feedback during treatment, the learning process had the shape of an inverted U that showed a gradual decline in the need for feedback. Finally, with regard to the role of task-essentialness, NF and NTE groups performed similarly indicating no clear benefit of task-essentialness.
DEDICATION

A mi padre, a quien tanto quiero y tanto echo de menos
ACKNOWLEDGEMENTS

These acknowledgements need to be both in English and Spanish as my daily life is. First, I would like to thank my wonderful mentor, Ron Leow for taking me as his mentee and teaching me so many things along the way. Our mentor-mentee relationship happened by chance; looking back I think it was the best that could happen to me and I really believe it was meant to be. Ron, your patience, guidance and advice were always the best. You are a first class researcher and mentor. I’m also deeply grateful to the other members of my committee, Cristina Sanz, Lourdes Ortega and Luis Cerezo. Cristina, I have learnt greatly from your multifaceted roles as a teacher, a language program director, and a researcher, and I am very grateful to have had the opportunity of working closely with you in all these aspects. Thank you for all your help and advice. Lourdes, your perspective and scope as the exceptional scholar that you are have been of incalculable value and have helped me develop as a researcher. Luis I admire your dedication, detailed work and passion for research. I came to know more about your research through your dissertation and I learnt tremendously from you. Your detailed feedback, comments and input pushed me to rethink and question some things to ultimately, I believe, improve this research. You, as the most recent PhD in my committee, gave me important advice to finally complete my studies. I see in you the young scholar I want to be.

Quiero agradecer a mis amigos y colegas del departamento de español en Georgetown: Sergio, Charlie, John (mis compañeros de promoción, con quienes he compartido tanto), Germán y Julito (de quienes he aprendido tanto como investigadores y como compañeros), Jessie (tan amable y con quien es y ha sido un placer trabajar en
research y enseñando), Ellen (que me ha ayudado siempre tanto), Sarah Grey (cuya ética del trabajo, inteligencia y amabilidad para dedicarme unos minutos o mandarme algún artículo relevante me ha enseñado mucho). A mis compañeros e instructores que me permitieron recoger datos en sus clases: Allison Caras, Ross Karlan, Tim McCormick, Mª José Navia, Colleen Moorman, Ana Levenson. A mis “RAs” que me ayudaron a codificar los datos: Celia Zamora, Meagan Driver y Chrissy Blistine-Bonilla.

To my dearest Lauren, who is an amazing researcher and a wonderful friend. I enjoy every minute that I’m with her, either doing and talking about Linguistics or simply chatting about our lives. We have been in class together, have developed research studies together, have presented work together and have written papers together. We have also spent Christmas together and celebrated birthdays together :) I truly treasure our friendship.

To Nick…what can I say about Nick Pandža? He knows. I owe him big. I am truly very grateful for him. For his kindness to spend endless hours with me, teaching me an entire new world of Statistics; for being with me until pretty late at night (11pm…); for his patience with my slow understanding and sometimes poorly-worded questions. I hope he truly knows what an integral role he has played in making this study possible. And along the way, we became closer friends.

To my Michigan family (Rod, Linda, Joel, my sis Missy, and Zach) who has always supported us and helped us to achieve whatever we wanted to pursue.

A mi otra familia de Michigan, a mis annarboritas, a quienes conocí hace más de 15 años y con quienes he pasado momentos inolvidables, muy buenos y algunos un poco menos. Mis annarboritas, que no dudaron en ayudarme en momentos muy difíciles para
Les doy las gracias porque aunque no podemos vernos mucho, siempre es como si nunca me hubiera ido, como si siguiera allí. Y los echo mucho de menos. A mi Cal-li, Alberto, Susanna, Roberto, Dio, Cristina, Ricardo, Esther, Mara, Ana, Ofe, Mª Dolores.

A mi rubia, Bohumira. Empezamos casi juntas y hemos terminado casi juntas :) Y en el camino ha habido una boda, un nacimiento, varias fiestas, confesiones, llantos y muchas risas :) Tú, como sabes, eres una verdadera amiga, amable, completamente “selfless”, dispuesta a todo por su gente :) Nos hemos ayudado en todo, dentro y fuera de la universidad y espero que siga siendo así para siempre. Eres de lo más valioso que me llevo de Washington.

To Dan and Josephine, my friends and partners in the adventure of being a parent while doing a PhD… We got so excited when we heard we were going to be parents about the same time! And we have shared so much about childcare, child rising, Linguistics, languages, and life goals. And our kids are best buddies! What else could we ask for?

A mi mamá, mi madre. Al principio de empezar el doctorado fue muy difícil para nosotros, nos faltaba un pilar muy importante en nuestras vidas: mi padre y tu compañero de vida. Con mucho cariño, paciencia y amor, poco a poco pudimos ir disfrutando de DC en tus venidas cuando yo recibía extra mimos :) Y cuando Nico llegó, nos trajo una felicidad inmensa. Y tú te dedicaste a cuidarnos a él y a mí, y cuando yo tenía que seguir con este monstruo al que llamamos tesis, tú estabas allí cuidando a Nico entregada en cuerpo y alma. Si no hubiera sido por ti, por tu generosidad, dedicación y entrega con Nico, todavía estaría escribiendo la tesis! Te agradezco de corazón todo lo que has hecho y sigues haciendo por nosotros. Y te quiero mucho.
A mi hermano y Natalia “mi cuñí”. Hemos pasado mucho juntos, nos hemos reído y hemos llorado juntos y siempre nos hemos apoyado. Cuando no hemos podido ir a España, aquí vinisteis sin pensarlo dos veces a seguir juntos, a disfrutar como titos de Nico :) Para escribir una tesis hay que mantener cierta cordura y serenidad y vosotros me habéis ayudado a eso.

A mis amigas, a “las niñas”, mis amigas de toda la vida. Cada una vivimos en una ciudad diferente, en un país diferente, pero ahí seguimos, juntas por más de 15-20 años. En este tiempo que he pasado haciendo la tesis, hemos pasado de todo juntas: ha habido bodas, fallecimientos, nacimientos, “coming outs”, risas y llantos, y sobre todo y siempre apoyo, mucho apoyo. Sin vosotras mi vida no sería mi vida, yo no sería quien soy. Y os agradezco de corazón que sigáis ahí, que sigamos ahí, sin importar las distancias, haciendo que nuestra amistad sobreviva todo.

A mi Nico. El sol de mi vida. Tú encarnas la felicidad y tu pura existencia, incluso cuando eras tan chiquitito como un granito de arroz en mi barriga, me llena, centra todo y me da perspectiva. Gracias por existir, por ser como eres. Que incluso tan chiquitito y joven como eres, ya me dices, “mamá, ¿tú a trabajar?”

And finally to Chad, my love, my best friend, my partner in crime. Without you, this dissertation would have never been written, really. You are so selfless and kind, you have always always supported me in anything I have ever wanted to do. In this journey of grad school, you have always been there, you have heard me and advised me on what to do or not to do. In every bit of the way, in every little detail and in any big detail. You have been so patient when I was spending very long nights working, when you listened carefully to my ideas (yes, Linguistics ideas), when you cared for our Nico so that I could
keep working. You have taught me so much here in graduate school and in all other
aspects of our life together. I am so lucky to be able to spend my life with such a
wonderful human being. And I love you forever.
## TABLE OF CONTENTS

**CHAPTER 1: INTRODUCTION**

The Statement of the Problem ................................................................. 1
Potential Contributions ............................................................................. 6
Definition of Terms .................................................................................. 8
  E-tutor ..................................................................................................... 8
  Feedback ................................................................................................. 8
  Rule explanation feedback .................................................................... 8
  “Right/wrong” feedback ......................................................................... 9
  Amount of feedback .............................................................................. 9
  Practice .................................................................................................. 9
  Task-essentialness ................................................................................. 9

**CHAPTER 2: REVIEW OF THE RELEVANT LITERATURE**

Feedback in SLA ...................................................................................... 10
  Terminology .......................................................................................... 10
  Modes and types ................................................................................... 11
  Amount of feedback ............................................................................ 16
  Summary of feedback section ............................................................... 17

Empirical Findings on Type of Feedback in CALI Studies ......................... 18
  Rules in the input .................................................................................. 19
  Rules provided only in the feedback ................................................... 25

Empirical Findings on Amount of Feedback Studies .................................... 31
  Summary of empirical findings on amount of feedback studies section .... 36
Practice.........................................................................................................................38
  Summary of practice section ......................................................................................42
Empirical Findings on Task-essential Practice Studies .................................................43
  Summary of empirical findings on task-essential practice studies section ..........50
Empirical Findings on Adverbial Clauses on L2 Spanish ............................................52
  Summary of empirical findings on adverbial clauses on L2 Spanish section ......61
Research Questions ......................................................................................................63

CHAPTER 3: THE PRESENT STUDY ..............................................................................65
Method ............................................................................................................................65
  Participants ................................................................................................................65
  Targeted structure .......................................................................................................68
Experimental Design ....................................................................................................73
  L2 development .........................................................................................................73
  Type and amount of feedback ..................................................................................74
  Task-essentialness .....................................................................................................76
Materials ........................................................................................................................78
  Language background questionnaire .......................................................................78
  Review session ..........................................................................................................78
  Treatment ....................................................................................................................81
Assessments ....................................................................................................................88
  Picture-matching interpretation test .........................................................................88
  Controlled written production test ...........................................................................90
  Exit questionnaire ......................................................................................................91
Procedure ..............................................................................................................................92
Coding and Scoring ..................................................................................................................93
CHAPTER 4: ANALYSES AND RESULTS ..............................................................................95
Statistical Analyses ..................................................................................................................95
Results .......................................................................................................................................96
  Research question #1 .............................................................................................................96
  Research question #2 ..........................................................................................................117
  Research question #3 ..........................................................................................................127
CHAPTER 5: DISCUSSIONS AND CONCLUSIONS .................................................................136
Discussion ................................................................................................................................136
  Research question 1 .............................................................................................................136
  Research question 2 ..........................................................................................................141
  Research question 3 ..........................................................................................................150
Conclusions ..............................................................................................................................152
Limitations and Future Research ............................................................................................157
APPENDIX A: Rules in the input ..........................................................................................161
APPENDIX B: Rules only in the feedback ..............................................................................165
APPENDIX C: Amount of feedback studies .........................................................................170
APPENDIX D: Review session ...............................................................................................173
APPENDIX E: Review session answer sheet .........................................................................202
APPENDIX F: Stimuli of the treatment ..................................................................................206
APPENDIX G: Stimuli of the interpretation test ....................................................................256
APPENDIX H: Stimuli of the production test .........................................................................301
LIST OF FIGURES

Figure 1: Example of feedback with rules [R] after an incorrect answer of an item referring to the present ................................................................. 75

Figure 2: Example of feedback with rules [R] after an incorrect answer of an item referring to the future ................................................................. 75

Figure 3: Example of right/wrong feedback condition [Y/N] ................................................................. 75

Figure 4: Relationship among four basic dimensions of structure-based tasks (Loschky & Bley-Vroman, 1993, p. 142) ................................................................. 77

Figure 5: Excerpt of the review session asking participants to locate in time the actions of the sentence ................................................................. 80

Figure 6: Excerpt of the review session showing participants an example of the format of the picture with present indicative with future value ................................................................. 80

Figure 7: TE condition in treatment with targeted items referring to the future ......................... 83

Figure 8: TE condition in treatment with counterpart items referring to the present ...... 83

Figure 9: TE condition in treatment with distractor items referring to the past ................. 84

Figure 10: Design of the treatment presentation ................................................................. 85

Figure 11: Example of stimulus of the NTE condition with items referring to the future 86

Figure 12: Example of stimulus of the NTE condition with items referring to the present ................................................................. 87

Figure 13: Example of stimulus of the NTE condition with items referring to the past.... 87

Figure 14: Example of assessment test with items referring to the future ......................... 88

Figure 15: Example of assessment test with items referring to the present ...................... 89

Figure 16: Example of assessment test with items referring to the past ...................... 89

Figure 17: Example of the controlled written production test with future and present items ........................................................................................................ 91

Figure 18: Results from the interpretation test ................................................................. 102
Figure 19: Results from the written production test with subjunctive and present tense items........................................................................................................................................111

Figure 20: Amount of feedback predicted for each tense by block.................................................................119

Figure 21: Trajectory of the feedback predicted for each tense by trial.........................................................120

Figure 22: Results from the interpretation test ...............................................................................................................130

Figure 23: Results from the production test ...............................................................................................................133

Figure 24: Trajectory of the feedback predicted for each tense by trial .........................................................147
# LIST OF TABLES

Table 1: Participant background .......................................................................................... 66  
Table 2: Distribution of participants per condition ................................................................. 67  
Table 3: Characteristics of feedback per groups ................................................................. 76  
Table 4: Summary of the traits of each experimental condition ........................................... 78  
Table 5: Distribution of TE items ....................................................................................... 82  
Table 6: Summary of the distribution of temporal items in the interpretation task ............ 90  
Table 7: Summary of the distribution of temporal items in the production task ............... 91  
Table 8: Procedure ........................................................................................................... 93  
Table 9: Mean accuracy (SDs) in the Interpretation test with subjunctive items ............... 99  
Table 10: Results of logistic multilevel model modeling Accuracy on the Interpretation test ................................................................................................................. 100  
Table 11: Mean accuracy (SDs) for the interpretation data with present tense items ....... 104  
Table 12: Summary of the results of the interpretation test ................................................ 107  
Table 13: Mean accuracy (SDs) in the Production test with subjunctive items ............... 108  
Table 14: Results of logistic multilevel model modeling Accuracy in the Production test ................................................................................................................. 110  
Table 15: Mean accuracy (SDs) from the production test with present tense items ......... 113  
Table 16: Summary of the results of the production test .................................................... 117  
Table 17: Mean amount of feedback received during treatment .......................................... 119  
Table 18: Results of quadratic multilevel model modeling Amount of feedback ........... 122  
Table 19: Predictions of the needed feedback .................................................................... 123  
Table 20: Results of correlations between feedback and assessment tests ..................... 125  
Table 21: Summary of the results of the Research Question 2 ........................................ 127
Table 22: Mean accuracy (SDs) in the interpretation test with subjunctive items ..........128

Table 23: Results of logistic multilevel model modeling Accuracy in the Interpretation test ..................................................................................................................................................129

Table 24: Summary of the results of the interpretation test .........................................................131

Table 25: Mean accuracy (SDs) in the Production test with subjunctive items ..........132

Table 26: Results of logistic multilevel model modeling Accuracy in the Production test ..................................................................................................................................................134

Table 27: Summary of the results of the production test .................................................................135

Table 28: Predictions of the needed feedback ..................................................................................147
CHAPTER 1: INTRODUCTION

The Statement of the Problem

We live in an era where technology has reached all aspects of our lives, including education. When it comes to learning another language, today’s technology provides several resources that include platforms and applications through the Internet (e.g., online chats, email, blogs, videos, etc.), and software that works as electronic instructors (e.g., Rosetta Stone). In formal settings of language instruction, universities and schools are enhancing – or even replacing – traditional face-to-face (FTF) instruction with computer-assisted language learning (CALL) (Cerezo, Baralt, Suh, & Leow, 2013). In the field of second language acquisition (SLA), the vast majority of research has involved FTF interactions, although investigations involving technology-mediated language learning have grown (e.g., Baralt, 2013; Bowles, 2008; Camblor, 2006; Cerezo, 2010; Hsieh, 2007; Loewen & Earlam, 2006; Medina, 2008; Moreno, 2007; Morgan-Short & Bowden, 2006; Rosa & Leow, 2004; Sachs & Suh, 2007; Sanz & Morgan-Short, 2004; Shekary & Tahiririan, 2006 to name a few), and promise to continue growing as SLA doctoral students rely more and more on technology (Sanz, Morales-Front, Zalbidea, & Zarate, 2015).

With regard to the effectiveness of technology-mediated language learning, it has been argued that CALL can be as effective as FTF instruction, if not more so (Grgurovic, Chapelle, & Shelley, 2013; Zhao, 2003). As acknowledged by Cerezo (2010), research using CALL technology has considered for the most part the computer as a medium of communication in asynchronous platforms (e.g., email, forums) and synchronous platforms (e.g., chat), and lately a growing number of studies have used the computer as a tutor, where, as described by Cerezo (2012) “learners work independently with technology without interacting with a teacher or a
peer” (p. 51) (e.g., in Spanish: Bowles, 2008; Camblor, 2006; Cerezo, 2010; Hsieh, 2007; Lin, 2009; Medina, 2008; Moreno, 2007; Morgan-Short & Bowden, 2006; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004; English: AbuSeileek, 2009; AbuSeileek & Rababah, 2007; Nutta, 1998; Petersen, 2010; Torlakovic & Deugo, 2004; Japanese: Nagata, 1993, 1996, 1998a, 1998b; Sachs, 2011; French, Watts, 1989; and Latin: Lado, Bowden, Stafford, & Sanz, 2013). This is not surprising since, from a practical point of view, current educational institutions are increasingly advocating for technology-enhanced curricula (Goertler, 2011). As pointed out by Cerezo (2012, p. 50), this could include technology-mediated components, where learners use the technology to interact with other humans, as well as technology-based components, where learners use technology to work independently. Interesting for the CALL strand of research is that the results of the empirical studies investigating the effectiveness of e-tutors against other instructional technologies have shown that CALI can successfully promote L2 grammar development (e.g., Watts (1989) comparing e-tutor vs. non-interactive videos; Nagata (1996) comparing e-tutor vs. instruction-based workbooks; Nutta (1998), Torlakovic & Deugo (2004), AbuSeileek & Rababah (2007), AbuSeileek (2009), and Petersen (2010) comparing e-tutor vs. FTF instruction).

Furthermore, as is argued to happen in CALL studies (Sagarra, 2007; Li, 2010), e-tutors may provide more salient and more consistent feedback than human teachers (Cerezo, 2012), which could arguably lead to a better quality of treatment.

The role of feedback in SLA has been extensively investigated in several empirical published studies (e.g., in FTF interactions: Ammar & Spada, 2006; Carroll & Swain, 1993; Carroll, Swain, & Roberge, 1992; DeKeyser, 1993; Ellis, Loewen, & Erlam, 2006; Ellis, 2007; Leeman, 2003; Mackey, 1999; Sheen, 2007; inter alia; in CALL: Baralt, 2013; Bowles, 2008; Loewen & Erlam, 2006; Morgan-Short, & Bowden, 2006; Nagata, 1993; Rosa & Leow, 2004;
Sagarra, 2007; Sanz & Morgan-Short, 2004; Sauro, 2009; Yilmaz, 2012, *inter alia*), as well as meta-analyses (e.g., Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), and the general finding is that feedback is positively related to language development. However, issues such as what type of feedback (more or less explicit), when to provide the feedback (immediately or delayed in the interaction), and how much feedback is more conducive to language learning are still being debated. The great majority of studies have considered the role of type of feedback in the context of FTF interactions (e.g., Carroll & Swain, 1993; Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Leeman, 2003; Loewen & Nabei, 2007; Sheen, 2007; *inter alia*). Some other studies have investigated type of feedback in CALL where the computer is used as a medium (e.g., Loewen & Erlam, 2006; Sauro, 2009; Yilmaz, 2012), and there are also several published studies (e.g., Bowles, 2008; Morgan-Short & Bowden, 2006; Nagata, 1993; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004, Stafford, Bowden, & Sanz, 2012; Lado, Bowden, Stafford, & Sanz, 2013) and dissertations (e.g., Camblor, 2006; Cerezo, 2010; Lin, 2009; Moreno, 2007; Sachs, 2011) that have examined the role of type of feedback in e-tutors. The results from these studies are not conclusive. While some of them have found 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 between both types of feedback in immediate posttests (e.g., Camblor, 2006; Hsieh, 2007; Sanz & Morgan-Short, 2004) and delayed posttests (e.g., Camblor, 2006; Hsieh, 2007). Only one study found no significant difference on the immediate posttest but an advantage for less explicit feedback in oral delayed tests (Moreno, 2007).

While the study of type of feedback has triggered an exponential amount of research in SLA, a potentially significant characteristic of feedback, namely *amount of feedback* has often
been overlooked in previous investigations (Li, 2010; Loewen, 2012). The studies that have investigated amount of feedback can be classified into two strands. The first strand includes descriptive studies where amount of feedback provided was not manipulated. These investigations examined how L2 FTF interactions have generated different amounts of feedback depending on type of task (e.g., problem-solving, spot the difference, and free conversations in Crookes & Rulon, 1988), learners’ characteristics (e.g., age in Oliver, 2000), contextual settings (e.g., dyads vs. teacher-fronted context in Oliver, 2000), interlocutor’s characteristics (e.g., native speakers vs. nonnative speakers in Mackey, Oliver, & Leeman, 2003), and type of error (e.g., Havranek, 2002). In the second strand amount of feedback was an independent variable. To my knowledge, Ellis, Sheen, Murakami, & Takashima (2008) is the only study that has manipulated the amount of feedback by providing a different number of feedback episodes on the same error to two different groups. The results of this study indicated that there was no advantage for single or multiple corrections of the same linguistic error. While the results of this study do seem to suggest that amount of feedback does not affect learning but rather just the provision of corrective feedback, more research is warranted, as this is only one study. Moreover, as suggested by Loewen (2012) “(g)iven that many negative feedback studies have targeted one or two English structures there is a need for additional studies of other languages and linguistic structures” (p.36).

Researchers in SLA have also been interested in creating materials that can provide extensive computerized practice. From a cognitive-interactionist approach, practice has been regarded as a prerequisite for interaction (DeKeyser, 2007), and it particularly is considered necessarily present during the learning process: “Indeed, one can choose to include rule presentation or feedback in a treatment, but both input and practice – task or activity in more
pedagogical terms—must be present” (Sanz & Morgan-Short, 2005, p. 237). Practice in SLA has been defined as “specific activities in the second language engaged in systematically, deliberately with the goal of developing knowledge of and skills in the second language” (DeKeyser, 2007, p. 8.). In form-focused instruction a particular instantiation of deliberate practice are tasks, and a desirable and optimal feature when designing a task is task-essentialness (TE) (Loschky & Bley-Vroman, 1993). Task-essentialness refers to the idea that structural accuracy is essential to the meaning in the task (p. 131). In order to successfully complete a communicative task, “it is essential to attend to the relevant structure” (p.138). Recently, some CALI studies have included TE in their designs (e.g., Cerezo, 2010; Hsieh, 2007; Lado et al., 2013; Moreno, 2007; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004, Stafford et al., 2012). In all of these studies, with the exception of Cerezo’s (2010) and Hsieh’s (2007), TE was provided along with feedback, thus the unique and contributive role of TE was not clear. One of the purposes of Hsieh’s (2007) and Cerezo’s (2010) dissertations was to examine the role of actively completing TE practice as opposed to passively being exposed to TE practice while receiving feedback (more explicit, less explicit or none). The results of both studies seem to indicate that it is the combination of active TE practice and feedback as opposed to TE practice only that promotes language learning. While these two empirical studies offer evidence that being actively engaged in TE practice is more beneficial than the mere exposure to it, the isolated role of TE practice compared to any other kind of practice has not yet been studied. The present dissertation seeks to fill this gap.

In light of all of the above, the present dissertation sought to fill the previously identified gaps by investigating relationships between corrective feedback and development of Spanish subordinate clauses during input-based practice by native speakers of English. Using computer-
assisted instruction, this study specifically investigated the effects of type of feedback provided (i.e., “right/wrong” [Cerezo, 2016] vs. rule explanation) during computerized input-based practice (task-essential vs. non task-essential), as well as the effects of taskessentialness. In addition to examining the type of feedback provided, the present study also tracked the amount of feedback provided by recording the number of feedback episodes needed to obtain the correct form. The goal of this was twofold: (1) to investigate whether there was a threshold from which learners started to require fewer feedback episodes to achieve accuracy; and (2) to investigate the correlational relationship between amount of feedback and L2 development.

Potential Contributions

The results of this dissertation promise to make both theoretical and practical contributions to the field of SLA and language pedagogy. From a theoretical point of view, psycholinguistic dimensions of SLA theory can be expanded by the study of the new types of interactions that computer-assisted language learning provides (Heift & Chapelle, 2012). The current dissertation is grounded in a cognitive-interactionist approach to adult SLA whose main goal is to assess the role of learner-internal and learner-external factors as influencers of language learning outcomes. As learner-external factors, this study looks at the effects of type of feedback with two different grammatical forms while completing task-essential practice. By implementing task-essential computerized practice that pushes learners to pay attention to the relevant form coupled with computerized feedback to help them test their hypotheses about how the L2 works, the current dissertation aims to expand the psycholinguistic dimensions of SLA theory.

In the framework of the Interactionist Approach (Gass & Mackey, 2007) several processes come to work (e.g., interaction –as a form of practice that provides opportunities to
make connections between form and meaning–feedback, negotiation for meaning, modified output), and research under this framework has provided evidence of positive effects of such processes in language learning (e.g., Ayuon, 2001; Leeman, 2003, Long, Inagaki, & Ortega, 1998; Mackey, 1999; Mackey, Gass, & McDonough, 2000; McDonough, 2001, 2005; inter alia.) The design of the present dissertation provides opportunities for interaction in the form of computerized practice and computerized feedback with the goal to advance our understanding of the effectiveness of input-based task-essential practice and different types of feedback. It also provides the opportunity to examine the isolated role of task-essentialness, something that no former study has done. Furthermore, although previous investigations have sought to discriminate what type of feedback with regards of degrees of explicitness may be more beneficial under certain learning conditions, no definitive conclusions have emerged. The present study seeks to contribute to this debate by investigating whether the possible benefits of different types of feedback may be affected by the grammatical form being taught. Moreover, another goal of this study is to examine whether grammatical forms that differ in degree of complexity require different amounts of feedback from learners before achieving accuracy of the target form. The results of this dissertation promise to advance our understanding of the effectiveness of feedback types and feedback quantity, the latter being a potentially significant characteristic of feedback that has been previously overlooked (Li, 2010; Loewen, 2012) in relation to linguistic complexity.

From a pedagogical and practical point of view, we cannot ignore the fact that more and more educational institutions are incorporating technology into their curricula with many seeking to develop hybrid programs where face-to-face instruction is supplemented with computer-assisted instruction (Cerezo et al., 2013). As Heift and Chapelle (2012) state, “the primary goal
for researchers in CALL is to better understand the new landscape with its implications for language teaching and impact on language learning” (p. 565). The findings from this line of research can guide material developers, teachers, and learners on maximizing the effectiveness of the computer as a language-learning tool. With this in mind, we seek to a) contribute to this line of research by advancing our understanding of the effectiveness of computer-assisted language instruction under certain conditions; and b) provide opportunities for learners that help them create valuable computer-learner interactions (Heift & Chapelle, 2012, p. 557).

**Definition of Terms**

**E-tutor**

E-tutor is also referred to as Computer Assisted Language Instruction (CALI). Following Cerezo (2012) an e-tutor is “software that, in addition to displaying language input, has the capability of engaging learners in some sort of pedagogical practice and evaluating their responses, parsing them as correct or incorrect, and reacting accordingly by, for example, providing modified input or corrective feedback” (p. 51). The present study used an e-tutor that provided task-essential tasks and corrective feedback tweaked to the needs of the learners.

**Feedback**

In the current dissertation feedback refers to the pedagogical information learners received regarding the choices they make when completing a task. It can be delivered in a more or less explicit way depending upon the degree of information that is provided about the target form.

**Rule explanation feedback**

In this study rule explanation refers to the written metalinguistic explanation that learners receive about the targeted structure.
“Right/wrong” feedback

The term “right/wrong” feedback (Cerezo, 2016) refers to the ‘correct’/ ‘incorrect’ answers (without grammatical information) to inform participants about the correctness of their choices.

Amount of feedback

In the present study, amount of feedback refers to the number of feedback episodes each learner requires before obtaining the correct form.

Practice

We considered practice as specific activities that are systematic and deliberate, and have the goal of developing the second language (DeKeyser, 2007). Particularly, the participants of the present study are engaged in receptive practice as defined by Leow (2007):

“any exposure to manipulated L2 input that provides not only many exemplars of targeted L2 forms or structures upon which learners’ attention to (and/or awareness of) is directly or indirectly premised but also some form of opportunity to perform a limited productive or nonproductive task or activity (e.g., selecting one out of two options, completing a problem-solving task, translating) during the exposure.” (p. 21-22)

Task-essentialness

Task-essentialness refers to the idea that structural accuracy is essential to the meaning in the task (Loschky & Bley-Vroman, 1993, p. 131). In order to successfully complete a communicative task, “it is essential to attend to the relevant structure” (p.138). Task-essentialness draws a learner’s attention to the target form so that he or she may form hypotheses about the target language.
CHAPTER 2: REVIEW OF THE RELEVANT LITERATURE

Feedback in SLA

Feedback is a construct borrowed from the literature of cognitive psychology on L1 acquisition (e.g., Marcus, 1993; Morgan, Bonamo, & Travis, 1995) and brought to SLA. While in cognitive psychology feedback has been understood as reactive and as a chance to provide positive reinforcement of a successfully completed task (positive feedback) or acknowledgement of an error (negative feedback), most of the SLA research has been done around the notion of negative feedback. Very often, negative feedback can be found in the literature as a synonym of negative evidence, error correction or corrective feedback, the latter being the most used term in SLA literature. SLA researchers have defined feedback in different ways, but there is a general consensus regarding its core description. For example, Lighbown and Spada (1999, p.171) defined feedback as “[a]ny indication to the learners that their use of the target language is incorrect”, or while Mackey and Abbuhl (2005) described it in more broadly as “the information learners receive regarding their communicative efforts” (p. 210).

Terminology

It is common for L2 researchers to use different terms interchangeably when referring to feedback, although lately the community tends to prefer certain terms over others. Thus, there are studies referring to feedback as evidence, particularly as positive evidence or negative evidence. Positive evidence can be described as exemplars of the L2 or “information that certain utterances are possible in the target language” (Leeman, 2007, p. 112), while negative evidence is considered information that certain utterances are impossible in the L2. Similarly, researchers have used the terms negative feedback and positive feedback, as acknowledgement of an error versus positive reinforcement of a successfully complete task, respectively. At first it may seem
that negative evidence and negative feedback overlap since they are both described as information regarding the ungrammaticality of certain utterances in the L2; however, as Leeman (2007) notes, negative feedback may contain positive evidence, negative evidence, or both. A term that has fallen into disuse is error correction, that implies a clear pedagogical intention to correct (Ortega, 2009) and it is usually avoided as it seems to suggest that the feedback provided leads to the elimination of errors, which, as Leeman (2007) points out, is a question to be answered empirically. Finally, a term that is widely seen in the SLA literature is corrective feedback, described as information regarding the ungrammaticality of learners’ utterances with the pedagogical intention to correct (Ortega, 2009, p. 71). Recently, it has been noted that since it is not always clear whether an interlocutor provides negative information with the intention to correct, negative feedback should be the preferable term (Ortega, 2009, p. 71).

While some researchers have argued the benefits or even the necessity of negative feedback for acquisition to take place (e.g., Krashen, 1981; Schwartz, 1993; Truscott, 2007), most of the SLA community seems to agree on the positive effects of feedback on the learning process. Indeed, the results of different recent meta-analyses (Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006, Mackey & Goo, 2007) point toward the idea that L2 learning can be promoted by the use of feedback.

**Modes and types**

Feedback can be provided by a human or by a computer in both oral and in written modes, and it can happen in the classroom or in the laboratory. While most research to date has involved feedback provided by humans either in the classroom or in the laboratory (e.g., Ammar & Spada, 2006; Carroll & Swain, 1993; Ellis, 2007; Herron & Tomasello, 1998; Leeman, 2003; Loewen & Nabei, 2007; Long et al., 1998; Lyster, 2004; Mackey & Oliver, 2000; Mackey &
Philp, 1998; McDonough, 2005, 2007; Sheen, 2008; inter alia), there is a growing line of research investigating the provision of feedback through the computer (e.g., CALL studies: Ayoun, 2001; Baralt, 2013; Loewen & Erlam, 2006; Sagarra, 2007, Sauro, 2009, CALI studies: ; Bowles, 2008; Camblor, 2006; Cerezo, 2010; Hsieh, 2007, 2008; Lin, 2009; Medina, 2008; Moreno, 2007; Morgan-Short & Bowden, 2006; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004, inter alia).

In contrast to the consensus regarding the concept of feedback, the controversy about feedback lies on determining its types and components. Several attempts to classify feedback have been done by different researchers (e.g. Leeman, 2000; Long & Robinson, 1998; Ranta & Lyster, 2007). Long and Robinson’s (1998) work described the types of data learners are exposed to in dichotomy terms (i.e., type of evidence, positive or negative; timing of the evidence, preemptively or reactively; and explicitness of the negative evidence, explicit or implicit), where feedback falls as reactive negative evidence that can differ in degrees of explicitness. Feedback has also been categorized as positive evidence, negative evidence, and metalinguistic rule (e.g., Leeman, 2000), while other researchers have considered that feedback should be classified solely on the bases of whether it pushes learners to repair their errors or not (e.g., Ranta & Lyster, 2007; Nassaji, 2007).

Much of the heated debate about feedback has centered around the notion of its degrees of explicitness. Carroll (2001) claims that the degree of explicitness can be determined by how it helps learners to perform the following actions: (a) detect the corrective purpose of feedback (i.e., there is an error); (b) spot the locus of the error (i.e., in what part of the utterance is the error located); and (c) recognize the nature of the error (i.e., error about vocabulary, morphosyntax, phonology, etc.). Thus, the label of ‘explicit feedback’ has been used to describe
feedback that overtly states that there is an error, while ‘implicit feedback’ requires the learner to infer the mere existence of the error (e.g., Carroll, 2001; Carroll & Swain, 1993; Leeman, 2000). An example of ‘explicit feedback’ is provided by Carroll and Swain (1993) by what they call *utterance rejection group*: “That’s wrong”, “You can’t say that”, or “That is not good English” (Carroll, 2001, p. 366). On the other side of the spectrum, recasts are shown as the typical example of implicit feedback, since the mere use of a recast does not guarantee that the learner perceives it (e.g., Lyster & Ranta, 1997). Based on these definitions of explicit and implicit feedback, the simple provision of negative evidence by means of gestures (e.g., a shake of the head) or verbal/written utterances (e.g., ‘no’, ‘wrong’, ‘incorrect’) classify as instances of explicit feedback.

Most of the research investigating degrees of explicitness of the feedback has been done in the context of face-to-face (FTF) interactions (e.g., Ammar & Spada, 2006; Carroll & Swain, 1993; Ellis, 2007; Ellis, Loewen, & Erlam, 2006; Leeman, 2003; Loewen & Nabei, 2007; Lyster & Mori, 2006; Lyster & Ranta, 1997; Sheen, 2007). Indeed, while all the four recent meta-analyses investigating the effectiveness of feedback did examine feedback delivered by a human during face-to-face interactions (Russell & Spada, 2006, Lyster & Saito, 2010), only two of them also included computerized feedback (Mackey & Goo, 2007; Li, 2010). The results of these meta-analyses regarding the effectiveness of feedback based on the degree of its explicitness are inconclusive. Russell and Spada (2006) did not draw any conclusions about the effectiveness of feedback types on the grounds that there were insufficient studies to arrive at any conclusion. Mackey and Goo (2007) did not make any claims, as they believed there is a “need for greater theoretical specificity or practical motivations in making claims about the superiority of one

---

1 Although one of their goals was to expand previous inclusion criteria used in other meta-analyses (e.g., Keck, Iberri-Shea, Tracy-Ventura, & Wa-Mbaleka, 2006) to include computer-based studies, this resulted in the inclusion of only two studies (Ayoun, 2001; Sagarra, 2007).
feedback over another” (p. 440). Lyster and Saito (2010) investigated recast, prompts and explicit correction and found larger effects for prompts over recast. Interestingly, though, following Lyster and Ranta (1997) they classified prompts as any type of negotiation of form (i.e., clarification requests, repetition of error, elicitation, and metalinguistic clues). Considering the explicit/implicit categorization more of a continuum than a dichotomy, Lyster and Saito (2010) placed recast towards the implicit end and explicit correction towards the explicit end. However, due to all the feedback moves that prompts included (e.g., clarification requests and metalinguistic clues), this type of feedback was placed along the entire continuum, thus making the results of the meta-analysis regarding the explicitness of the type of feedback less informative. In her meta-analysis, Li (2010) coded feedback types as implicit (i.e., clarification request, elicitation, repetitions, and any feedback not intended to draw learner’s attention to the error), and as explicit (i.e., metalinguistic feedback, explicit correction, and any feedback intended to draw learner’s attention to the error). The results indicated that there was a medium overall effect for corrective feedback, and that effect was maintained over time. However, the effect of implicit feedback was better maintained than that of explicit feedback. With regards to human-delivered feedback versus computerized feedback, this was the only meta-analysis that studied mode of delivery as a possible variable. While Mackey and Goo (2007) also included computer-delivered feedback, the number of studies that met the inclusion criteria and had computerized feedback was only two (i.e., Ayoun, 2001; Sagarra, 2007). Furthermore, this meta-analysis did not investigate mode of delivery as a possible mediating factor. Li (2010), on the other hand, did investigate mode of delivery and interlocutor as independent variables. For mode of delivery, she classified studies based on whether the feedback was delivered in face-to-face communication or in computerized communication. Moreover, she also classified the empirical
studies according to interlocutor type, and coded them based on whether the feedback was
provided by native speakers, a computer, or language teachers. While results indicated that
computer-delivered feedback and face-to-face feedback did not differ substantially in affecting
L2 development, she found that feedback provided by native speakers or embedded in the
computer were more effective than feedback provided by language teachers.

**Degrees of explicitness in computerized feedback.** Recently there have been a number
of computer-based studies that have empirically examined the role of different types of feedback
with regard to their degree of explicitness. Some of these studies have explored computer-
mediated communication environments (e.g., Loewen & Erlam, 2006; Sauro, 2009; Yilmaz,
2012). As is the case with FTF interaction studies, results regarding what feedback type is more
effective are not conclusive. While both Loewen and Erlam (2006) and Sauro (2009) found no
differences between feedback groups (metalinguistic prompts and recasts in Loewen and Erlam
(2006), and metalinguistic information and recast in Sauro (2009)), Yilmaz (2012) found a clear
advantage for explicit correction over recast.

A similar pattern is found in computer assisted language instruction (CALI) studies,
where the computer takes the role of a tutor. The overall results of these CALI studies
underscore a positive effect of feedback in language development (e.g. Nagata & Swisher, 1995;
Rosa & Leow, 2004; Sanz & Morgan-Short, 2004). However, different results have emerged as
to what type of feedback seems to be more beneficial. Some have found an advantage of more
explicit feedback over implicit (e.g. Bowles, 2008; Lado et al., 2013; Nagata, 1993; Nagata &
Swisher, 1995; Rosa & Leow, 2004). Others found no difference in the effects of explicit and
implicit feedback from pre- to posttests (e.g. Camblor, 2006; Hsieh, 2007; Moreno, 2007; Sanz
& Morgan-Short, 2004). And interestingly, one doctoral dissertation (Moreno, 2007), found an
advantage of implicit feedback over explicit in delayed measures. These studies are specifically 
relevant to the present dissertation, therefore they will be discussed more in depth in the review 
of the literature section.

Amount of feedback

So far I have examined the variable of feedback in general and feedback type in 
particular, that latter having earned a good amount of attention in SLA literature. However, a 
potentially contributing variable that has been overlooked in previous investigations is amount of 
feedback (Li, 2010; Loewen, 2012), being understood as the intensity of the feedback. In focus 
on form (FonF) instruction, researchers have considered length of treatment or intensity as a 
variable mediating its effects. In fact, Norris and Ortega (2000) identified length of instruction as 
a possible moderating variable. Intensity of FonF can be operationalized as “how frequently a 
specific linguistic form is focused on, with repeated focuses being more intensive and possibly 
more salient to learners” (Loewen, 2011, p. 582). Studies reporting the amount of time spent in 
the treatment have typically ranged between 30 to over 120 minutes. In her meta-analysis, Li 
(2010) found that treatments of 50 minutes or less had a significantly higher effect size than did 
longer treatments. However, as noted by both Li (2010) and Loewen (2011; 2012), the length of 
the instruction may not reflect the amount of feedback received. And although, according to 
Loewen (2011), recent studies in FonF have made an effort to report the length of treatment or 
the number of feedback episodes that occurred, the variable of amount of feedback has hardly 
been empirically studied.

There are two strands of studies examining amount of feedback, descriptive studies and 
experimental studies. In descriptive studies researchers have investigated how much feedback is 
generated based on certain aspects of the interaction, for example, type of task (e.g., problem-
solving, spot the difference, and free conversations in Crookes & Rulon, 1988), learners’ characteristics (e.g., age in Oliver, 2000 and Mackey, Oliver, & Leeman, 2003), contextual settings (e.g., dyads vs. teacher-fronted context in Oliver, 2000), interlocutor’s characteristics (e.g., native speakers vs. nonnative speakers in Mackey et al., 2003), and type of error (e.g., Havranek, 2002). In the second strand, amount of feedback has been studied as an independent variable. To date only one study has examined the effects of different amounts of feedback episodes on the same type of error (e.g., Ellis, Sheen, Murakami, & Takashima, 2008). The results of Ellis et al. (2008) indicated that there is no advantage for single or multiple corrections of the same linguistic error, and that simply the provision of feedback is beneficial for language learning. However, as this is only one study more research is warranted.

Summary of the feedback section

Feedback is a construct brought from cognitive psychology literature on L1 acquisition (e.g., Marcus, 1993; Morgan, Bonamo, & Travis, 1995) into SLA research. Empirical studies on the effects of feedback as well as recent meta-analyses have found evidence that feedback promotes second language development. One of the most debated points regarding feedback effectiveness is related to its degrees of explicitness. Several studies, mostly conducted in face-to-face interactions, have investigated whether explicit feedback is more effective than implicit feedback (recently researchers have rephrased the dichotomy as more explicit vs. less explicit feedback), and the results have been inconclusive. Similarly, computer-based studies examining the effects of types of feedback have reached different outcomes, with the scale tipping more towards explicit feedback. A potentially contributing variable that has been overlooked in previous investigations is amount of feedback, understood as the number of feedback episodes generated during interaction. To my knowledge, there are just a few descriptive studies (e.g.,
Crookes & Rulon, 1988; Oliver, 2000; Mackey et al., 2003), and only one experimental study (e.g., Ellis et al., 2008) that addresses amount of feedback. The results of this study seem to suggest that learning outcomes are not affected by different amounts of feedback. However, due to the paucity of studies investigating this variable, more research is warranted.

Considering all the heated debate surrounding feedback, it is interesting how little research has been carried out on the possible interaction it may have with other variables that could play a role in its effectiveness. Indeed, only recently some researchers have been interested in how the effects of feedback may be mediated by learner-internal variables, such as working memory (e.g. Mackey, Philp, Fujii, Egi, & Tatsumi, 2002) or learner-external variables, such as complexity of the linguistic form (e.g., Cerezo, 2010; de Graaff, 1997; Ellis, 2007; Nagata & Swisher, 1995). Of special interest for the present dissertation is this latter variable, thus a more comprehensive look at these studies will be presented in the review of the literature section.

Empirical Findings on Type of Feedback in CALI Studies

In this section I review CALI studies that have examined the role of type of feedback in L2 development. The review first looks at studies that included rules in the input (e.g., Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004; Stafford, Bowden, & Sanz, 2012), and then moves on to studies where the rules were not provided via a pre-practice explanation but rather via feedback (e.g.; Bowles, 2008; Camblor, 2006; Cerezo, 2010; Hsieh, 2007; Lado et al., 2013; Moreno, 2007). Tables summarizing these studies can be found in Appendices A and B.
**Rules in the input**

With a sample of 32 participants, Nagata and Swisher (1995) investigated how the effects of different types of CALI feedback (intelligent and traditional) could be mediated by two different Japanese structures (less complex, verb inflection in passive constructions; more complex, particles in passive constructions). Participants were divided into two groups depending on the type of feedback they received. Traditional CALI feedback was more implicit in nature as it informed about missing or unexpected words without providing explicit grammar explanation. On the other hand, intelligent CALI feedback informed about errors and gave information about appropriate rules. Results indicated learners that received intelligent feedback (explicit) performed significantly better than learners that received traditional feedback on the more complex structures. However, no difference between the two types of feedback was found for the less complex form. Due to this edge of the intelligent CALI feedback, the authors concluded “metalinguistic computer feedback can be very helpful when the linguistic domain to be learned entails significant complexity and may be much less helpful when it does not.” (p. 342). The results of this study must be interpreted with caution, however, since feedback was not isolated from grammatical explanation prior to practice so the possible contribution of explicit and implicit feedback is confounded with grammar explanation.

Investigating types of language instruction within the attentional strand, Rosa and Leow (2004) studied the effects of different task conditions on learners’ recognition and production of Spanish contrary-to-fact conditionals in the past. One hundred participants were divided in five

---

2 I follow the terminology regarding implicit and explicit feedback used in the original studies, while acknowledging the fact that such terminology can be problematic. As Cerezo (2010) argued, a detailed analysis of some feedback studies suggests that the implicit feedback definition used by some researchers does not seem to correspond to a true concept of implicit feedback. In the implicit feedback condition of these studies, learners were informed that they had made a mistake, and as I pointed out earlier implicit feedback requires the learner to infer the mere existence of the error (e.g., Carroll, 2001; Carroll & Swain, 1993; Leeman, 2000).
experimental groups and one control group. The conditions manipulated task-essentialness, immediate feedback provision (implicit or explicit), and pre-task information. Task-essentialness was held constant in all experimental groups, while prior to practice grammar explanation (+/−) and type of feedback (explicit/implicit) differed among the groups. The researchers operationalized explicit feedback as metalinguistic explanation provided upon learners’ responses, and implicit feedback as ‘right/wrong’ answers provided after learners’ performance. The instructional task was a computerized jigsaw puzzle that contained 18 critical items and 10 distractors. In each puzzle, participants had to complete the subordinate section of conditional sentences by choosing one out of four options that best conveyed the contrary to fact meaning in the past. The characteristics of the groups were the following: (1) EPEFE (prior grammar, practice, feedback [+ TE practice, +grammar, + explicit feedback]); (2) EPIFE (prior grammar, practice, feedback [+ TE practice, +grammar, + implicit feedback]); (3) EFE (no grammar, practice, feedback [+ TE practice, -grammar, + explicit feedback]); (4) EP (prior grammar, practice, no feedback [+ TE practice, +grammar, - feedback]); (5) IFE (no grammar, practice, feedback [+ TE practice, -grammar, + implicit feedback]); and (6) control: no grammar (- TE practice, no feedback [-TE practice, -grammar, -feedback]).

Assessment instruments consisted of a recognition test and a written production test and they were used at posttest (immediate after instructional session) and delayed posttest (3 weeks). The results showed that all the experimental groups that completed task-essential (TE) tasks (i.e., EPEFE, EPIFE, EFE, EPE, IFE) outperformed the control group in both recognition and production assessment tests, even after 3 weeks. In addition to this, learners in the explicit feedback condition performed significantly better than learners in the implicit feedback condition. In general, learners in the explicit learning conditions (with either explicit grammar
explanation and/or explicit feedback) performed better than learners in the implicit learning conditions (with either no explicit grammar and/or implicit feedback). Relevant for the present study are the results regarding the role of explicit/implicit feedback without metalinguistic information in the form of a pretask. While there was no difference between IFE and EFE groups in the recognition of old items and in the production of new items, differences were found in the production of old items and the recognition of new items where the group with explicit feedback (i.e., EFE) outperformed the group with implicit feedback (i.e., IFE). These results present a complicated landscape for the more/less explicit feedback debate.

The authors concluded that learners performed better when two sources of information on the target structure were provided (i.e., a pretask and explicit or implicit feedback), or when only one source of information in the form of explicit feedback was available.

Within the processing instruction strand and with the aim to investigate the effects of explicit conditions on the acquisition of Spanish word order, Sanz and Morgan-Short (2004) examined the effects of presence or absence of pre-task metalinguistic information, presence or absence of metalinguistic feedback, and task-essential practice. The 69 participants were randomly assigned to one of the following conditions: grammar explanation and explicit feedback [+E, +F], grammar explanation and no explicit feedback [+E, -F], no grammar explanation and explicit feedback [-E, +F], and no grammar explanation and no explicit feedback [-E, -F]. The instructional instruments were referential and affective structured input activities, and the assessment materials included interpretation and production tests. They found that all groups (regardless of the degree of explicitness) improved significantly and similarly. The authors interpreted that explicit information (prior to practice or during practice) was not necessarily conducive to language development. The lack of difference between types of
feedback groups is especially relevant for this dissertation since the present study investigates the effects of more versus less explicit feedback. Sanz and Morgan-Short (2004) concluded that task-essential practice was the sufficient element to promote language learning and that explicit information may not necessarily facilitate second language acquisition. Nevertheless, this conclusion needs to be taken with caution. Although the less explicit group (-E, -F) showed a significant improvement from pre- to posttests leading the researchers to their conclusion, it must to be noted that both [-F] groups did receive some kind of implicit feedback (as they were informed that their answers were ‘right’ or ‘wrong’) allowing them to make hypothesis about the underlying rule. Therefore, the isolated contributions of task-essentialness and feedback are not teased out.

In a recent study by Stafford, Bowden, and Sanz (2012), the researchers investigated the effects of explicit treatment conditions in the learning of Latin agent and patient roles with a sample of heritage and advanced learners of Spanish. Participants were divided into four different groups similar to those in Sanz and Morgan-Short (2004): [+GE, +EF] where participants received preemptive grammar explanation along with during practice metalinguistic feedback; [+GE, -EF] with preemptive grammar explanation and right/wrong feedback; [-GE, +EF] with no grammar explanation but metalinguistic feedback; and [-GE, -EF] with no grammar explanation and only right/wrong during practice feedback. With a pre-post-delayed test design, researchers operationalized learning, which they called initial language learning, by measuring the accuracy data from three different sources. First they administered two interpretation tests –one aural and one written– similar to those in the practice tasks of the treatment. Then, a grammaticality judgment test where participants decided if the sentences presented to them were well formed or not, and finally a production test to assess whether the
possible language skills developed during the input-based treatment could be transferred to produce Latin sentences.

To investigate if processing presented a different pattern according to the cues available for learners to assign thematic roles, the researchers assessed interpretation accuracy by sentence type. Thus, they examined interpretation accuracy for sentences where SVO cue was always available to determine the thematic roles; sentences where SVO was not available but the verb agreement cue was; and finally sentences where neither SVO nor verb agreement cues were available, but the noun case morphology cue was the only cue available.

Results suggested that TE practice coupled with less explicit feedback was sufficient for interpretation, but more explicit feedback was necessary for production. For this dissertation the results of the [-GE, +EF] and [-GE, -EF] groups, the two groups receiving no preemptive grammar explanation and more vs. less explicit feedback, are particularly relevant. Results for these two groups showed a wide pattern. Some results indicated no group differences (e.g., in the aural and written interpretation tests, and interpretation test with only agreement cues available), a better performance for the [-GE, +EF] group (e.g., in the written production test), and a better performance for the [-GE, -EF] group (e.g., in the grammaticality judgment test). In light of this, no clear conclusion regarding the effectiveness of one feedback type over the other can be drawn.

**Summary of rules in the input section.** All the above studies provided practice, pre-practice grammar explanation, and different types of feedback. In light of these studies, the picture of whether explicit conditions as opposed to implicit conditions are more beneficial to language development is not yet clear. While Nagata and Swisher (1995), and Rosa and Leow (2004) seem to find an edge for more explicit conditions, Sanz and Morgan-Short (2004), and
Stafford et al. (2012) argue that the pre-practice grammar explanation that usually accompanies the more explicit conditions in their experiments is not necessary for language development when TE practice and feedback are provided.

Regarding the role of type of feedback, the results of the reviewed studies show an unclear landscape. First, Nagata and Swisher’s (1995) study suggests that the effectiveness of one feedback type over the other lies in the grammatical form being taught. But this conclusion does not seem complete since both feedback groups had pre-practice grammar explanation. It is unclear how much can be attributed to feedback type and how much to grammar explanation. Rosa and Leow (2004) seemed to find an advantage for more explicit conditions. With regard to the role of feedback type without pre-practice grammar explanation, the effectiveness of one type of feedback over the other during TE practice changed depending on the assessment test (i.e., recognition vs. production) and whether the item was previously encountered (old) or new rendering any possible conclusion on feedback type relatively difficult to draw. It should be noted that in the design of this study the role of TE practice was combined with feedback. The lack of group difference in the Sanz and Morgan-Short (2004) study seemed to indicate that there was not a special advantage for one feedback type over the other. In fact, the authors concluded that TE practice, a common factor in all groups, was the sufficient element to promote language learning. However, as it has been pointed out, in this study TE practice was always coupled with feedback, therefore, it is unclear how much can be attributed to feedback and how much to TE practice. Stafford et al.’s (2012) results of the two feedback groups without pre-practice grammar explanation also showed an unclear picture. While in some tests one group performed better than the other, in other tests the pattern was reversed or no group difference was found. In this study,
like in previous studies, TE practice and feedback were always present so the partial role of each variable remains unclear.

In the next section I review CALI studies that included grammatical explanation only in the form of feedback. Some of them are doctoral dissertations while others are published studies.

**Rules provided only in the feedback**

In a CALI study, Bowles (2008) designed an output-based TE maze where participants had to translate a sentence from English to Spanish. The necessary elements (along with some unnecessary elements as well) were provided. Participants had to choose the correct items in the correct order to accurately translate a sentence containing *gustar*.

Since Bowles was also interested in type of verbalization, she collected data using verbal protocols. One hundred and fifty participants were randomly assigned to four different conditions with regard to type of instruction plus two more conditions as controls for verbal protocols. The six conditions were: A [+metalinguistic verbalization, +explicit feedback]; B [+metalinguistic verbalization, -explicit feedback]; C [-metalinguistic verbalization, +explicit feedback]; D [-metalinguistic verbalization, -explicit feedback]; E [no verbalization, +explicit feedback]; and F [no verbalization, -explicit feedback]. Because participants were all beginners of Spanish and data collection happened early in the semester, it was necessary for all of them to receive a computerized lesson on indirect object pronouns.

The [+explicit feedback] groups were provided with grammar explanation every time they chose an element, whether correct or incorrect, while the [-explicit feedback] group had no grammar explanation. The nature of the maze informed participants in the [-explicit feedback] groups that they had made an error, since only by choosing the correct words in the correct order could they proceed in the task. The results of written production tests revealed that participants
receiving explicit feedback performed significantly better on the immediate posttests than those who received implicit feedback. However, by the time of delayed posttests, these differences disappeared and all feedback groups performed statistically similar.

In a CALI study and under a cognitive, attentional framework, Camblor (2006) examined whether exposure to L2 input via different types of feedback has a differential effect on learners’ production of the Spanish noun-adjective agreement. The treatment task was an output-based, picture-description task that included masculine, feminine, singular, and plural nouns and adjectives. For her dissertation, Camblor created 24 pictures and for each picture participants had to write a sentence describing the object of the picture, its color, and its location in the living room.

Seventy-seven first-year college-level students were randomly assigned to four groups with the following conditions: A) Implicit feedback, where the errors were capitalized, and participants were prompted to rewrite their sentences; B) Explicit feedback in the form of metalinguistic explanation, where participants had their mistakes capitalized and, after receiving explanation about the source of the mistakes, were prompted to rewrite their sentences; C) Interactive feedback in the form of an interactive game that guided participants to the correct form of the grammatical item; after completing the game, participants were also required to rewrite their sentences; and D) Control (no feedback).

Assessment tasks were a controlled fill-in-the-blank task and a semi-open production test where participants were told that a couple had made changes in their living room and were provided with a picture of both before and after the redecoration. Participants’ task was to describe the changes, naming the objects, their color and their location in the room. Results
suggested that feedback provision was beneficial, but no conclusions could be drawn as for the beneficial effects of one type of feedback over another.

To address the conflation of task-essential practice and feedback present in previous studies (e.g., Rosa & Leow, 2004; Sanz & Morgan-Short, 2004), the doctoral dissertation by Hsieh (2007) included a control group for feedback in a study comparing practice versus exposure-to-practice, and feedback (explicit and implicit) versus no feedback. Fifty-two college-level students enrolled in first semester Spanish courses at an American university were randomly assigned to six different groups with the following conditions: PRONLY [practice only, no feedback]; PRIFB [practice, implicit feedback]; PREFB [practice, explicit feedback]; EXPONLY [exposure to practice, no feedback]; EXPIFB [exposure to practice, implicit feedback]; EXPEFB [exposure to practice, explicit feedback]. Implicit feedback was embedded in the task, since learners could not continue performing the task if they chose the wrong items to form an accurate sentence with gustar. The explicit feedback groups received metalinguistic explanations when choosing their answers. Performance was measured by oral and written production tests and recognition tests.

With regard to practice, her results indicated that learners in the practice groups performed significantly better than those in the exposure-to-practice groups as measured by oral production tests. However, no group differences were found in written production and recognition tests. As for the role of feedback, the feedback groups performed significantly better than the control group, but no differences between the explicit feedback and the implicit feedback groups were found. Hsieh concluded that practice only did not have a significant role, and that it was feedback, and not practice that facilitated language development.
The doctoral dissertation by Moreno (2007) also investigated the effects of practice and feedback. This dissertation directly addressed task-essentialness as an independent variable, as does the present study. A detailed description of its design is provided in the next section, but a brief summary with regard to Moreno’s results on feedback is offered here. The author found no differences between feedback groups on immediate posttests, but significantly greater gains of less explicit feedback in oral delayed measures. Moreover, the more explicit feedback group experienced a significant greater loss from immediate posttests to delayed posttests, while the less explicit group maintained their gains.

The next study is particularly relevant for this dissertation. It examined some of the independent variables that the present study investigates and is therefore also referred to in future sections of this review of the literature. In his doctoral dissertation, Cerezo (2010) studied the role of agency (active participation in practice versus exposure to practice), and how its effects could be moderated by type of feedback and complexity of the linguistic form during TE tasks. Feedback differed with respect to the degree of explicitness (error signaling vs. grammar explanation [spot vs. explain]) and opportunity for error repair (+ or – [prompt vs. continue]). The less complex form chosen was Spanish prepositional phrases in relative clauses, while the more complex form was Spanish present subjunctive in relative clauses. There were eight different groups, including two control groups, that resulted from the combination of these conditions. Participants received 15 mini-episodes with 15 items (10 target forms and 5 distractors) per structure. The instructional tasks consisted of a fill-in-the-blank activity within two communicative situations, one per targeted form. Assessment tools were a grammaticality judgment test (GJT), a written production test, and an oral production test. In terms of feedback, when both grammatical forms were considered together, results indicated that all feedback
groups showed some learning, in contrast to the two No Feedback groups (practice and exposure to practice). This seems to suggest that in the absence of metalinguistic information, be it prior to or during practice, no learning occurs. Moreover, considering the results of immediate and delayed posttests, differences between the more explicit feedback and less explicit feedback groups were found, with more explicit feedback yielding better results.

Finally, Lado, Bowden, Stafford, and Sanz (2013), within the Latin Project framework (Sanz, Bowden, & Stafford), investigated Latin assignment of thematic functions in a CALI study with a group of 58 college students enrolled in a second year Spanish program. The researchers provided TE computerized practice and two types of feedback, namely more explicit feedback that included negative evidence and metalinguistic information (NE+MI), and less explicit feedback without metalinguistic information (NE-MI). Accuracy and reaction times (RTs) were measured as well as performance on old and new items. Results showed that participants in both experimental conditions improved in accuracy and RTs on immediate posttests, with the NE+MI group showing an advantage over the NE-MI group. However, by the time of delayed posttests, the differences between groups dissipated. In their conclusion, the researchers determined that both types of feedback lead to better accuracy and faster responses. In terms of the effectiveness of one feedback type over the other, they underscored that while providing metalinguistic feedback led to an initial advantage, this edge disappeared after two weeks. In contrast, they concluded that providing learners with negative evidence in the form of right/wrong feedback in combination with TE practice leads to more sustained yet modest gains.

**Summary of rules provided only in the feedback section.** All experimental studies of the above review included metalinguistic information about how the target language works only in the form of feedback. The results of these studies do show a constant: that providing
participants with feedback, regardless of the type, proves to be beneficial as learners perform more accurately and faster (e.g., Lado et al., 2013). As in the case of the studies that included rules via pre-practice grammar explanation, the results of these feedback only studies do not show a clear tip of the scale in favor of one feedback type over another. At first, it appears that more explicit type of feedback yields better results, but the picture is not as clear. In some of the studies where more explicit feedback showed an advantage in immediate posttests, that advantage disappeared with time (e.g., Bowles, 2008; Lado et al., 2013). Other studies did not find any differences at immediate or delayed posttests (e.g., Camblor, 2006; Hsieh, 2007). One additional study found an edge for more explicit feedback at both immediate and delayed posttest (Cerezo, 2010), while another found no differences at immediate posttests, but an edge of the less explicit feedback group at delayed posttests (Moreno, 2007).

A common limitation of some of the reviewed studies is the lack of control group in their designs (e.g., Bowles, 2008; Lado et al., 2013; Moreno, 2007). Camblor (2006) had a control group for feedback; however, as noted by Cerezo (2010), the nature of her experimental tasks (i.e., output-based) as well as the design of the more explicit feedback condition (i.e., no models provided) left participants with no possible input of the language to shape their hypotheses, thus making learning very challenging. It should also be noted that in several of these studies the design of the experimental conditions had feedback and practice conflated (Bowles, 2008; Lado et al. 2013; Moreno, 2007). Two studies that avoided the conflation of practice and feedback by including a control group are Hsieh’s (2007) and Cerezo’s (2010). Hsieh (2007) compared the control group (PRONLY) with the two feedback groups (PRIFB and PREFB). Cerezo (2010) also compared the control group (No Feedback) with the all feedback groups (EC, EP and SC, SP). While the results of Hsieh (2007) indicated that no feedback group performed better than
the other, Cerezo (2010) found an advantage for more explicit types of feedback. Given that
these are the only two studies that were able to properly examine the role of type of feedback,
and given their different results, more research is warranted.

**Empirical Findings on Amount of Feedback Studies**

Most studies that have investigated amount of feedback are interaction descriptive studies
where the amount of feedback provided was not manipulated. These investigations examined
how L2 FTF interactions generated different amounts of feedback depending on type of task
(e.g., problem-solving, spot the difference, and free conversations in Crookes & Rulon, 1988),
learners’ characteristics (e.g., age in Oliver, 2000), contextual settings (e.g., dyads vs. teacher-
fronted context in Oliver, 2000), interlocutor’s characteristics (e.g., native speakers vs. nonnative
speakers in Mackey, Oliver, & Leeman, 2003), type of feedback (e.g., Loewen & Nabei, 2007),
and type of error (e.g., Havranek, 2002). What it follows is a review of the most relevant studies
that are also summarized in Appendix C.

Whereas the focus of the present study is not to examine the role of type of error
correction (i.e., grammatical error correction, vocabulary error correction or pronunciation error
correction), nor the role of elicitations, the design of this investigation does include elicitations
and grammatical error correction thus, an in-depth review of Havranek’s (2002) study is
warranted. In her classroom-based exploratory study Havranek investigated six different levels
of proficiency with 207 German learners of English. One thousand and seven hundred instances
of corrective feedback were analyzed according to type, length, and utterance focus. Different
amounts of feedback emerged depending on whether the correction was directed to grammar,
vocabulary, or pronunciation. Corrections were also analyzed to determine whether they allowed
for elicitations or not. The length of these corrections was measured in turns. The results
regarding the effects of feedback with and without elicitation indicated that corrections without elicitation were least successful while corrections with elicitation that lasted between 5-7 turns were most successful. However, corrections with elicitation longer than 7 turns were less successful, which she interpreted as an indication that the learner was not developmentally ready for that structure. The type of error that was most frequently corrected was grammatical errors whose correction was also found to be successful when looking at the assessment tests. Havranek (2002) concluded that learners profited more from feedback when errors involved “relatively simple rules” (p. 269) that she described as rules involving “redundant features like verb endings or use of auxiliary do in questions and negation” (p. 268). The classification of these two rules as simple can be controversial. First, the author did not explain how complexity is defined or whether she follows a particular author’s description of linguistic complexity. Second, while some researchers agree with the classification of negation as a simple rule (e.g., Housen et al., 2005), some other researchers disagree with the classification of question formation as a simple rule. Thus, authors such as Spada and Tomita (2010) when carrying out their meta-analysis and following Hulstijn and de Graaff’s (1994) description of linguistic complexity classified question formation as complex. Similarly verb endings could involve the use of the 3rd person singular –s, and as mentioned earlier, this form has been categorized as simple (for example, by Krashen, 1982), but also as complex (for example, by DeKeyser, 1998). Considering all this, Havranek’s conclusion that learners benefited more from feedback when the feedback involved simple rules seems inadequate. With regard to the effect of repeated corrections, the author claimed that further correction on the same item did not contribute much to their learning with hardly any difference in the accuracy scores of items that were corrected once, twice or even three times, and a decrease in the accuracy of errors that were corrected on more than three occasions. It
needs to be noted that there was no correlational analysis between the amount of feedback each individual received and their particular performance in the assessment tests that could more clearly determine the relationship between amount of feedback and performance.

Loewen and Nabei (2007) investigated the efficacy of different types of feedback (i.e., recast, clarification requests, and metalinguistic) on English question formation. They also looked at the different amount of feedback generated by each feedback type. In this review I primarily focus on information relevant to amount of feedback. Their 66 participants were assigned to four different groups: recast, clarification, metalinguistic, and no feedback. They all completed two tasks: a spot-the-difference task and a guess-the-storyline task. The assessment tests were an untimed GJT, a timed GJT, and an oral production task. The oral data from the two tasks were recorded and each feedback episode was tallied and classified. The results showed that recast and clarification requests were the most frequent type of feedback with an average frequency of 18 episodes each (recast: 18.7; clarification: 18), while metalinguistic feedback was much less frequent (5.6 episodes). Note that these numbers are merely descriptive, as no statistical analyses to compare the amount of feedback per group were run. As for participants’ performance in the assessment tests, all feedback groups performed similarly regardless of the amount of feedback received. The authors concluded that "the fact that the metalinguistic group performed as well on the posttests as the other two groups suggests that less metalinguistic feedback may be necessary to achieve the same results as compared to recasts or elicitations" (p. 374). However, this conclusion assumes two things: 1) that recasts and the clarification groups significantly had more feedback episodes than the metalinguistic group, which we do not know since no statistical analyses were run; and 2) that recasts episodes and metalinguistic episodes can be compared in terms of amount, which I am going to argue cannot be done. To do a just
comparison on the amount of feedback that each group could effectively take in to check and form their hypotheses of the language, it would be necessary to count the feedback episodes that were noticed by the participants. It is widely known that the implicit nature of the recasts can make them go unnoticed, so it is not possible to know how many of those 18 recasts were noticed and how many were not. The explicit nature of the metalinguistic feedback makes them very noticeable as it implies that a teacher or researcher explicitly states there is an error and what the grammatical rule is, but we cannot be sure that the participants noticed all of the 18 episodes. It could be the case that only five episodes were noticed in the recast group. Thus, the conclusion of the researchers based on different amounts of feedback should be taken with caution.

Another interaction study, Mackey, Oliver and Leeman (2003) investigated the effects of interlocutor type on the provision and incorporation of feedback in task-based interaction. They collected data from adult versus child dyads, and native speakers (NSs) versus non-native speakers (NNSs) dyads. Their interactions were analyzed to assess the effects of interlocutor on three factors: (1) amount of feedback, (2) opportunities for modified input, and (3) immediate incorporation of feedback. They specifically focused on nontarget utterances that contained grammatical errors. The two research questions relevant to amount of feedback were: (1) Do NSs and NNSs differ in the amount of implicit negative feedback they provide to NNSs during task-based interaction? and (2) Do adult and child dyads differ in the amount of implicit negative feedback provided to NNSs during task-based interaction? The results of the first research question showed that when comparing NSs and NNSs regardless of age, NSs in both age groups provided more feedback than NNSs, but it was statistically significant only for adult NSs. In other words, adult NSs provided significantly more feedback than adult NNSs. With respect to
the second research question, it was found that adult and child dyads, regardless of nativeness level, provided the same amount of feedback.

This was one of the first studies to identify the potential contributions of amount of feedback and to explore this as a variable. They underlined the importance of researching several aspects of feedback such as amount, its possible correlation with L2 development, and the possible need of a minimum amount of feedback to promote L2 development. As they put it:

“(…) further research is clearly warranted to determine the effects of feedback, and in particular, to investigate whether amount of feedback directly correlates with linguistic development. In other words, if we establish that feedback can promote development, we then need to study whether more feedback leads to more development, whether there is a minimal level that is sufficient, and whether timing, quality, type, output opportunities, and context of feedback are relevant factors, as we suspect they are”.

(Mackey, Oliver, and Leeman, 2003, pp. 56-57; Emphasis from the original text)

To my knowledge, only one study, Ellis, Sheen, Murakami, & Takashima (2008) has investigated amount of feedback as an independent variable. In this quasi-experimental study, 35 Japanese learners of English were part of three intact classes serving as two experimental groups and a control group. The experimental groups received focused (N= 11) and unfocused written feedback (N= 13), while the control (N= 11) did not receive any feedback. The focused group was provided with corrective feedback only on errors related to definite and indefinite articles, while the unfocused group received feedback on articles as well as other types of errors. Assessment tests were an error correction test and a free writing task. The error correction test contained twelve article items and four distractors, so the maximum possible score was twelve. For the writing task, participants were provided with picture compositions containing four
pictures. Students’ task was to provide a title for the story and a detailed story. Writing test scores were calculated using obligatory occasions analysis (Ellis & Barkuizen, 2005). Each obligatory occasion for the use of the articles were identified and then scored. As the researchers expected by the study design, the focused group received more frequent correction on article errors than the unfocused group. The results showed that both feedback groups improved significantly from pretests to posttest on both assessment tests. However, the researchers did not find any differences between the feedback groups. At delayed posttests, the unfocused group maintained their level of accuracy, while the focused group increased their level of accuracy on article use, but yet again, these differences between groups were not significant. Nevertheless, the authors explained these differences referring to the fact that the focused group received more feedback on article errors than the unfocused group. They concluded that there was no advantage for multiple corrections or fewer corrections of the same linguistic error, and that it was just the provision of written corrective feedback that was effective at least for English articles. This study bore some limitations. As the authors acknowledged, the final sample size was small due to the fact that some participants showed a good level of accuracy at pretests and were therefore excluded from the study. Also at pretests, the standard deviations of the focused group’s scores were about half those of the unfocused and control groups, suggesting than the three groups might not have been comparable even before the experiment took place.

Summary of empirical findings on amount of feedback studies section

Most studies that investigated amount of feedback are descriptive studies where the amount of feedback provided was not manipulated. These investigations examined how different learner-external factors (e.g., type of task, or contextual settings) and learner-internal factors (e.g., learners’ characteristics) generated different amounts of feedback. One such study is
Havranek (2002) who investigated 207 German learners of English. Her results indicated that learners benefited more from feedback when errors involved “relatively simple rules” (p. 269) compared to more complex rules. However, the author did not follow a clear description of what is simple and what is complex. The results of repeated corrections indicated that more feedback did not relate to more learning, that is, that further correction on the same item did not contribute much to their learning with hardly any difference in the accuracy items corrected up to three times and a decrease of the accuracy of items corrected more than three times. In another ESL study investigating the efficacy of different types of feedback (i.e., recast, clarification requests, and metalinguistic), Loewen & Nabei (2007) found that recast was more frequent than metalinguistic feedback, but both groups, the recast and the metalinguistic feedback groups, performed statistically similarly. In light of these results, the authors concluded that less metalinguistic feedback might be necessary to achieve the same results as the recast and the clarification groups obtained. However, this conclusion assumes that the amount of feedback of the metalinguistic group was significantly lower than that of the other two groups, something that was not statistically tested. It also assumes that the instances of recasts and metalinguistic feedback they tallied can be directly compared, while it is possible that some instances of the recast were not noticed, and hence the amount of recast and the amount of metalinguistic feedback should not be compared. An interesting study that was one of the first to call attention to the possible effects of amount of feedback was Mackey, Oliver and Leeman (2003) who investigated the effects of interlocutor type on the provision and incorporation of feedback in task-based interaction. While the researchers were not directly interested in investigating amount of feedback, they found that adult NSs provided more feedback than adult NNS. However, since they did not investigate the developmental effects of feedback, we cannot compare their results
with those of Havranek or Loewen & Nabei. Finally, to my knowledge, as of the writing of this
dissertation, Ellis et al. (2008) is the only study that manipulated the amount of feedback that
learners received. In this study two groups received focused written feedback only on article
erors and unfocused written feedback on several types of errors, while a control group did not
receive any feedback. Results indicated that while the focused group received more feedback
than the unfocused group, both feedback groups benefited from feedback but no group
differences arose. Limitations of this study included a small sample size and lack of homogeneity
between groups before treatment. Altogether the results of Havranek (2002), Loewen and Nabei
(2007), and Ellis et al. (2008) seem to suggest that more feedback does not translate into better
performance. However, none of these studies carried out correlational analyses nor did they set
out to determine the role of amount of feedback, so more research on the possible role of
different amount of feedback is warranted.

**Practice**

That input exposure is required for second language acquisition seems to be an accepted
consensus in the field of SLA (e.g., Leow, 2007; Sanz & Morgan-Short, 2005; Stafford, Bowden
& Sanz, 2012). However, variables embedded within input exposure such as degrees of
manipulation of the input, (type of) practice, presence or absence of metalinguistic information,
or timing of this information (prior-, ongoing- or post-task) are still debatable. Several studies
have investigated some of the above-mentioned variables (such as manipulated input, rule
presentation, feedback), and their interaction (see Norris & Ortega, 2000; Sanz & Morgan-Short,
2005; and Spada & Tomita, 2010 for detailed reviews). However, as Sanz and Morgan-Short
(2005) noted, the practice that contextualizes input, feedback and rule presentation has not been
the focus of SLA research. This has been considered a limitation of the field. Indeed, as Sanz and
Morgan-Short (2005) put it: “Indeed, one can choose to include rule presentation or feedback in a treatment, but both input and practice—task or activity—in more pedagogical terms—must be present.” (p. 237). An issue that immediately arises from this statement, though, is how to define practice and what type of practice is the most beneficial for second language learning. Rutherford and Sherwood Smith (1988) advocate for ‘consciousness raising’ activities that deliberately “attempt to draw learner’s attention specifically to the formal properties of the target language” (p. 107), while Doughty and Williams (1998, p. 3) note that a previous engagement in the meaning needs to be established prior to attention to the form.

Despite the wide use of the word ‘practice’ in the field of SLA, only a relatively small number of researchers have attempted to provide a definition of this concept. One of the first researchers to provide a definition of practice was Ellis in his 1993 article. According to Ellis, practice "involves an attempt to supply the learner with plentiful opportunities for producing targeted structures in controlled and free language use in order to develop fully proceduralized implicit knowledge." (p. 109). Other researchers, such as Lightbown (2000), define practice as “opportunities for meaningful language use (both receptive and productive) and for thoughtful, effortful practice of difficult linguistic features” (p. 443). Sanz and Morgan-Short (2005) emphasize the importance of interacting with the target form: “practice requires learners to interact in some way with the target form in the input, to respond to the information” (p. 237). As part of a series of articles edited in a book by DeKeyser (2007) with the purpose of examining the concept of practice in second language, Leow (2007) described practice—receptive practice in particular—from an attentional perspective:

“any exposure to manipulated L2 input that provides not only many exemplars of targeted L2 forms or structures upon which learners’ attention to (and/or awareness of) is
directly or indirectly premised but also some form of opportunity to perform a limited productive or nonproductive task or activity (e.g., selecting one out of two options, completing a problem-solving task, translating) during the exposure.” (pp. 21-22)

In his book DeKeyser (2007) defined practice “as specific activities in the second language engaged systematically, deliberately, with the goal of developing knowledge of and skills in the second language.” (p. 8). In her contribution to this book, and based on Lightbown’s (2000) notion of practice, Lourdes Ortega (2007) described meaningful practice as one that enables peer collaboration and is purposefully selective but relative unobtrusively focused on the language code within a wider context of communication. She related her notion of practice to the proposal of task-based language learning developed by Robinson (1995, 2003).

When defining practice in SLA, two distinctive types can be distinguished based on its focus: meaning-focused practice and form-focused practice. Meaning-focused practice is in nature more incidental, with no pre-established language-learning goals. Form-focused practice – used predominantly in form-focused instruction approaches – involves deliberate practice with a preset language learning goal. It also involves the presence of a tutor – human or computer-mediated – that, according to Bransford, Brown, and Cocking (1999), helps learners “to rehearse appropriate practices that enhance performance” (p. 178).

From a form-focused instruction approach, over the past several decades SLA scholars have been investigating what is the most beneficial practice for language learning. Rutherford and Sherwood Smith (1988) advocated for ‘consciousness raising’ activities that through hypothesis testing and inference facilitate language development. In their 1993 paper, Loschky and Bley-Vroman suggested that meaningful communication tasks can successfully be used as form-focused practice and are the best way to incorporate grammar in language instruction.
Loschky and Bley-Vroman (1993) introduced the notion of task-essentialness, and since then several studies have incorporated task-essential practice into their designs (e.g., Leow, 1997a, 1997b, 1998a; 1998b; Rosa & O’Neill, 1999). Task-essentialness refers to the idea that structural accuracy is essential to the meaning in the task (Loschky & Bley-Vroman, 1993, p. 131). In order to successfully complete a communicative task, “it is essential to attend to the relevant structure” (p.138). They also argue for the inclusion of feedback, which is considered necessary to help learners to hypothesize about the underlying rules or to confirm or disconfirm their previous hypotheses.

Task-essentialness has also been one of the features considered in the meta-analysis performed by Keck, Iberri-Shea, Tracy-Ventura, and Wa-Mbaleka (2006). Due to the proliferation of studies on task-based language teaching, Keck et al. (2006) carried out a meta-analysis investigating the link between task-based interaction and acquisition. Although there are numerous primary investigations that include communication tasks as treatment in their research designs, Keck et al. (2006) had to narrow the scope of their investigation and only reviewed face-to-face interaction tasks studies. They coded primary research studies based on task types (jigsaw, information gap, problem-solving, decision-making, opinion-exchange, and narrative), and task features (task-utility, task-essential, task-natural). Interestingly, they found that studies that included task-essential tasks yielded larger effects over time.

Despite these apparent benefits, task-essentialness has hardly been studied as an independent variable. It can be found as part of research designs (e.g. Leow, 1997a, 1997b, 1998a, 1998b; Rosa & O’Neill, 1999), and recently, with the proliferation of computer-based studies, TE has also been included in some CALI studies (e.g., Cerezo, 2010; Hsieh, 2007; Lado et al., 2013; Moreno, 2007; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004, Stafford, et al.,
Among these studies, only one doctoral dissertation, Moreno (2007), included TE as an independent variable with the purpose to examine the contributive role of TE in instruction. Due to some methodological issues in the design, such as the conflation of feedback and practice, the isolated role of TE could not be determined. Two more dissertations, Hsieh (2007) and Cerezo (2010), included task-essentialness in their designs with the purpose to examine the role of agency, that is, the role of actively completing TE practice as opposed to passively being exposed to TE practice. The results of both studies seem to indicate that it is the combination of active TE practice and feedback as opposed to TE practice only that promotes language learning. These two studies showed an advantage for active TE practice over exposure to it, but the isolated role of TE practice compared to any other kind of practice has not yet been studied. The present dissertation seeks to fill this gap.

**Summary of practice section**

Traditionally in SLA literature, a clear definition of practice has been avoided, although recently there have been some attempts to address this. Based on its focus, practice can be considered incidental or deliberate. By nature, form-focused instruction approaches to language acquisition utilize deliberate practice, as this involves a pre-determined learning goal. Among some of the features of form-focused practice, task-essentialness has been defined as a desired feature when designing practice, since it orients learners’ attention to the linguistic form, and it may help them, if further processed, to connect form to meaning. TE has been present in several computer-based and non-computer-based studies (e.g., Cerezo, 2010; Hsieh, 2007, 2008; Lado et al., 2013; Leow, 1997a, 1997b, 1998a, 1998b; Rosa & O’Neill, 1999; Moreno, 2007; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004, Stafford, et al., 2012), as well as in a meta-analysis (e.g., Keck et al., 2006). The results of both types of studies suggest that TE may help language
development but given the conflation of TE and feedback the contributive role of TE is still unclear. Although there have been three different unpublished doctoral dissertations that have studied the role of TE, their results are inconclusive, warranting further investigation into the unique contributive role of TE. The present study aims to address this issue by investigating the isolated role of TE in a CALI study.

**Empirical Findings on Task-essential Practice Studies**

As seen in the previous section, several investigations have included TE practice as part of their instructional materials under the premise that TE practice makes learners pay attention to the target form and interact with it in a meaningful way. Since Loschky and Bley-Vroman (1993) introduced the notion of task-essentialness, several studies have incorporated TE practice into their designs (e.g. Leow, 1997a, 1997b, 1998a, 1998b; Rosa & O’Neill, 1999). What follows is a review of said studies. First, I analyze paper-and-pencil format studies and then CALI studies. Within the attentional strand, Leow (1997a; 1997b; 1998a; 1998b) published a series of studies whose research designs included TE practice. In these studies the researcher manipulated a crossword puzzle in such a way that the learners needed to attend to the grammatical form in order to complete the task. The overall results of these studies showed that learners engaged in TE practice attended more to the targeted item (e.g., Leow, 1998a, 1998b) and reached higher levels of awareness that are considered to lead to learning (Schmidt, 1990). Attending to the relevant structure to successfully complete the task is one of the desired features that a communicative task should have (Loschky & Bley-Vroman, 1993). However, it can be argued that the crossword puzzle task common in these studies was less meaningful, as it did not necessarily push the learners to make form-meaning connections.
This limitation was overcome by Rosa and O’Neill’s (1999) study with a jigsaw puzzle task that pushed learners to make form-meaning connections. With a pool of 67 learners enrolled in fourth semester Spanish at an American university, Rosa and O’Neill (1999) examined the development of Spanish contrary-to-fact conditional sentences in the past. Participants were randomly assigned to one of five conditions. The conditions differed based on two aspects: (1) whether they were provided with formal instruction about the targeted structure or with a reading passage with no conditional sentences (±FI); and (2) whether they had to search for the rule or they were asked to memorize some content while completing the puzzle task (±RS). The fifth condition served as a control, [T-only]. This group was provided with the same reading passage that (−RS) groups received and no formal instruction on the targeted structure. They had to complete the jigsaw puzzle for whose completion no directions were given and also complete a vocabulary test.

Results from the multiple choice recognition test indicated that participants in the (+FI) groups improved significantly more than did the group in the most implicit condition, namely [−FI, −RS]. However, no differences were found between the performance of the [+FI, +RS], [+FI, −RS], [−FI, +RS], and [T-only] groups in the multiple-choice recognition task. The [T only] group performed almost significantly better than the group that received no formal instruction and was asked to memorize content during the task, namely [−FI, −RS]. This was not expected by Rosa and O’Neill (1999) who explained this result arguing that the [−FI, −RS] group could have been exposed to a cognitive overload having to figure out the rule underlying the grammatical form, complete the task, and memorize content, whereas the control group was more free to direct its attention to the task and the targeted form. Their experimental task used with all groups—the jigsaw puzzle—had two features, task-essentialness and feedback. The
puzzle was manipulated so that the learners had to attend to the form in order to successfully complete the task. Feedback was inherent to the task since the puzzle pieces did not fit if the wrong choice was selected. In fact, the researchers did point to the combination of these features as the possible cause of the positive results experienced by their learners. As it was not possible to create a condition with only the TE feature in the task designed by the experimenters, the isolated contribution of task-essentialness and feedback is not clear.

After the work by Rosa and Leow (2004) and Sanz and Morgan-Short (2004) reported in the previous section, several doctoral studies included TE practice in their CALI designs (e.g., Bowles, 2008; Cerezo, 2010; Hsieh, 2007; Lado et al., 2013; Moreno, 2007), all of them with positive results.

Whereas most of these dissertations did not have the goal to investigate the contributive role of TE practice, one study (Moreno, 2007) directly addressed ‘task-essentialness’ as an independent variable. She sought to investigate the effects of task-essentialness and feedback on preverbal direct object pronouns (OVS Spanish word order). Her fifty-seven participants were randomly assigned to one of four groups that combined presence or absence of task-essentialness [+TE/-TE] and more versus less explicit feedback [+EF/-EF]. In the non task-essential condition (-TE), after seeing a picture, participants had to complete an interpretation task by selecting one of two sentences. The two sentences contained the same structure form (direct object pronoun); therefore, attention to the target structure was not necessary since the only difference between both sentences was on the subjects and not on the objects. In the task-essential condition (+TE) students were required to carefully pay attention to the direct object pronoun to successfully complete the task. According to the pictures, participants had to choose either masculine/feminine and singular/plural direct object pronouns when navigating a tree that
successively branched out with a set of object pronouns. Moreno (2007) did not find significant effects for task-essentialness, leading her to conclude that task-essential practice (vs. non task-essential practice) did not play a significant role in language learning. However, this study presents some limitations. First, the isolated role of TE was not clear since all groups received feedback and no control group for feedback was included. Second, her sample size was small with a consequent low power. Due to these limitations, Moreno (2007) could not draw a definitive conclusion regarding the role of task-essentialness.

TE practice was a task feature present in two more doctoral studies (Hsieh, 2007; and Cerezo, 2010). Both of these studies looked at the role of agency when completing TE practice, and both had larger sample sizes than Moreno’s (2007) study and a control group for feedback. Hsieh designed her study with input-based TE practice that three groups completed while receiving more explicit feedback (PREFB), less explicit feedback (PRIFB), and no feedback (PRONLY). In Cerezo’s (2010) study, participants completed output-based TE practice. His design also included a group that did not receive any feedback (No Feedback). However, these studies were not interested in examining the role of TE, and no group with non-TE practice was included. Consequently the role of TE practice remained unclear.

After these doctoral dissertations, other published studies have also included TE practice into their designs. Some of these have already been reviewed in previous sections (e.g., Stafford, Sanz, & Bowden, 2010; Stafford et al., 2012; Henshaw, 2012; Lado et al., 2013; Sanz, Lin, Lado, Stafford & Bowden, 2014). Henshaw’s (2012) study is particularly relevant for the current dissertation given its design and its targeted form.

Following the tenets of Processing Instruction (VanPatten, 2002), the purpose of Henshaw’s (2012) study was to evaluate the effects of referential and affective activities—both
in isolation and in combination with one another—on the learning of adverbial clauses of time.

One hundred and three participants were randomly assigned to one of the following three groups: AFF, where learners received affective activities only; REF, where learners received referential activities only; and RA, where learners received referential activities followed by affective activities. By definition, referential activities are considered TE since 1) attention to the relevant structure and 2) structural accuracy are necessary in order to successfully complete the task. However, affective activities are not TE since they can be successfully completed overlooking the relevant structure.

The treatment consisted of explicit information on how to form the present subjunctive, where in the sentence it can be found, and when it is used. As PI proposes, participants were also informed about a possible strategy that could negatively affect the way they make their choices. They then completed a recognition task and an interpretation task with 48 items. In the recognition task all learners heard the first half of a sentence (main clause and conjunction), and had to select from the options on their screens the phrase that best completed it. These options varied from one group to another: the REF group had to choose between a phrase in indicative or subjunctive (only one was correct); the AFF group had to choose between two grammatically correct phrases in subjunctive; and the RA group completed half referential activities and half affective activities. In the interpretation task all learners heard the subordinate clause of a sentence and had to select one of two possible main clauses that best start the sentence. The choices varied depending on the group as in the recognition task. For referential activities participants had to choose between a habitual action in present tense and a future event in periphrastic future tense, where only one was correct. For affective activities they had to choose
between two grammatically correct phrases (for example in periphrastic future). The researcher provided the same amount of verbs in indicative and subjunctive and in written and aural mode. After treatment, performance was measured by a multiple-choice recognition and interpretation tests. There were 20 recognition items (12 in subjunctive and 8 in indicative) and 20 interpretation items (12 in subjunctive and 8 in indicative) with six different options: three were semantically appropriate and three were not. Out of the three semantically appropriate, only one was grammatically correct. The three choices had present indicative, present subjunctive, and a periphrastic future.

Results from the Repeated Measures ANOVA indicated that type of practice did not have a differential effect on performance evidenced by the lack of statistical differences between groups. From pretests to immediate posttests, all groups made equally significant gains in recognition and interpretation tests. From immediate posttests to delayed posttests, all groups experienced a significant loss in the recognition test; however, only the REF group experienced a significant decline in delayed interpretation test. Henshaw (2012) argued that even when practice does not require learners to attend to the form, if they are structured input (SI) practice, learners may still establish the appropriate form-meaning connections. She also explained this lack of group difference on the fact that all groups received explicit information about the form. According to the author, her study diverged from Sanz and Morgan-Short’s (2004) as these researchers “assumed that some type of corrective feedback was necessary” (Henshaw, 2012, p. 410). Her investigation did not provide evidence for the effectiveness of task-essentialness or the need of feedback given that the AFF group did not receive any pedagogical feedback on the grammatical form and still experienced learning. It should be noted that Sanz and Morgan-Short (2004) did not assume a role for feedback; something that Sanz and colleagues have done in later
investigations (e.g., Stafford et al., 2012; Lado et al., 2013). However, they did acknowledge a limitation of their study regarding feedback: since all groups received TE practice and feedback, they could not determine whether it was TE practice alone or the combination with feedback that leveled the field for all groups (p. 72).

Henshaw’s (2012) results can also be explained not by the lack of role of pedagogical feedback, but rather by the design of the study. By this design, the AFF group received more instances of positive evidence since in the treatment they always had to choose between two grammatically correct options. That could partially explain why the two groups that engaged in affective activities (RA and AFF) were more successful at maintaining their gains at delayed interpretation tests.

Another limitation of this study is related to the exposure of the targeted form. In the treatment, the author provided the same amount of verbs in indicative and in subjunctive. However, in the assessment tests performance on indicative was only used to control for overextension of subjunctive to ensure participants were not selecting subjunctive in all cases. Nonetheless, accuracy scores of recognition and interpretation tests did not include any cases of indicative. In my view, accuracy on indicative should also be measured since temporal clauses are used in both moods, and the instruction of the new mood leads learners to restructure their hypotheses about how Spanish expresses temporality.

It can be assumed that before the study if a participant wished to express the temporality of an event in Spanish—whether referred to the present or to the future—he or she would use indicative as it is the only resource known to them. The use of indicative would work as some kind of ‘default’ setting. As participants go through the study they test and re-structure their hypotheses about how the language works. They have to find a way to include this new mood
(subjunctive) and its use in their interlanguage, thus modifying the ‘default’ setting. The same amount of exposure to both moods is not necessary as participants have had plenty of exposure and instruction on indicative prior the experiment. However, exposure to exemplars in indicative and control of performance of this mood need to be present in order to truly assess any development on temporal clauses.

A final limitation is related to the participants who had to learn for the first time the present subjunctive form. By including learners who are already familiar with the form of the subjunctive, the validity of the study is increased as it reduces the processing load and the possibility of low accuracy due to lack of morphology knowledge. To accurately use present subjunctive learners of Spanish need to overcome two possible hindrances: 1) to master the form of present subjunctive, which hardly differs from the form of present indicative; and 2) to understand in what contexts the subjunctive is necessary and what it means.

**Summary of empirical findings on task-essential practice studies section**

Since Loschky and Bley-Vroman (1993) introduced the notion of task-essentialness, several paper-and-pencil format studies (e.g. Leow, 1997a, 1997b, 1998a, 1998b; Rosa & O’Neill, 1999) and CALI studies (e.g., Rosa & Leow, 2004; Sanz & Morgan-Short, 2004) have incorporated TE practice into their designs. In CALI studies, after the work done by Rosa and Leow (2004) and Sanz and Morgan-Short (2004), several doctoral studies included TE practice in their designs (e.g., Bowles, 2008; Cerezo, 2010; Hsieh, 2007; Lado et al., 2013; Moreno, 2007). From them, only one (Moreno, 2007) examined task-essentialness as an independent variable. The results from Moreno’s study indicated that TE practice did not have a significant role in language learning. However, some
methodological limitations such as small sample sizes and conflation of task-essentialness and feedback render these results inconclusive.

Following Moreno’s (2007) dissertation, other unpublished (e.g., Cerezo, 2010; Hsieh, 2007) and published studies (e.g., Lado et al., 2013; Sanz, Lin, Lado, Stafford & Bowden, 2014; Stafford, Sanz, & Bowden, 2010; Stafford et al., 2012) investigated language learning through TE practice. However, none of these investigations had the goal to determine the role of task-essentialness, and therefore none of them included a group that could control for it thereby leaving the role of TE practice undetermined.

One recent published study that manipulated task-essentialness as a variable is Henshaw (2012). Following the tenets of PI, three groups received explicit information about the subjunctive in temporal clauses. While receiving feedback they had to complete: (1) referential activities (that are TE) in the REF group; (2) affective activities (non TE) in the AFF group; and (3) a combination of both (TE & non TE) in the RA group. Results on recognition and interpretation tests did not show group differences, thus indicating that type of practice did not have a differential effect on performance. The author argued that this homogeneity between groups was due to both the explicit information they all received and the SI activities they all completed. She also assumed no role for feedback since the AFF group did not received any feedback on the form and still experienced learning. However, rather than by a lack of role for feedback, the gains of the AFF group could be explained by the extra exposure to positive evidence this group had during treatment.

In Henshaw’s study, performance on temporal clauses requiring indicative was not considered. However, in my view, accurate competence in temporal clauses should be measured in the moods that the subordinate clause can appear (i.e., indicative and subjunctive). Moreover,
the participants of this investigation had to learn for the first time the present subjunctive form, potentially leading to a high processing load and a low level of accuracy.

The present study overcomes some of these limitations by taking the following measures:

1. The Non TE group has the same exposure to the targeted forms as the TE groups.
2. One of the TE groups does not receive feedback so as to avoid any conflation between task-essentialness and feedback.
3. The Non TE group does not receive feedback.
4. The treatment and the assessment tasks include exemplars in indicative and subjunctive. The numbers of indicative exemplars are fewer than the number of the subjunctive exemplars since the latter is the new form for the participants.
5. Participants are familiar with the present subjunctive form to avoid high processing load and low accuracy due to lack of verb morphology knowledge. They are naïve regarding the use of the subjunctive mood in temporal clauses.

**Empirical Findings on Adverbial Clauses on L2 Spanish**

Only recently have the Spanish temporal adverbial clauses received more attention in the SLA research (e.g., Benati, Lee, & McNulty, 2010; Henshaw, 2012; Kanwit & Geeslin, 2014; Lee & McNulty, 2013). While researchers have shown interest in other uses of subjunctive such as (in its pedagogical term) the subjunctive of doubt (e.g., Collentine, 1998; Farley, 2001a, 2001b, 2004a, 2004b, Fernández, 2008; Lee & Benati, 2007a, 2007b), the use of subjunctive in temporal clauses by L2 learners has hardly been examined. In this section I review the recent developmental and exploratory studies that had subjunctive in temporal clauses as their targeted structures, with the exception of Henshaw (2012) that was already reviewed in the previous section.
Following the principles of Processing Instruction (PI), Benati, Lee, and McNulty (2010) investigated the same targeted structure as the current study: temporal sentences with the adverb *cuando* referring to the future to express uncertainty as to when something will happen in the future versus *cuando* followed by indicative to express a habitual event or action.

Thirty-six participants were enrolled in the first semester of an intermediate level Spanish class and were new to the target form of subjunctive and all of its uses. The first week, participants completed pretests that consisted of a written interpretation test, an aural interpretation test, a written sentence-level production test, an aural sentence-level production test, and a guided-composition. The second week, they were given a lecture on the function of the subjunctive, formation of subjunctive, and processing strategies. After that, participants completed four referential activities where they had to detect the differences between –e and –a in an unstressed final syllable. These activities were presented in a lecture format and the researchers asked participants to volunteer the answer. After the volunteer’s answer of each item, explicit feedback was provided. In the third week, participants met in tutorial groups and received a packet with 12 activities (4 referential and 8 affective). In addition to this, participants were asked to evaluate the information included in the activity by first determining if they thought their instructor would do something and then, using a Likert scale, they had to state how (dis)similar they were to the instructor. Finally, in the fourth week participants completed posttests. As they were coding and analyzing the data, the researchers decided to examine two more variables: whether having participants with different language backgrounds would affect the effects of PI, and whether participants with different language backgrounds would create more or fewer contexts for *cuando* + subjunctive referring to the future to occur during the guided-composition. To this end, they divided their participants into three groups: native
speakers of English studying a foreign language (Spanish) for first time, native speakers of English with experience in other foreign languages, and non-native speakers of English for whom Spanish was their third or fourth language.

The results of the interpretation test revealed that all three groups improved significantly from pre- to posttests, but were not different from each other. The results of the sentence-level production test mirrored those of the interpretation test: all groups significantly improved and were not different from each other. The researchers coded all the subjunctive produced in the guided composition as correct or incorrect and tallied all the correct forms of the subjunctive after *cuando*. The results were also similar to the other two tests: all groups improved significantly but were not different from each other. The researchers concluded that the effects of PI extends to *cuando* + subjunctive not only when learners are interpreting the L2 data, but also when they produce the L2 at the sentence-level and at the discourse level.

This research, however, presented some limitations. The first refers to the low number of participants in each cell. Although the researchers did not set off with the intention to investigate language background as an independent variable, they ultimately decided to examine if such a variable had an impact on the effects of PI. Thus, they had 15 participants whose L1 was English and Spanish was their only L2; 13 participants whose L1 was English but they had knowledge of other languages besides Spanish; and 7 participants whose L1 was not English and due to their language background they were considered trilingual. The researchers did not report the effect sizes in the statistical report, but it could be that with such low number of participants per cell, particularly the one with only 7 participants, the power of the statistical tests was low. Second, given that the study was part of the regular curricular classes, when participants carried out the practice they did so in a lecture format. Researchers presented each item of the task in a slide of a
Power Point presentation, they asked a volunteer to give the answer of each item, and then explicit feedback was provided for everyone to hear. This context may not be the most optimal to foster deep processing from each student individually. When engaged in practice, each learner takes different time to read, think about it, process, and come out with an answer. When the practice is done in a choral way, some learners may take “a passenger seat” approach, that is, let others do the activity without putting a lot of their own effort into it. Computerized individual materials can allow each learner to individually work on the practice, arguably providing a context that better promotes more cognitive effort. Third, the score for the sentence-level production came from two different tests: a written sentence-level production test and an aural sentence-level production test. However, when analyzing the data, the researchers pooled the results of both tests into one score, therefore any potential information on the effects of the modality of the test were lost. Modality could play a role, so it seems more advisable and informative to analyze the scores of each task separately. This, however, could be problematic since each separate test had a very low number of targeted items (only 3). Finally, the researchers did not analyze learners’ production of indicative, which would have been informative to assess whether overproduction of subjunctive in contexts where indicative is necessary occurred.

These last two limitations were addressed in Lee and McNulty (2013), which is a complementary study of Benalti et al. (2010). With the same participants and treatment, Lee and McNulty (2013) analyzed the effects of PI as measured by a written interpretation task, an aural interpretation task, two sentence-level production tasks, and a guided-composition. They separately analyzed participants’ performance of the two-modality sentence-level production tests, which were combined in Benati et al. (2010). In the original study, participants completed two sentence-level production tests: in one, participants were presented with a written fill-in-the-
blank test. The other was a sentence-fragment task for which the participants heard half of a sentence and had to complete it. They also addressed overproduction of subjunctive in indicative contexts, something that was not done in Benati et al. (2010). Finally, they examined learners’ mechanisms to express futurity and uncertainty before and after treatment. To address the first research question about the effects of instruction, they scored the pretests and posttests for correct use of the subjunctive items. There were six targeted items across the two interpretation tasks and six across the two production tasks. The highest score possible for interpretation was six. Since the fill-in-the-blank task and the sentence-fragment task had each three targeted items, three was the highest score possible of each task.

Here I only focus on the outcomes of the tests and variables that were expressively addressed in this study, and for the sake of concision, I do not repeat the results already revised in the summary of Benati et al. (2010). Those results indicated that all learners equally improved (i.e., regardless of language background) in the two tests that were now analyzed separately: in the fill-in-the-blank test and the sentence-fragment task. As for the overgeneralization of subjunctive in non-subjunctive contexts, they found that only five learners out of 35 produced subjunctive when indicative was required, and these learners represented each language background, indicating that language background did not play a role. The researchers argued that these five learners did not overgeneralize the use of the subjunctive, but rather developed a test-taking strategy since they used subjunctive in all contexts possible in the two controlled-production tests. Finally, the researchers analyzed the forms participants produced in the three production tests (i.e., fill-in-the-blank test, sentence-fragment task, and the guided-composition) to examine learners’ mechanisms to express futurity and uncertainty. They found that before treatment, most learners did not recognize or produce futurity or uncertainty contexts that would
require subjunctive, and they mostly and erroneously produced present tense for those contexts. However, at posttests, the production of subjunctive in futurity and uncertainty contexts was significantly greater (79% to 92%).

Since both Lee and McNulty’s and Benati et al.’s studies had the same participants and treatment, some of the limitations of the methodology, for example, the low number of participants per cell, are shared. In both studies the scores from the two interpretation tasks (written and aural) were collapsed into one score, losing any relevant information about modality of the task. Finally, to examine if overgeneralization occurred, the researchers analyzed the items that required indicative in each of the two production tests. However, these were only three items per test, six in total, which is a very small number for any pattern, if there is one, to surface. These two developmental studies in conjunction with Henshaw’s suggest that instruction, in particular processing instruction, seems to have a positive effect on the learning of temporal clauses.

The most recent study on Spanish temporal clauses comes from a L2 variationist approach. Kanwit and Geeslin (2014) examined how non native speakers (NNSs) interpreted indicative and subjunctive forms in temporal sentences, whether the way they interpreted them changed across levels of proficiency, and whether learners differed from native speakers (NSs). They also investigated what linguistic and extralinguistic factors predicted NNSs mood interpretation, whether the factors changed across proficiency levels, and whether the same factors predicted NS mood interpretation. With all this their primary goal was to propose stages of development in the interpretation of subjunctive and indicative forms in adverbial clauses. Ninety-seven participants were divided in four different groups according their proficiency levels: Level 1 (fifth-semester language learners; n=29), Level 2 (400-linguistics course learners;
n= 35), Level 3 (graduate students of Spanish; n=17), and the NSs group (n=16). Participants completed a 24-item task in which they had to indicate if an event took place habitually, if it had not yet occurred, or if both interpretations were possible. Of the 24 items, 12 contained the verb in subjunctive and 12 contained the verb in indicative. The three adverbs under study were cuando “when”, después de que “after”, and hasta que “until”, and all three appeared with indicative and subjunctive. Prescriptively the indicative forms in the subordinate clause would imply a habitual interpretation while the subjunctive form would imply a future, ‘not yet occurred’ interpretation. Researchers manipulated the location of the subordinate clause that could appear preposed or postposed, and the regularity of the verb morphology (i.e., whether the verbs were regular or irregular in their subjunctive form). Moreover, only present tense was used in the verbs of the main clause to express futurity, in order to force the participants to base their interpretations on the mood of the verb in the adverbial clause. Stimuli included items like the following:

(1) Salimos del campus cuando venga María.
   a. María regularly arrives.
   b. María has not yet arrived.
   c. Both interpretations are possible.

(2) Cuando Marta tiene tiempo libre, Julio la visita.
   a. Marta regularly has free time.
   b. Marta has not yet had free time.
   c. Both interpretations are possible.

The results indicated that advanced learners (Level 3) behaved like NSs understanding samples such as (1) as events that have not occurred yet. Interesting, however, is the finding that
advanced participants (Level 3) behaved more prescriptively than NSs in context with present indicative in the subordinate clause, assigning a habitual interpretation in sentences such as (2), whereas the NSs could accept both interpretations. Kanwit and Geeslin (2014) explained this by the fact that most of their advanced participants were also teachers acting more prescriptively, and by the fact that variability is developmental and may only occur at very high levels of proficiency (Regan, Howard, & Lemée, 2009). However, looking at (2) as an example of indicative context, it could be difficult for a native speaker to choose the habitual interpretation (a) that would be what the authors would consider the prescriptive option. It could be argued that (2) has an interpretation such as “every time that Marta has free time, Julio visits her”, which may not, in a clear and distinctive way, denote a habitual meaning, leaving options (b) and (c) as better choices for the native speakers. In fact, the options offered by the authors relate only to the subordinate clause, leaving any information expressed in the main clause as unimportant for the interpretation of the entire sentence. A more suitable option could have been a sentence that summarizes the entire item, such as (3). Options (3a) and (3b) could be considered better interpretations of what is expressed in the stimulus sentence. With all this in mind, the finding that advanced participants behaved more prescriptively than NSs in context with present indicative may be related to the nature of the items employed in the task rather than clear evidence of differences between advanced NNSs and NSs.

(3) Cuando Marta tiene tiempo libre, Julio la visita.

a. Julio’s visit occurs regularly upon Marta’s free time.

b. Marta has not yet had free time and Julio’s visit hasn't occurred.

c. Both interpretations are possible.
To determine which of the linguistic and extralinguistic independent variables contributed to the prediction of response choice, a binomial, step-up/step-down regression for each one of the participant groups was performed. The results of this general model indicated that predictors changed across levels of proficiency, and thus the researchers established four different stages of acquisition that were represented by each group. For Level 1 learners, the lowest level group that represented stage 1, interpretations were predicted by the regularity of the verb morphology and the order of the clause, and adverbs played a role in the interpretation of the verbs in indicative. At stage 2, verbal mood began to play a role, adverbs continued to play a role in interpretation, and a lot of variability characterized this stage. At stage 3, mood played an essential role in interpretation, clause order continued to be relevant in the interpretation of indicative, and individual variability decreased. Finally, at target level, mood played an important role, but it was not the only factor to predict mood interpretation, since lexical factors (i.e., the adverb) and individual factors (i.e., age) conditioned interpretation as well.

This is an interesting study since it shows evidence of different developmental stages in the interpretation of temporal clauses, and it contributes to research on the acquisition of Spanish subjunctive. It also shows evidence of language variation that can inform L2 acquisition. While prescriptively NSs could be expected to base their interpretation of the temporal clauses solely and mostly on verbal mood, it was found that verbal mood was not the only factor to predict mood interpretation, and other factors such as the presence of an adverb and age of the speaker also seemed to play a role. This finding is interesting because it shows that learners at first and second stages shared a predictor (i.e., the presence of an adverb) with NSs to base their mood interpretations. Finally, Kanwit and Geeslin’s (2014) study is noteworthy because it provides
evidence that even native speakers rely on the presence of an adverb when interpreting temporal sentences.

**Summary of empirical findings on adverbial clauses on L2 Spanish section**

While the study of Spanish subjunctive in noun clauses has traditionally been more explored in the SLA field, the Spanish subjunctive in temporal clauses is recently receiving more attention. Most of the studies investigating subjunctive in temporal clauses fall under the PI strand of research, while one other exploratory study follows a variationist approach to L2 acquisition. Thus, Benati et al. (2010), following the tenets of PI, provided instruction on the formation and temporal use of subjunctive to 36 learners. They divided their participants into three groups according to language background: (a) native speakers of English studying Spanish for the first time; (b) native speakers of English with experience in other foreign languages; and (c) non native speakers of English for whom Spanish was their third or fourth language. The results of all assessment tests (i.e., interpretation, sentence-level production, and guided composition) revealed that all groups improved equally, and none were different from each other. The researchers drew positive conclusions with regard to the effects of PI in interpretation and production of *cuando* + subjunctive. This study presented a few limitations, such as the low number of participants per cell, or the fact that the treatment had a lecture format, far from ideal. Two other limitations were addressed in Lee and McNulty (2013), namely, learners’ performance of indicative to account for possible overgeneralization of subjunctive, and modality of the two production tests (aural production and written production, which in Benati et al. (2010) were previously collapsed). Like in Benati et al. (2010), the results indicated that all learners, regardless of language background, performed equally, and only five out of 35 produced subjunctive in indicative contexts, which led the researchers to conclude that no
overgeneralization really occurred. Some of the limitations regarding methodology or low number of participants per cell that Benati et al. had are shared in Lee and McNulty’s study. In addition, although accounting for overgeneralization of subjunctive is a strength to determine real learning of subjunctive, the low number of items designed for this (only three items per test, six in general), could have made it difficult for any pattern to surface. All in all, the results of these processing instruction studies seem to suggest that PI has a positive effect in the learning of temporal clauses.

From a variationist approach to L2 acquisition, Kanwit and Geeslin (2014) had the goal to propose stages of development in the interpretation of subjunctive and indicative forms in adverbial clauses. They specifically examined how NNSs and NSs interpreted indicative and subjunctive forms in temporal sentences, and whether there were changes in interpretation across levels of proficiencies. They also investigated what linguistic and extralinguistic factors predicted mood interpretation. Their 97 participants were divided into different groups according to levels of proficiency (Level 1, Level 2, Level 3, and NSs), and they completed a task where they had to indicate if an event took place habitually, if it had not yet occurred, or if both interpretations were possible. The results indicated that advanced learners (Level 3) behaved like NSs understanding samples that included subjunctive in the subordinate clause as events that have not occurred yet. They also behaved more prescriptively than NSs in context with present indicative in the subordinate clause, to which they assigned a habitual interpretation, whereas the NSs could accept both interpretations (i.e., not occurred yet and habitual). Whereas the authors explained this result based on the fact that their participants in the Level 3 were mostly Spanish instructors, who acted more prescriptively, a closer look at their materials may reveal that some items did not unequivocally adhere to one meaning, namely, the habitual meaning, which could
have led the NSs to choose other options as more suitable answers. With regard to stages of acquisition, the researchers established four stages that paired with each proficiency level. An interesting result of this is that while verbal mood started to play a role at stage 2 (an intermediate level of proficiency), it is not until stage 3 (a more advanced level) that it becomes essential to interpret indicative and subjunctive in temporal sentences. However, for NSs verbal mood was not the only source to interpret temporal sentences, and the presence of an adverb and the age of the speaker also seemed to play a role.

While it is clear that the study of Spanish subjunctive in temporal clauses is gaining more attention, these are only a few studies, from only two research strands (PI and L2 variation), which leave the field still quite open for more research on the acquisition of temporal clauses by Spanish learners.

**Research Questions**

The current study aims to address the above mentioned gaps by investigating the effects of different types of feedback (more versus less explicit) and the amount of feedback episodes required in the development of learners’ Spanish temporal and causal clauses, while completing input-based practice (TE versus non TE). The following research questions will inform the present investigation:

1. RQ1. A) Does type of feedback (Rules vs. Right/Wrong) have a differential effect on L2 grammar development, as measured by interpretation and production tests, immediately after the treatment?
   
   RQ1. B) Is this effect maintained after 4 weeks?

2. RQ2. A) Is there a minimum threshold regarding the amount of feedback from which learners start to attain accuracy in a more consistent way?
RQ2. B) If so, does the threshold differ depending on the type of feedback?

RQ2. C) Is amount of feedback significantly related to development?

3. RQ3. A) Does type of input-based practice (TE vs. NTE [non-taskessential]) have a differential effect on L2 grammar development, as measured by interpretation and production tests, immediately after the treatment?

RQ3. B) Is this effect maintained after 4 weeks?
CHAPTER 3: THE PRESENT STUDY

Method

Participants

Seventy-eight students of Spanish enrolled in a fourth semester Intermediate II course Spanish Language Instruction Program at a university in the United States participated in the study. In this program students met three times a week and completed 150 minutes of instruction per week (approx. 37.5 hours in total). The Intermediate II proficiency level had been chosen to assure that participants had been previously exposed to the Spanish present subjunctive, but not to the temporal subordinate clauses. It was not the purpose of this study to introduce the learners to the present subjunctive form for the first time, but rather the use of present subjunctive and present indicative in adverbial temporal subordinate clauses.

The learners who participated in this research had the following characteristics: age (18-25 years old), English as (one of) native language(s), little to no prior knowledge of any other Romance language, and intermediate proficiency level. Table 1 summarizes these traits. From the original pool of seventy-eight participants, four participants were discarded due to their L1s (i.e., Hindi, Chinese, Korean, and Russian). One other participant reported to have Spanish spoken at home from birth until the age of 5 but having received formal instruction in Spanish at the age of 12. That same participant also reported to have spent 1.5 years in France at the age of 5 and having received instruction in French at that time and in high school senior year. The rest of the participants reported English as their native language from among whom six also reported speaking other non-Romance native languages (i.e., Tigrinya, Pashto, Urdu, Yoruba, Bulgarian, German, and Tamil). After further correspondence with these participants, all except one reported native language dominance in English, therefore that one participant was also discarded.
Additionally, some other participants reported different levels of proficiency in other languages that included Russian (1), Chinese (3), Arabic (1), French (2), Hebrew (2), Latin (1). The data from those claiming to have some proficiency and some exposure in French were further examined to determine their level of proficiency. Based on the self-reports and correspondence with these participants, it was decided to exclude one more person, who had had 3 years of formal exposure to French right before starting college level studies.

Table 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Age</td>
<td>18-25 years old</td>
</tr>
<tr>
<td>b) Native Language(s)</td>
<td>English (+ Non-Romance language)</td>
</tr>
<tr>
<td>c) Prior Study: Romance Languages</td>
<td>Little to no formal study</td>
</tr>
<tr>
<td>d) Proficiency</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Thus, the final number of participants was seventy-one. All these participants completed the entire study, that is, during the pre-, immediate posttest, and delayed posttest stages, with a 100% retention level. They were between 18 and 22 years old ($M = 19.29$) and their proficiency level was determined based on the institutional enrollment. Participants were randomly assigned to four different groups resulting from the combination of the two between-subject variables, namely, type of feedback (Rules [R], Right/Wrong feedback [R/W] or none [NF]), and task-essentialness (presence [TE] or absence [NTE]). Table 2 shows the distribution of participants per condition.
Table 2

*Distribution of participants per condition*

<table>
<thead>
<tr>
<th>Group</th>
<th>Features</th>
<th>N</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Rules [R]</td>
<td>Task-essential w/ rules in the feedback</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>(2) Right/Wrong [R/W]</td>
<td>Task-essential w/ right/wrong feedback</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>(3) NoFeedback [NF]</td>
<td>Task-essential w/o feedback</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>(4) No TE [NTE]</td>
<td>Non task-essential practice w/o feedback</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

All participants completed a review session prior to carrying out the experiment. In the review, participants practiced some drills of present indicative, present subjunctive, and preterit. No instruction on the usage of present subjunctive was provided. The review also covered the usage of present tense with a future value, which was crucial for the experiment of the present study. Moreover, in the review, participants got trained on the format of the pictures used in the experiment and were eventually informed on the meaning of each format. A detailed description of the review session is included in the material section. Participants also completed an exit questionnaire at the end of the study to find out whether they had consulted any source to obtain information about any grammatical point related to the study. Participants reported studying grammar for their Spanish classes, but no intentional study of any structure they saw in the study.

For their time, participants were awarded three extra credit points on the lowest grade of one of their class tests. Permission to award extra credit was obtained from the Director of the Spanish Language Program. Permission to conduct research with human subjects was obtained from the Social and Behavioral Sciences committee of the Institutional Review Board of the university (IRB approval # 2015-0029).
Targeted structure

The targeted structure chosen was temporal adverbial clauses referring to the present and the future (e.g., Cuando llegue a casa, preparo la cena ‘When I get-SUBJ home, I prepare-PRES w/future value dinner’). In the following sections the targeted structure is discussed in detail.

Adverbial temporal clauses. As is the case with all adverbial subordinate clauses, temporal clauses are subordinate clauses that function as adverbs and occupy the same position in the sentence. They serve as a modifier of the main clause and they help to locate in time the action expressed in the main clause (see example 1) (Campos, 1993, p. 150).

(1) María se ducha cuando se levanta

‘María showers when she wakes up’

In both English and Spanish, temporal clauses define the time when the event in the main clause occurs through relations of simultaneity, succession, or anteriority. These relations are expressed through conjunctions such as when (cuando), while (mientras), as soon as (tan pronto como), before (antes (de) que), after (después (de) que), etc.

In general, temporal clauses require the indicative mood when the subordinate clause denotes factual events (past, present, or habitual) as in (2) and (3). However, when the subordinate clause denotes non-factual events referring to a future eventuality, the subjunctive is required, as observed in (4) (Pérez Saldanya, 1999).

(2) Cuando estuvo en Madrid, visitó muchos museos.

‘When she was in Madrid, she visited many museums.’

(3) Salgo a pasear por la playa cuando acabo de cenar.

‘I go for a walk on the beach when I finish dinner.’

(4) Díselo en cuanto puedas.
‘Tell him as soon as you can.’

The main clause requires a variety of tenses depending on the time of the events expressed in the subordinate clause. Thus, when the subordinate clause denotes factual events referring to the present or the past, the verb of the main clause can be conjugated in present, preterit, imperfect, present perfect, etc. With non-factual events referring to the future in the subordinate clause, the main clause uses simple future, periphrastic future, imperative, and present tense with future value. The present study focuses on temporal clauses that denote factual events (present or habitual) and non-factual events (future). These two uses have been selected due to the similarity in the form of the tenses required in the subordinate clause, namely present indicative and present subjunctive.

**Present indicative with future value in the main clause.** As previously mentioned, with non-factual events a variety of tenses can appear in the main clause: simple future, periphrastic future, imperative, and present tense with future value. For the present study, present tense with future value has been chosen. One of the reasons behind this choice is to enhance the task-essentialness of the task. More details about task-essentialness will follow. Another reason is that present with future value is considered an expression of futurity, and has been recently studied in SLA with learners of Spanish (e.g., Gudmestad & Geeslin, 2013; Kanwit & Geeslin, 2014).

The goal of Gudmestad and Geeslin’s (2013) study was to explore the developmental path of acquiring variable future-time reference in Spanish. They included learners at several proficiency levels and offered predictive accounts of how multiple linguistic factors (i.e., lexical temporal indicators, (un)certainty makers, and temporal distance) influence verb selection for native speakers (NSs) and learners. The three future expressions under study were periphrastic future, morphological future, and present indicative. Their results indicated that participants
(both native speakers and nonnative speakers, NNSs) used the three future expressions at different rates. Even though present indicative was not as frequently selected as periphrastic future or morphological future by the L2 learners of any proficiency group, their data showed that as proficiency increased, learners demonstrated an understanding of the fact that the present indicative can refer to events in the future, rather than only to present-time events (p. 70).

In another study, Kanwit and Geeslin (2014), discussed in the review of the literature section, specifically examined present indicative with future value in temporal clauses with both NSs and advanced NNSs at three different levels of proficiency. The goal of their study was to examine how NNSs interpret indicative and subjunctive forms in temporal clauses, how this changes across levels of proficiency, and how learners differ from NSs. Prescriptively, the indicative forms in the subordinate clause would imply a habitual interpretation while the subjunctive form would imply a future, ‘not yet occurred’ interpretation. Participants were presented with sentences and were asked to choose the correct interpretation of the sentence. Their stimuli included items like the following:

(5) Salimos del campus cuando venga María.
   d. María regularly arrives.
   e. María has not yet arrived.
   f. Both interpretations are possible.

(6) Cuando Marta tiene tiempo libre, Julio la visita.
   d. Marta regularly has free time.
   e. Marta has not yet had free time.
   f. Both interpretations are possible.
The results indicated that advanced learners behaved like NSs understanding samples such as (5) as events that have not occurred yet. This result is particularly relevant to this dissertation as it yields evidence that present tense with future value is clearly understood as an expression of futurity in temporal clauses. The authors also found that advanced participants behaved more prescriptively than NSs in items with present indicative in the subordinate clause, assigning a habitual interpretation in sentences such as (6), whereas the NSs could accept both habitual and ‘not occurred yet’ interpretations. Kanwit and Geeslin (2014) explained this by the fact that most of their advanced participants were also teachers acting more prescriptively, and by the fact that variability is developmental and may only occur at very high levels of proficiency (Regan, Howard, & Lemée, 2009). However, looking at item (6) that is presented as an example of the habitual interpretation encoded in the indicative mood, it could be difficult for a native speaker to choose option (a), the habitual interpretation, which would be what the authors consider the prescriptive option. It could be argued that item (6) has an interpretation such as “every time that Marta has free time, Julio visits her”, which may not, in a clear and distinctive way, denote that Marta regularly has free time, leaving options (b) and (c) as suitable choices as (a) for the native speakers. In fact, the options offered by the authors relate only to the subordinate part of the sentence, leaving any information expressed in the main clause as unimportant for the interpretation of the entire sentence. A more suitable option could have been a sentence that summarizes the entire item, such as in (7a) and (7b). Options (7a) and (7b) offer an interpretation of what is expressed in the entire stimulus sentence. With all this in mind, the finding that advanced participants behaved more prescriptively than NSs in context with present indicative may be related to the nature of the items employed in the task rather than a clear evidence of differences between advanced NNSs and NSs.
(7) Cuando Marta tiene tiempo libre, Julio la visita.

d. Julio’s visit occurs regularly upon Marta’s free time.
e. Marta has not yet had free time and Julio’s visit hasn't occurred.
f. Both interpretations are possible.

Despite this fact, both studies, Gudmestad and Geeslin (2013) and Kanwit and Geeslin (2014), show evidence that present tense with future value is a valid expression of future for both native speakers and L2 learners, and denotes events that have not yet occurred, thus being a valid expression in non-factual temporal clauses.

**Challenges of the structure.** Temporal clauses that denote factual events (present or habitual) should not pose a challenge for the participants of this study, who already know how these sentences are formed in Spanish. However, temporal sentences that denote non-factual events (future) are new to them and the introduction of this new expression should lead to a re-structuring of the paradigm.

Temporal clauses that denote non-factual events require the use of subjunctive in the subordinate clause. Although participants are not new to the morphology of the present subjunctive, mastering this form can pose a challenge for them. Both present indicative and present subjunctive are formally very similar and sometimes differ only in one vowel (e.g., él habla vs. él hable), making the present subjunctive form relatively nonsalient when compared to the present indicative form.

Problems due to transfer can also be expected with the introduction of this new expression. While English uses one form for both temporal clauses that denote factual and non-factual events, Spanish has two forms, one for each use. Thus, the verb ‘finishes’ refers to a habitual action in (8), and a future action in (9).
(8) I go to the gym when my English class finishes.

(9) I’ll go to the gym when my English class finishes.

Items (10) and (11) are Spanish examples of factual and no-factual events respectively, each one with different morphology, unlike English. In (10) the present indicative is used (‘termina’), and in (11) the present subjunctive is used (‘termine’).

(10) Voy al gimnasio cuando termina mi clase de inglés.
(11) Voy al gimnasio cuando termine mi clase de inglés.

Considering all of this, it is expected that L1 English learners of Spanish (a) will transfer their patterns (i.e., a single verb ending that encodes both present and future meaning) and assume there is no difference between the expression of present/habitual actions and future actions in temporal clauses; and (b) will not notice the very subtle different verb endings of the present indicative and present subjunctive in the subordinate clause. Here is where task-essentialness and feedback may play an important role in aiding learners to notice the targeted items and test their hypotheses.

**Experimental Design**

The present study follows a pretest-posttest-delayed posttest design. What follows is a description of the variables studied in this investigation.

**L2 development**

L2 development in this study is operationalized as performance on an interpretation test and controlled written production test administered immediately after treatment and again four weeks later. A detailed explanation of each test is provided in the Materials section.
Type and amount of feedback

Since this study addresses the effectiveness of feedback in general, and type of feedback in particular, type of feedback is an independent between-subject variable with three levels: feedback with rules, right/wrong feedback, and no feedback. The feedback provided was both positive and negative, and used English as the vehicular language. It contained no positive evidence of Spanish, which is only present in the input-based practice that all participants complete. By doing this, exposure to positive evidence is controlled in all the experimental groups.

During the treatment phase participants had to choose among three different options in a picture-matching interpretation task. In the feedback with rules group [R], participants who chose a wrong answer received grammatical explanation in English on the Spanish temporal clauses. When participants chose the correct answer, a smiley face and a sentence with positive feedback appeared in the middle of the screen (e.g., :-) That is correct!). The grammar explanation provided differed depending on the time frame of the item. That is, when they chose a wrong answer for an item referring to the present, only grammatical rules relevant to present were presented (see Figure 1); conversely, when they chose a wrong answer for an item referring to the future, only grammatical rules relevant to future were presented (see Figure 2). In the right/wrong feedback condition [Y/N], participants who chose a wrong answer received a message such as ‘uh-oh ☹ Press any key to try again’ (see Figure 3). When they chose the correct answer, the same type of positive message that the feedback with rules group obtained was displayed.
Consider:

**What is the time frame of this event?**

- When the time frame of the event is the PRESENT (the sentence talks about normal, habitual actions or known truths, like in this case), the PRESENT TENSE should be used in both the main and the subordinate clauses.

Please, press any key to TRY AGAIN

*Figure 1. Example of feedback with rules [R] after an incorrect answer of an item referring to the present*

Consider:

**What is the time frame of this event?**

- When the time frame of the event is the FUTURE, like in this case, (the sentence talks about something that will occur in the FUTURE), the PRESENT SUBJUNCTIVE should be used in the subordinate clause and future or PRESENT INDICATIVE with future value can be used in the main clause.

Please, press any key to TRY AGAIN

*Figure 2. Example of feedback with rules [R] after an incorrect answer of an item referring to the future*

*Figure 3. Example of right/wrong feedback condition [Y/N]*
In addition to type of feedback, this dissertation sought to investigate the amount of feedback necessary in order to obtain the correct response. Therefore, in the feedback groups (regardless of feedback type) participants could repeat each trial as many times as they needed until they chose the correct answer. This feature of ‘try again’ allowed different amounts of feedback for different learners. Table 3 shows the characteristics of the feedback provided to each group.

Table 3

<table>
<thead>
<tr>
<th>Characteristics of feedback per groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback groups</td>
</tr>
<tr>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Rules [R]</td>
</tr>
<tr>
<td>Right/Wrong [R/W]</td>
</tr>
</tbody>
</table>

Task-essentialness

While input-focused practice is present in the four conditions, including the control group (NTE), task-essentialness is present in three out of the four groups thereby making this a between-subjects variable. Task-essentialness is one of the three types of involvement of a grammatical structure in a task, the other two being task-naturalness and task-utility (Loschky & Bley-Vroman, 1993, p. 132). As explained in their article, these three differ in the degree of manipulation the task designer exerts, with task-naturalness being the less manipulated, and task-essentialness being the most manipulated. With this in mind and as diagrammed in Figure 4, rather than a dichotomous feature that indicates presence or absence, task-essentialness can be considered a continuum of a feature of a task where the less controlled the task is the more
natural (and the less essential) it is, and the more controlled the task is the most essential (and the less natural) it is.

<table>
<thead>
<tr>
<th>Essential</th>
<th>Useful</th>
<th>Natural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis formation</td>
<td>Automatization</td>
<td></td>
</tr>
<tr>
<td>More control</td>
<td>Less control</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>Production</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4. Relationship among four basic dimensions of structure-based tasks (Loschky & Bley-Vroman, 1993, p. 142)*

With this in mind, task-essentialness is operationalized in the present study in relation to one condition (TE) versus the other (NTE). It is by comparing the characteristics of the tasks and what participants have to do in the TE condition versus the non-TE condition that the tasks in the TE condition can be considered task-essential. In order to control for lack of task-essentialness in the control group, participants in this condition have to match a sentence with one of three possible pictures. By its nature, the task does not direct them to pay attention to the targeted items. In this condition, participants can fulfill the task by just processing some key vocabulary words. The other three groups have to complete a TE task, where they need to match a picture that is described in only one of three sentences provided. A more elaborated description of the TE and NTE tasks is detailed in the materials section. Table 4 shows a summary of the characteristics of each experimental condition.
Table 4

Summary of the traits of each experimental condition

<table>
<thead>
<tr>
<th>Groups</th>
<th>Positive evidence in the form of input-based practice</th>
<th>Task-essentialness</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules [R]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Right/Wrong [R/W]</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>NoFeedback [NF]</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>[NTE]</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

Materials

Language background questionnaire

This questionnaire elicited biodata information about participants such as age, gender, and native language(s). Participants reported information about any languages they indicated knowing, age of first exposure of first and subsequent languages, amount of formal instruction, self-rated proficiency in reading, speaking, listening, and writing skills, and frequency of use of any languages other than native. The questionnaire also addressed how many hours a week they spent outside of classroom using their nonnative language(s) in academic (e.g., homework, preparation for quizzes and exams) and non-academic activities (e.g., social interactions, watching T.V., browsing the Internet).

Review session

After reporting to the language laboratory and before starting the treatment, participants were presented with a review on the morphology of two Spanish tenses, the time reference of tenses such as present, preterit, and future, the values of present indicative as an expression of present actions and future actions, and the format of the pictures. The language used during the presentation was English.
The tenses reviewed were the present subjunctive and preterit. As preterit is in the treatment as a distractor, it was also included in the review session. Participants were presented with verb stems and endings for each tense, and then asked to conjugate regular and irregular verbs of the different conjugations. One tense was reviewed at a time. After conjugating the verbs for each tense, participants were provided with the correct answers so that they could check their work. The review of preterit was brief since this tense was expected to be well known (although maybe not mastered) by the participants, and it only functioned as a distractor. The review of the subjunctive was longer as this tense was newer to them. Since the aim of this dissertation was not to teach the morphology of a new tense (i.e., present subjunctive), participants were recruited once they had formally studied this tense (but not its temporal use) in class. This review session also aimed to minimize any possible problems with the verb morphology. The review of the present subjunctive included explicit information on verb endings for the three conjugations, regular verbs, and irregular verbs, but nothing related to when subjunctive is used in Spanish.

Once the morphology was reviewed, the presentation went on to the usage of present, future and preterit tenses to indicate present, future, and past actions respectively with the aim to introduce the present and future value of present indicative. It then asked participants to locate in time the actions expressed in a sentence and depicted in a picture (see Figure 5). After checking their answers, a brief explanation of the format of the pictures was provided. The presentation again explicitly stated the two uses of present indicative and presented graphic examples (Figure 6). Appendix D includes all the slides of the review section, and Appendix E includes the answer sheet participants completed during the review.
Figure 5. Excerpt of the review session asking participants to locate in time the actions of the sentence

Figure 6. Excerpt of the review session showing participants an example of the format of the picture with present indicative with future value
**Treatment**

The treatment consisted of a picture-matching interpretation task with temporal clauses. To avoid any potential impediment from lack of vocabulary knowledge, the vocabulary used during the treatment included high frequency words that are typically known by third semester Spanish students, and glosses of less frequent words. The materials for both the TE condition and NTE condition are explained below. Appendix F includes all the stimuli used during the treatment.

**TE practice.** For the TE practice of this treatment participants were presented with a picture and three different sentences. Their task was to choose the sentence that best describes what the picture conveys. The three choices provided are grammatical, valid, and meaningful, but only one is the correct option given the context of the event. The only conjunction included in the subordinate clauses was **cuando** (when). This decision was made to eliminate possible overload of learners’ cognitive resources trying to distinguish when to use different conjunctions, which is not the goal of this study.

The practice consisted of a total of 49 items and the distribution’s break down is as follows: 29 targeted items (items referring to the future), 12 counterpart items (items referring to the present), and 8 distractors (items referring to the past). Items in present tense were included as counterparts of the future items. Participants were expected to know how to express temporality referring to the present, something that is seen early in language classes, but they were not expected to know how to express temporality referring to the future. The difference in the number of items referring to the future and the present is so that participants were not overexposed to a structure already known (i.e., temporal sentences referring to the present) and
were provided with enough exemplars of the newer structure (i.e., temporal sentences referring to the future). The eight distractors are pictures conveying temporal clauses referring to the past. All the treatment items only used 1<sup>st</sup> and 3<sup>rd</sup> person singular. Table 5 summarizes the distribution.

Table 5

*Distribution of TE items*

<table>
<thead>
<tr>
<th>Temporal items</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Targeted items referring to the future</td>
</tr>
<tr>
<td>12</td>
<td>Counterpart items referring to the present</td>
</tr>
<tr>
<td>8</td>
<td>Distractors referring to the past</td>
</tr>
<tr>
<td><strong>49</strong></td>
<td><strong>TOTAL</strong></td>
</tr>
</tbody>
</table>

Pictures conveying temporal clauses referring to the present are drawn seven times to indicate a daily routine. Pictures conveying temporal clauses referring to the future are drawn in a thought bubble to indicate it is a future thought. Dates and times are also provided and are necessary to help locate the events in time. See Figures 7, 8 and 9 below.
a. Voy a Francia cuando aprendo francés.
b. Fui a Francia cuando aprendí francés.
c. Voy a Francia cuando aprenda francés.

Figure 7. TE condition in treatment with targeted items referring to the future

a. Cuando hago ejercicio me ducho.
b. Cuando hice ejercicio me duché.
c. Cuando haga ejercicio me ducho.

Figure 8. TE condition in treatment with counterpart items referring to the present
Figure 9. TE condition in treatment with distractor items referring to the past

The design of the treatment presentation was semi-randomized. It had a strict presentation of tense blocks in order to examine a possible threshold in the amount of feedback and practice, but the items within each block were randomized. Figure 10 shows the design.
<table>
<thead>
<tr>
<th>Present block 1 (2 items)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjunctive block 1 (4 items)</td>
<td></td>
</tr>
<tr>
<td>Present block 2 (2 items)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 2 (4 items)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 3 (3 items)</td>
<td></td>
</tr>
<tr>
<td>Present block 3 (1 item)</td>
<td></td>
</tr>
<tr>
<td>Present block 4 (2 items)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 4 (3 items)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 5 (2 items)</td>
<td></td>
</tr>
<tr>
<td>Present block 5 (1 item)</td>
<td></td>
</tr>
<tr>
<td>Present block 6 (1 item)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 6 (3 items)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 7 (2 items)</td>
<td></td>
</tr>
<tr>
<td>Present block 7 (1 item)</td>
<td></td>
</tr>
<tr>
<td>Present block 8 (1 item)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 8 (4 items)</td>
<td></td>
</tr>
<tr>
<td>Present block 9 (1 item)</td>
<td></td>
</tr>
<tr>
<td>Subjunctive block 9 (4 items)</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 10. Design of the treatment presentation*

**NTE practice.** Participants in the control group also completed a picture-matching interpretation task where they were presented with one sentence and three pictures as options. The pictures depicted events that can be easily distinguished by knowing basic vocabulary. No
connection between the targeted structure and the meaning of the picture needs to be made in order to successfully complete the task; therefore this task does not direct learners’ attention to the relevant targeted items. There is a total of 49 items with the same distribution as the treatment. Figure 11, 12, and 13 provide examples of items referring to the future, present and past, respectively.

*Figure 11. Example of stimulus of the NTE condition with items referring to the future*
Figure 12. Example of stimulus of the NTE condition with items referring to the present

Figure 13. Example of stimulus of the NTE condition with items referring to the past
Assessments

Two different tests serve as assessment tools. In both tests, as in the treatment, the vocabulary used included high frequency words and glosses of less frequent words. Below is a description of both tests. Appendices G and H include all the stimuli of the interpretation test and the written production test respectively.

Picture-matching interpretation test

Similar to the treatment task, this interpretation test is a picture-matching task with 44 items. There are 21 items referring to the future, 19 items referring to the present, and 4 distractors referring to the past (Table 6). In this test participants were presented with one picture and half of a sentence that needed to be completed with one of four choices (i.e., present indicative, present subjunctive, past, and future). Future tense was included in the testing phase in order to gauge whether participants have understood the rule regarding the use of present subjunctive with future events. Figures 14, 15, and 16 provide examples of items referring to the future, present, and past respectively.

Figure 14. Example of assessment test with items referring to the future
**Figure 15.** Example of assessment test with items referring to the present

- a. ... cuando preparo el desayuno.
- b. ... cuando prepararé el desayuno.
- c. ... cuando preparé el desayuno.
- d. ... cuando prepare el desayuno.

**Figure 16.** Example of assessment test with items referring to the past

- a. ...cuando voy a Australia.
- b. ...cuando fui a Australia.
- c. ...cuando vaya a Australia.
- d. ...cuando iré a Australia.
As in the treatment, the frequency with which the subordinate clause with *cuando* appears before and after the main clause is controlled for with half of the items having the subordinate clause before the main clause and the other half after. Also, as in the treatment, the order of the four answer options and the order in which each tense appeared were randomized across the experiment. Items referring to three points of time were randomized.

Two versions of the test were developed. One was used for the pretests, the other was used for the posttests and the first version was reutilized for the delayed posttests. Participants randomly completed one version or the other at pretest and posttest, in such a way that no one version was specifically used for the pretests or posttests. At the delayed posttests, I made sure to provide each participant with the version they completed at the pretests.

Table 6

*Summary of the distribution of temporal items in the interpretation task*

<table>
<thead>
<tr>
<th>Distribution of temporal items in the interpretation task</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 21 targeted items referring to future:</td>
</tr>
<tr>
<td>• 19 counterpart items referring to present:</td>
</tr>
<tr>
<td>• 4 distractors referring to past</td>
</tr>
<tr>
<td>44 Total</td>
</tr>
</tbody>
</table>

**Controlled written production test**

A fill-in-the-blank test with 44 items served as the controlled written production test. There were 20 sentences referring to the present, 20 referring to the future and four distractors referring to the past (Table 7). The frequency with which the subordinate clause appears before and after the main clause was counterbalanced. Participants had to produce the missing verb in
the subordinate clause. As in the treatment, there were only two forms addressed in this study: 1st and 3rd person singular. In order to control that the point in time each item referred to was clear, temporal makers were included. See Figure 17 for examples of the controlled written production test with items referring to the future and to the present.

El año que viene empiezo una nueva vida en otra ciudad cuando __________________ (yo, terminar) la universidad.

Cada semana cuando __________________ (yo, limpiar) mi habitación encuentro cosas perdidas.

*Figure 17. Example of the controlled written production test with future and present items*

Table 7

*Summary of the distribution of temporal items in the production task*

<table>
<thead>
<tr>
<th>Distribution of temporal items in the production task</th>
</tr>
</thead>
<tbody>
<tr>
<td>• 20 targeted items referring to future</td>
</tr>
<tr>
<td>• 20 counterpart items referring to present</td>
</tr>
<tr>
<td>• 4 distractors referring to past</td>
</tr>
<tr>
<td>44 Total</td>
</tr>
</tbody>
</table>

**Exit questionnaire**

At the end of the last session at week 6, participants completed an exit questionnaire. The questionnaire elicited information such as whether the review session was helpful and why, whether the pictures of the treatment and assessment tools were clear, whether the dates in the
pictures were helpful and why, whether the vocabulary was an impediment to understand the sentences, and whether they had any extra exposure to the targeted items between sessions.

**Procedure**

This study follows a pretest (week 1) – treatment - immediate posttest (week 2) - delayed posttest (week 6) design. In the first session, participants reported to the language laboratory where they completed the review session. After that, they completed the pretests that comprised the interpretation picture-matching test and the controlled written production test. SuperLab 5 is the software used to create the treatment and assessment tests while the review session was supported in Microsoft PowerPoint. Upon completion of the pretest, participants filled out a language background questionnaire. A week later, in session two, participants were randomly assigned to one of the four experimental groups and completed the treatment and control tasks accordingly. Immediately after completing these tasks, the assessment tests were administered. During the last session, four weeks later, participants reported to the laboratory to complete the third set of assessment tests and the exit questionnaire. Table 8 summarizes the procedure of the study.
Table 8

Procedure

<table>
<thead>
<tr>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Review session (20 min)</td>
<td>Pretests (45 min)</td>
<td>Background questionnaire (10 min)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assessment tests (45 min)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment (30 min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assessment tests (45 min)</td>
<td>Exit questionnaire (10 min)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment tests (45 min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coding and Scoring

The scoring of both assessment tools involved a two-step analysis, in which the use of temporal clauses referring to the future was first examined followed by the use of temporal clauses referring to the present. This two-step analysis allows a better determination of the origin of the correct answers (i.e., whether the correct data is driven by the correct performance of the new targeted structure, the temporal clauses referring to the future, or by the correct performance of the already known structure, the temporal clauses referring to the present). It also helped to account for overgeneralization of subjunctive in contexts where indicative is required. Thus, in the interpretation test with items referring to the future, participants were granted 1 point for the answer that included the correct use of subjunctive and 0 for any other answer (i.e., answers with the verb of the subordinate clause in present tense, future tense or past tense). The same procedure was followed with items referring to the present: 1 point for the answer that
included the correct use of indicative and 0 for any other answer. The maximum score for this test was 40.

The scoring for the production test involved two different procedures: lenient and strict. For the lenient scoring, participants received 1 point as long as they used the correct mood (either indicative for the items referring to the present, or subjunctive for the items referring to the future), not considering any possible errors in conjugation such as verb-subject agreement mistakes or spelling mistakes. For the strict scoring, participants only received 1 point if they produced the exact correct answer. The maximum score for this test was 40.
CHAPTER 4: ANALYSES AND RESULTS

Statistical Analyses

Mixed-effects modeling was employed to analyze the data. In mixed-effects models, the independent variables are modeled with fixed effects and the random variation of the sample is modeled with random effects. As recently noted in the psycholinguistic and SLA literature (e.g., Linck, Schwieter, & Sunderman, 2012; Linck & Cunning, 2015), mixed-effects models offer a number of benefits over traditional Analysis of Variance (ANOVA) or multiple regression methods. First, by-subject ($F_1$) and by-item analyses ($F_2$) can be combined within one analysis, allowing results to generalize across people and items. While random variation can arise from the participants, it can also arise from the materials tested. Traditionally, as Linck and Cunning (2015) note, SLA researchers conduct independent analyses to account for this: one analysis with the data averaged over participants ($F_1$) and another analysis with the data averaged over items ($F_2$). Mixed-effects models offer the benefit of conducting both analyses simultaneously. Second, averaging across trials is not necessary in mixed-effects models, as it is for ANOVAs, as each individual trial for each participant is included in the analysis. Third, mixed-effects models allow for the examination of predictors at multiple levels (e.g., time as a trial-level variable vs. group as a subject-level variable), which in turn makes the model not subject to the assumption of independence of observations that are typical in an ANOVA.

Since mixed-effects models are still not widely used in the field of SLA, there are not clear standards as to model building and significance testing. Recently, Linck and Cunning (2015) offered some guidelines in this regard. They recommend a description of the structure of the model (fixed- and random effects), the software used to run the model, and the methods of estimation. Thus, the multilevel models (MLMs) of the present study were conducted with the
lme4 package version 1.1-9 (Bates, Maechler, Bolker, & Walker, 2015) in R version 3.2.0 (R Core Team, 2015) for logistic multilevel modeling using the “bobyqa” optimizer.

To analyze all the data of the present study, growth models were used to capture learning trajectories. Growth models are valuable for researchers interested in capturing change and development over time, while including information regarding within-person change, average change, between-person differences in change, and determinants of change (Baltes & Nesselroade, 1979; Bryk & Raudenbush, 1992; McArdle, 1986, 1988; Meredith & Tisak, 1990). Another advantage of growth models is that they show the probability of a correct or an incorrect response on an item given the predictors in the model instead of the mean accuracy.

Results

For the sake of clarity, I repeat each research question before explaining the models and the results.

Research question #1

RQ1-Does type of feedback (Rules vs. Right/Wrong) have a differential effect on L2 grammar development, as measured by interpretation and production tests, immediately after the treatment? Is this effect maintained after 4 weeks?

To investigate development over time, a series of multilevel models were posited and refined to find the best fitting model. Ultimately Piecewise conditional growth model was used to model the accuracy data from the interpretation and the production tests. This final model was the result of slowly building up on previous models and determining its goodness-of-fit via likelihood ratio tests (LRTs).
For the first research question, the dependent variable was accuracy (0, 1), and the fixed-effects were group (dummy-coded: Rules [task-essential practice with rule explanation], Right/Wrong [task-essential practice with right/wrong feedback], NF [task-essential practice without feedback], NTE [Non Task-essential practice]); tense (categorical variable coded as present or subjunctive); and time. To make growth as a piecewise function, the variable time was split into two different variables that were in complementary distribution, namely, time A (pre-to immediate posttests) and time B (immediate post- to delayed posttests). To address development from pre- to delayed posttests, a new model was built. As suggested in the literature (e.g., Barr et al. 2013; Linck & Cunning, 2015), I adopted a full maximum likelihood estimation process (Laplace Approximation). A full maximum random effect structure, however, can lead to non-convergence (i.e., the model does not produce a stable result), frequently due to an excessively complex model (Barr et al. 2013; Bates et al., 2015; Linck & Cunning, 2015). In such cases, a number of steps can be taken, including simplification of the random effects specification (Bates et al., 2015). Thus, random slopes of the model were tested one by one with likelihood ratio tests and only random slopes that significantly improved the model fit were retained.

The random variation that could arise from the two different versions of the assessment tests was taken into account. Version was considered a level 2 variable and it was itself nested in item, since there were two versions of the same item. Subject was also a level 2 variable, making subject and version within items crossed variables at the same level of sampling. Having subject and version as crossed random effects, mixed-effects models account for the variance of the subjects at the same time as the variance of the items. Thus, the random effects structure of the

---

3 More explanation on page 105.
model included cross-classified subject, version and item random intercepts, and random slopes for Time A (pre- to posttests), Time B (post- to delayed posttests), and tense.

As stated earlier, the model included two categorical variables (group and tense). In mixed-effects models one categorical factor can be used as a reference group against which all other levels of that factor are compared. For easiness of interpretation and to investigate different interactions, the factor of baseline can change, but this does not affect the goodness of the model. Although there is currently a controversy on how to assess significance with linear models (Baayen et al., 2008), that is not the case for logistic multilevel models that do produce usable $p$ values (significance at $p < .05$; marginal significance at $p < .10^4$).

**Results from the interpretation test.** Descriptive statistics with the mean accuracy from the interpretation test is presented in Table 9 to help visualize the effects presented in the model. On average, participants in both feedback groups were more accurate from pre- to immediate posttests, while the NF group did not seem to improve. Descriptively at a glance, Right/Wrong seemed more accurate than Rules at immediate posttests and at delayed posttests, and the NF group seemed less accurate at immediate posttests than at pretests. From immediate post- to delayed posttests, both feedback groups exhibited a loss in accuracy, but still a better performance than at pretests. The NF group performed better at delayed posttests as compared to their performance at immediate posttests.

---

4 The R software used to conduct the multilevel models analyses automatically sets the marginal level to $p < .10$. 

98
Table 9

*Mean accuracy (SDs) in the Interpretation test with subjunctive items*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules</td>
<td>.24 (.43)</td>
<td>.76 (.43)</td>
<td>.48 (.50)</td>
</tr>
<tr>
<td>Right/Wrong</td>
<td>.13 (.33)</td>
<td>.88 (.33)</td>
<td>.52 (.50)</td>
</tr>
<tr>
<td>No Feedback</td>
<td>.21 (.40)</td>
<td>.14 (.35)</td>
<td>.27 (.44)</td>
</tr>
</tbody>
</table>

*Note: Highest score possible = 1. Chance = .25*

Most of these patterns were confirmed by the model analysis. Piecewise logistic multilevel model was used for the accuracy data from the interpretation test (Table 10). First, the Rules group and subjunctive tense were baseline, thus all results were interpreted with respect to this baseline (e.g., a significant effect for Right/Wrong indicates this group is significantly different from the Rules group). For interpretation purposes, the baseline was changed at times to identify within-subject and between-subject comparisons.
Table 10

*Results of logistic multilevel model modeling Accuracy on the Interpretation test*

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>$b$</th>
<th>$\exp(b)$</th>
<th>$SE$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (Rules/subjunctive)</td>
<td>-1.73</td>
<td>-1.17</td>
<td>.46</td>
<td>.001*</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/W</td>
<td>-.90</td>
<td>.40</td>
<td>.66</td>
<td>.17</td>
</tr>
<tr>
<td>NF</td>
<td>-.39</td>
<td>.67</td>
<td>.68</td>
<td>.56</td>
</tr>
<tr>
<td>Pre- to posttest (Rules/subjunctive):</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre- to posttest x R/W</td>
<td>1.80</td>
<td>6.05</td>
<td>.75</td>
<td>.01*</td>
</tr>
<tr>
<td>Pre- to posttest x NF</td>
<td>-4.26</td>
<td>.01</td>
<td>.78</td>
<td>.001*</td>
</tr>
<tr>
<td>Post- to delayed posttest (Rules/subjunctive)</td>
<td>-1.84</td>
<td>.16</td>
<td>.46</td>
<td>.001*</td>
</tr>
<tr>
<td>Post- to delayed posttest x R/W</td>
<td>- .90</td>
<td>.40</td>
<td>.68</td>
<td>.18</td>
</tr>
<tr>
<td>Post- to delayed posttest x NF</td>
<td>3.05</td>
<td>21.11</td>
<td>.71</td>
<td>.001*</td>
</tr>
<tr>
<td>Present</td>
<td>3.58</td>
<td>35.87</td>
<td>.59</td>
<td>.001*</td>
</tr>
<tr>
<td>Present x R/W</td>
<td>2.00</td>
<td>7.39</td>
<td>.84</td>
<td>.01*</td>
</tr>
<tr>
<td>Present x NF</td>
<td>1.84</td>
<td>6.29</td>
<td>.88</td>
<td>.03*</td>
</tr>
<tr>
<td>Present x Pre- to posttest</td>
<td>-2.09</td>
<td>.12</td>
<td>.38</td>
<td>.001*</td>
</tr>
<tr>
<td>Present x Pre- to posttest x R/W</td>
<td>-3.57</td>
<td>.03</td>
<td>.57</td>
<td>.001*</td>
</tr>
<tr>
<td>Present x Pre- to posttest x NF</td>
<td>2.75</td>
<td>15.64</td>
<td>.63</td>
<td>.001*</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest</td>
<td>1.13</td>
<td>3.09</td>
<td>.37</td>
<td>.002*</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest x R/W</td>
<td>2.70</td>
<td>14.88</td>
<td>.59</td>
<td>.001*</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest x NF</td>
<td>-2.77</td>
<td>.06</td>
<td>.59</td>
<td>.001*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>$SD$</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>subject</td>
<td>3.30</td>
<td>1.81</td>
</tr>
<tr>
<td>Pre- to posttest</td>
<td>subject</td>
<td>3.67</td>
<td>1.91 -.52</td>
</tr>
<tr>
<td>Post- to delayed posttest</td>
<td>subject</td>
<td>3.08</td>
<td>1.75 .00 -.30</td>
</tr>
<tr>
<td>Present</td>
<td>subject</td>
<td>5.13</td>
<td>2.26 -.76 .22 .08</td>
</tr>
<tr>
<td>Intercept</td>
<td>version: item</td>
<td>.16</td>
<td>.40</td>
</tr>
<tr>
<td>Intercept</td>
<td>item</td>
<td>.05</td>
<td>.23</td>
</tr>
</tbody>
</table>

*Note: *Significant at $p<.05$; ^Marginal at $p<0.1$. Covariates are shaded in gray

The model intercept indicated that at pretests the Rules group was less likely than not to be correct with subjunctive items ($b = -1.73$, $SE = .46$, $p = .001$). Upon refitting the model with Right/Wrong as baseline and then NF as baseline to determine their performance at pretest, the predictions of the model indicated that both groups were less likely than not to be correct with
subjunctive items (Right/Wrong: $b = -2.63$, $SE = .48$, $p < .001$; NF: $b = -2.13$, $SE = .52$, $p < .001$), meaning they all performed statistically similarly before starting the treatment. This was confirmed by the comparisons between Rules and Right/Wrong ($b = - .90$, $SE = .66$, $p = .17$), Rules and NF ($b = - .39$, $SE = .68$, $p = .56$), and Right/Wrong and NF ($b = .50$, $SE = .70$, $p = .47$) that predicted no differences in the performance of all the groups at pretest.

Upon examining the data for the within-subject predictions from pre- to immediate posttests, the model predicted that the Rules group was more likely to answer correctly at immediate posttests than at pretests ($b = 3.48$, $SE = .50$, $p = .001$), and so was the Right/Wrong group ($b = 5.28$, $SE = .55$, $p < .001$), but that was not the case for the NF group, for which the model predicted no difference in their performance from pretest ($b = -.78$, $SE = .59$, $p = .18$). The between-subject comparisons indicated that the Right/Wrong group were about six times more likely to be accurate than the Rules group ($b = 1.80$, $SE = .75$, $p = .01$), and the NF group was significantly less accurate than the Rules group ($b = -4.26$, $SE = .78$, $p = .001$) (see Figure 18). In order to directly compare the predictions for the Right/Wrong group and the NF group, Right/Wrong and subjunctive were used as baseline, and the model showed a significant positive effect for Right/Wrong as compared to NF ($b = -6.07$, $SE = .81$, $p = .001$) from pre- to immediate posttests. Thus, the descriptive results presented in Table 9 seem to accurately capture the model’s predictions.
Figure 18. Results from the interpretation test

From immediate post- to delayed posttests, the model predicted a significant negative slope for the Rules group ($b = -1.84, SE = .46, p = .001$) and the Right/Wrong group ($b = -2.75, SE = .50, p < .001$) meaning that they both experienced a significant loss. For the NF group the model predicted a significant positive slope ($b =1.20, SE = .53, p = .02$), suggesting an improvement.

The between-group comparisons from the immediate post- to delayed posttests data between the NF group and the two feedback groups indicated that the slope for the NF was positive and significantly different from that of the Rules group ($b =3.05, SE = .71, p = .001$), and the Right/Wrong group ($b =1.21, SE =.54, p = .025$), whose slopes declined. Note, however, that the NF participants were performing slightly above chance at posttests, while both feedback groups were well above chance. The comparisons between the two feedback groups revealed that the slope for the Rules group was not significantly different from the slope of the Right/Wrong group ($b = - .90, SE = .68, p = .18$), suggesting that the significant difference they had at
immediate posttests was maintained at delayed posttests. While that was what the model predicted, this maintained difference between Rules and Right/Wrong did not seem to be captured in the graphical representation of the model (Figure 18). Upon looking at the model graphically and upon further inspections of the estimates in the model, it appeared that there was a suggested difference between groups at pretests and no difference at delayed posttests (not as the model predicted). Thus, it could be that perhaps there was not enough power in the model to observe a significant effect from immediate post- to delayed posttests that would suggest that the groups were not different at delayed posttests.

Post hoc analyses did not shed more light into this. To examine the overall change across time, a new logistic multilevel model was built where the time variable included data from pre- to delayed posttests. The predictions of the model revealed that both feedback groups improved significantly from pre- to delayed posttests (Rules, $b = 1.80$, $SE = .59$, $p = .002$; R/W, $b = 2.65$, $SE = .62$, $p < .001$), while the NF group did not show significant improvement ($b = .49$, $SE = .65$, $p = .44$). The between group comparisons showed some conflicting results. On one hand the model predicted no differences in the slopes of the Rules group and the Right/Wrong group ($b = .85$, $SE = .86$, $p = .32$), suggesting that they performed similarly at time of delayed posttests. On the other hand, the model predicted a significantly different slope for the Right/Wrong group when compared to the slope of the NF group ($b = 2.15$, $SE = .90$, $p = .01$), but no significant differences in the slopes of the Rules group and the NF group ($b = 1.30$, $SE = .88$, $p = .14$), indicating that these last two groups performed similarly and suggesting that the Rules group and the Right/Wrong group were different from each other from pre- to delayed posttests. It needs to be mentioned that in the process of model building and model testing for the post hoc analyses of the pre- to delayed posttests data, the information from posttests was not included in the final
model since it did not improve the goodness of fit of the model. This loss of information can lead to loss of power.

With the goal to examine whether participants had overgeneralized the subjunctive in contexts where the indicative mood is required, the present tense data was analyzed as well. Overgeneralization of the subjunctive would entail that the model would predict significant negative results with the present tense items. For easiness of interpretation, Table 11 shows the descriptive statistics with the average accuracy from the interpretation test with present tense items.

Table 11

*Mean accuracy (SDs) for the interpretation data with present tense items*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules</td>
<td>.78 (.41)</td>
<td>.86 (.34)</td>
<td>.80 (.40)</td>
</tr>
<tr>
<td>R/W</td>
<td>.91 (.29)</td>
<td>.84 (.36)</td>
<td>.87 (.33)</td>
</tr>
<tr>
<td>No Feedback</td>
<td>.91 (.27)</td>
<td>.88 (.31)</td>
<td>.87 (.33)</td>
</tr>
</tbody>
</table>

*Note: Higher score possible: 1. Chance = .25*

Upon refitting the model with present tense and the Rules group as baseline, it was found that at pretests all the groups were more likely than not to be correct with present tense items (Rules: $b =1.85$, $SE = .39$, $p < .000$; R/W: $b =2.95$, $SE = .43$, $p < .000$; NF: $b =3.30$, $SE = .47$, $p < .000$). A marginal effect for the Right/Wrong group ($b =1.10$, $SE = .56$, $p = .05$) and a significant effect for the NF group ($b =1.44$, $SE = .59$, $p = .01$) were found, meaning that both of these groups were more likely to answer correctly than the Rules group. Upon refitting the model with the Right/Wrong group as baseline to compare their performance with that of the NF group, no
significant differences were predicted by the model between these two groups at pretest ($b = .34, SE = .62, p = .58$).

From pre- to immediate posttests, the advantage exhibited by the Right/Wrong and the NF groups when compared to the Rules group at pretests was not observed by immediate posttests, as the slopes for Right/Wrong and NF were not significantly different from 0 (R/W, $b = -.37, SE = .56, p = .50$; NF, $b = -.12, SE = .62, p = .84$), while a positive effect was predicted for the Rules group ($b = 1.38, SE = .54, p = .01$), suggesting that all three groups performed similarly at immediate posttests (see Figure 16). The group comparisons confirmed this and showed (a) a significant negative slope for Right/Wrong as compared to the slope of Rules ($b = -1.76, SE = .77, p = .02$); and (b) a marginal negative slope for NF as compared to the slope of Rules ($b = -1.51, SE = .82, p = .06$), meaning that the Right/Wrong group and the NF group experienced a decline while the Rules group improved. With regard to the NF and Right/Wrong group, the model predicted no significant differences between these two groups ($b = .25, SE = .83, p = .76$). All in all, the results suggest that the three groups performed statistically similarly at immediate posttests.

From immediate post- to delayed posttests there was a marginal positive effect for the Right/Wrong group ($b = 1.08, SE = .55, p = .05$) but not for the NF group ($b = -.43, SE = .57, p = .44$) or the Rules group ($b = -.71, SE = .52, p = .17$), suggesting that the performance of these latter groups at delayed posttests was the same as at immediate posttests. The predictions for the between-group comparisons confirmed this pattern, and showed that (a) the slope for the Right/Wrong group was significantly better than that of the Rules group ($b = 1.79, SE = .75, p = .01$), (b) the Rules group and the NF group were not different from each other ($b = .27, SE = .76,$)
\( p = .71 \), and (c) the slope for the NF group was significantly worse than that of the Right/Wrong group \( (b = -1.51, SE = .79, p = .05) \).

To determine any significant improvements or declines in participants’ performance with present tense items across time, a new logistic multilevel model was built where the time variable only included data from pre- to delayed posttests. The model predicted no significant effects for the Rules group from pre- to delayed posttests \( (b = .84, SE = .63, p = .18) \) or for the NF group \( (b = .49, SE = .65, p = .44) \), but a marginal effect for the Right/Wrong group \( (b = 1.2, SE = .68, p = .07) \), suggesting a marginal probability for this group to be more correct at delayed posttests than at pretests. The results from the present data showed that the three groups performed similarly with no significant changes across time, suggesting that overgeneralization did not occur in a systematic way (see Figure 16).

**Summary.** Table 12 presents a summary of the results of the interpretation test. Altogether these results suggest that the feedback groups improved significantly across time with subjunctive items, while the NF group did not. From pre- to immediate posttests, a significant positive effect was predicted for both feedback groups, while no significant change was predicted for the NF group. The within-subjects comparisons showed that both feedback groups were more likely to answer correctly at immediate posttests than at pretests, while the NF group was predicted to be as accurate at immediate posttests as at pretests. A significant difference between groups from pre- to immediate posttests was found, with the model predicting significantly better performance for the Right/Wrong group than the Rules group with subjunctive items, and a significantly less accurate performance of the NF group when compared to the Rules and the Right/Wrong groups. The predictions of the model with the data from immediate post- to delayed posttests indicated that (a) both feedback groups experienced a
significant loss, which could be partially related to the long extension of time (i.e., four weeks) between the two sets of assessment tests, and (b) the NF group experienced a significant improvement; note, however, their low performance (Figure 16). The significant difference between feedback groups found at immediate posttests was maintained throughout delayed posttests. Such maintained difference did not seem to be captured in the graphical representation of the model (Figure 16). Upon further inspections of the estimates in the model, it appeared that there was a suggested difference between feedback groups at pretest (possibly not detected due to a lack of power in the model), and maybe not at delayed posttests, which is what the model predicted. Unfortunately, post hoc analyses could not determine if this was the case, since contradictory evidence pointing at both groups being different and being similar at delayed posttests was found. Finally, the present tense data showed that no overgeneralization occurred, which strengthens the subjunctive results.

Table 12

Summary of the results of the interpretation test

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre to posttest</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>RW&gt;R, R, RW&gt; NF</td>
</tr>
<tr>
<td>Post to delayed</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>RW&gt;R, NF&gt;RW, R</td>
</tr>
<tr>
<td>Pre to delayed</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>RW =R, RW&gt; NF, NF =R</td>
</tr>
</tbody>
</table>

Note. A= significant positive effect. B= significant negative effect; C= no significant effect
**Results from the written production test.** The data from the production test was coded following a lenient coding criterion and a strict coding criterion. I conducted the analyses with both sets of data and the results were the same, therefore here I am only reporting the data with the lenient coding.

As with the data from the interpretation test, a Piecewise logistic multilevel model was used for the data from the production test. This model had the same structure as the previous model: the dependent variable was accuracy (0, 1), and the fixed-effects were group, tense, and time. As earlier, the variable time was split into time A (pre- to immediate posttests) and time B (post- to delayed tests), in complementary distribution. To illustrate the results from the model a simplified characterization of the data as cell means is presented in Table 13.

Table 13

*Mean accuracy (SDs) in the Production test with subjunctive items*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules</td>
<td>.07 (.26)</td>
<td>.65 (.47)</td>
<td>.32 (.46)</td>
</tr>
<tr>
<td>Right/Wrong</td>
<td>.13 (.33)</td>
<td>.67 (.46)</td>
<td>.39 (.48)</td>
</tr>
<tr>
<td>No Feedback</td>
<td>.08 (.28)</td>
<td>.11 (.31)</td>
<td>.11 (.32)</td>
</tr>
</tbody>
</table>

*Note: Higher score possible: 1 Chance = .25*

Descriptively, both feedback groups showed high improvements and the NF group showed some improvement by the time of immediate posttests. At immediate posttests the Right/Wrong group seemed to perform better than the Rules group. Like with the interpretation test data, the Rules group and the Right/Wrong group seemed to experience a loss in accuracy by
the delayed posttest, but their performance was still better than at pretests. The NF group, however, demonstrated the exact same performance at the delayed posttest as at the immediate posttest.

Most of these patterns were confirmed by the Piecewise logistic multilevel model analysis (Table 14). Comparisons were made with the Rules group and subjunctive tense as baseline. The model intercept indicated that at pretest the Rules group was more likely than not to be incorrect ($b = -3.19, SE = .39, p = .001$) with subjunctive items, and so were the Right/Wrong group ($b = -2.40, SE = .37, p < .001$) and the NF group ($b = -3.14, SE = .43, p < .001$). The between-subject comparisons indicated the Right/Wrong ($b = .79, SE = 1.08, p = .13$) and the NF ($b = -.18, SE = .83, p = .74$) groups did not differ from the Rules group at pretest, and they were not different from each other ($b = .33, SE = .66, p = .61$). These results indicate that none of the groups had any substantial previous knowledge of the targeted structure before the treatment.
Table 14

Results of logistic multilevel model modeling Accuracy in the Production test.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>$b$</th>
<th>$\exp(b)$</th>
<th>$SE$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (Rules/subjunctive)</td>
<td>-3.19</td>
<td>.04</td>
<td>.39</td>
<td>.001</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right/Wrong</td>
<td>.79</td>
<td>1.08</td>
<td>.53</td>
<td>.13</td>
</tr>
<tr>
<td>NF</td>
<td>-.18</td>
<td>.83</td>
<td>.57</td>
<td>.74</td>
</tr>
<tr>
<td>Pre- to posttest (Rules/subjunctive):</td>
<td>4.12</td>
<td>61.56</td>
<td>.41</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Pre- to posttest x Right/Wrong</td>
<td>-.58</td>
<td>.56</td>
<td>.56</td>
<td>.29</td>
</tr>
<tr>
<td>Pre- to posttest x NF</td>
<td>-3.44</td>
<td>.03</td>
<td>.62</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Post- to delayed posttest (Rules/subjunctive)</td>
<td>-2.06</td>
<td>.12</td>
<td>.29</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Post- to delayed posttest x Right/Wrong</td>
<td>.23</td>
<td>1.26</td>
<td>.41</td>
<td>.57</td>
</tr>
<tr>
<td>Post- to delayed posttest x NF</td>
<td>1.71</td>
<td>5.53</td>
<td>.49</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Present</td>
<td>6.23</td>
<td>507.75</td>
<td>.62</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Present x Right/Wrong</td>
<td>.006</td>
<td>1.00</td>
<td>.89</td>
<td>.99</td>
</tr>
<tr>
<td>Present x NF</td>
<td>1.31</td>
<td>3.70</td>
<td>.94</td>
<td>.16</td>
</tr>
<tr>
<td>Present x Pre- to posttest</td>
<td>-5.18</td>
<td>.005</td>
<td>.39</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Present x Pre- to posttest x Right/Wrong</td>
<td>-1.83</td>
<td>.16</td>
<td>.55</td>
<td>.001</td>
</tr>
<tr>
<td>Present x Pre- to posttest x NF</td>
<td>3.18</td>
<td>24.04</td>
<td>.62</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest</td>
<td>1.95</td>
<td>7.03</td>
<td>.30</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest x Right/Wrong</td>
<td>1.47</td>
<td>4.35</td>
<td>.42</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest x NF</td>
<td>-1.56</td>
<td>1.08</td>
<td>.52</td>
<td>.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>$SD$</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>subject</td>
<td>1.75</td>
<td>1.32</td>
</tr>
<tr>
<td>Pre- to posttest</td>
<td>subject</td>
<td>1.61</td>
<td>1.27</td>
</tr>
<tr>
<td>Post- to delayed posttest</td>
<td>subject</td>
<td>.80</td>
<td>.89</td>
</tr>
<tr>
<td>Present</td>
<td>subject</td>
<td>5.28</td>
<td>2.30</td>
</tr>
<tr>
<td>Intercept</td>
<td>version: item</td>
<td>.07</td>
<td>.28</td>
</tr>
<tr>
<td>Intercept</td>
<td>item</td>
<td>.06</td>
<td>.24</td>
</tr>
</tbody>
</table>

Note: *Significant at $p < .05$; ^Marginal at $p < 0.1$. Covariates are shaded in gray

Figure 19 shows the graphical predictions of the model. From pre- to immediate posttests the model predicted the Rules group ($b = 4.12$, $SE = .41$, $p < .001$) and the Right/Wrong group ($b = 3.53$, $SE = .39$, $p < .001$) were significantly more accurate than at pretests, while the slope for the NF group did not significantly change from pre- to immediate posttests ($b = .66$, $SE = .47$, $p = .16$). No significant differences between feedback groups were predicted from pre- to
immediate posttests \( (b = -0.58, SE = 0.56, p = .29) \), but Rules was expected to be about eight times more accurate than NF \( (b = -3.44, SE = 0.62, p < .001) \). Upon refitting the model with the Right/Wrong group as baseline to compare it with the NF group, a significant positive effect was found for the Right/Wrong group from pre- to immediate posttests \( (b = 2.16, SE = 0.62, p < .001) \), suggesting the better performance of the Right/Wrong group over the NF group. These results suggest that feedback groups experienced development, while the NF group did not.

**Figure 19.** Results from the written production test with subjunctive and present tense items

The slopes from immediate to delayed posttests for the Rules group \( (b = -2.06, SE = 0.29, p < .001) \) and the Right/Wrong group \( (b = -1.82, SE = 0.30, p < .001) \) significantly changed in a negative way, suggesting a significant loss for these groups with respect to their performance at immediate posttests. No significant change was predicted for the NF group \( (b = -0.33, SE = 0.40, p = .40) \), indicating their performance at delayed posttests was similar than at immediate posttests.
The predictions for the group comparisons showed that the loss experienced by the Rules group was not significantly different from that of the Right/Wrong group ($b = .23$, $SE = .41$, $p = .57$), meaning that both feedback groups were not different at delayed posttests, while a significant positive effect was found for the NF group when compared to the Rules group ($b = 1.71$, $SE = .49$, $p = .000$), possibly due to their slope not being negative while the one for the Rules group was. After refitting the model with NF as baseline to compare this group with Right/Wrong, the model’s predictions indicated a significant positive effect for the NF group ($b = 1.47$, $SE = .50$, $p = .003$), which, like with the Rules group, may be due to the negative slope of the Right/Wrong group when compared to the almost flat slope of the NF group. These results in conjunction with the visual representation of the model in Figure 17 suggest that at delayed posttests both feedback groups performed equally and that the NF group was significantly less accurate than those groups, with no signs of learning.

In post hoc analysis a new model was built to investigate development across time from pre- to delayed posttests. The predictions of the new model indicated that both feedback groups improved significantly from pre- to delayed posttests (Rules: $b = 2.52$, $SE = .44$, $p < .000$; R/W: $b = 2.10$, $SE = .42$, $p < .000$), and for the NF group no improvement was predicted ($b = .53$, $SE = .49$, $p = .27$). No significant differences between Rules and Right/Wrong were predicted ($b = -.42$, $SE = .59$, $p = .47$), while the NF group was predicted to be significantly less accurate than the Rules group (-1.98, $SE = .64$, $p = .002$) and the Right/Wrong group (=-1.56, $SE = .63$, $p = .01$). Therefore, in spite of the loss experienced from immediate to delayed posttests, the feedback groups appeared to have demonstrated significant grammatical development, while the NF group did not.
As with the interpretation test, the data from present tense items was analyzed to see if overgeneralization had occurred. To help illustrate the results of the model, Table 15 shows descriptive statistics with the mean accuracy.

Table 15

*Mean accuracy (SDs) from the production test with present tense items*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules</td>
<td>.90 (.29)</td>
<td>.80 (.39)</td>
<td>.78 (.41)</td>
</tr>
<tr>
<td>Right/Wrong</td>
<td>.96 (.20)</td>
<td>.56 (.49)</td>
<td>.81 (.40)</td>
</tr>
<tr>
<td>No Feedback</td>
<td>.96 (.19)</td>
<td>.90 (.29)</td>
<td>.88 (.32)</td>
</tr>
</tbody>
</table>

*Note:* Higher score possible: 1

At a glance, the Right/Wrong group and the NF group seemed to display a better performance than the Rules group at pretests. On average, the Rules group decreased in their accuracy at immediate posttests and at delayed posttests. This decrease was more acute with the Right/Wrong group at immediate posttests, but at the time of delayed posttests both groups improved. The NF group gradually decreased in their performance from pre- to immediate posttests and from immediate to delayed posttests. All groups seemed to perform worse at delayed posttests than at pretests.

Most of these patterns were confirmed by the model’s results. With Rules and present tense items as baseline, the model predicted that the Rules ($b = 3.03, SE = .40, p < .000$) and the Right/Wrong ($b = 3.83, SE = .46, p < .000$) groups were more likely than not to be accurate with present tense items. Unexpectedly, when the model was refit to have NF as baseline in order to directly test the predictions of the model for this group in regard to their performance at pretests,
immediate posttests, and delayed posttests, the model did not converge so no reliable within-participant results with respect to the NF group’s performance can be reported. The predictions for the group comparisons indicated that there were no significant differences between the feedback groups at pretests ($b = -.79, SE = .59, p = .17$), but a marginal difference between Rules and NF ($b = 1.13, SE = .62, p = .07$). Interestingly, with Right/Wrong as baseline, no differences were predicted between NF and Right/Wrong ($b = .33, SE = .66, p = .61$)$^5$, maybe suggesting a lack of power for the model to predict a significant difference between Right/Wrong and Rules.

As represented in Figure 19 the model predicted a significant loss from pre- to immediate posttests for the Right/Wrong group ($b = -3.48, SE = .45, p < .000$), and for the Rules group ($b = -1.06, SE = .40, p = .008$). As shown graphically, a strong significant disadvantage was predicted for Right/Wrong over Rules ($b = 2.42, SE = .59, p < .000$) and over NF ($b = 2.16, SE = .68, p = .001$), and no significant differences were predicted between the NF group and the Rules group ($b = .25, SE = .65, p = .69$). These results suggest some confusion in the Rules group and a clear overgeneralization with the Right/Wrong group at the time of immediate posttests. By the time of delayed posttests, the model predicted an improvement for Right/Wrong ($b = 1.59, SE = .29, p < .000$) but not for Rules, whose slope did not change from immediate posttests ($b = -.11, SE = .30, p = .72$). The predictions for the between-participants comparisons showed a significant negative effect for the Rules group ($-1.70, SE = .42, p < .000$) and for the NF group ($-1.55, SE = .49, p = .001$) when compared to the Right/Wrong group, suggesting an improvement by the Right/Wrong group that is graphically captured in Figure 17.

$^5$ A prediction for a marginal difference between Rules and NF, and not significant difference between Yes/No and NF, seems contradictory when considering that no significant differences were predicted between Rules and Yes/No. However, the lack of differences between feedback groups could be due to a lack of power.
Finally, to determine whether there was overproduction of the subjective overall, a post hoc analysis was carried out with a new model that only included data from pre- to delayed posttests. The results indicated, unlike the interpretation data, a significant negative prediction for all groups (Rules: $b = -1.16$, $SE = .43$, $p < .006$; R/W: $b = -1.83$, $SE = .48$, $p < .000$; NF: $b = -1.06$, $SE = .52$, $p < .04$), suggesting that they all performed better at pretests than at delayed posttests.

**Summary.** Table 16 presents a summary of the results of the production test. Like the interpretation data, participants in the feedback groups significantly improved across time with subjunctive items. From pre- to immediate posttests the model predicted a significant effect for the Rules and the Right/Wrong groups, but no significant change for the NF group. The group comparisons indicated that (1) both feedback groups performed statistically similar to each other, (2) Rules was expected to be about eight times more accurate than NF, and (3) Right/Wrong was also statistically more accurate than NF. These results suggest that feedback groups experienced development from pre- to immediate posttests, while the NF group did not. From immediate post- to delayed posttests both feedback groups experienced a significant loss, while the NF group did not change significantly. The group comparisons showed that (1) the loss experienced by the Rules group was not significantly different from that of the Right/Wrong group, meaning that both feedback groups were not different at delayed posttests; (2) the NF group and the Rules group were significantly different from each other, and (3) the NF group and the Right/Wrong group were also significantly different from each other. These results, graphically represented in Figure 19, suggest that at delayed posttests both feedback groups that experienced a loss in accuracy performed equally, and that the NF group was significantly less accurate than those groups, with no signs of learning. The results from the post hoc analysis indicated that both
feedback groups improved significantly from pre- to delayed posttests, while no improvement was predicted for the NF group. The group comparison showed no significant differences between feedback groups, and the NF group was significantly less accurate than the Rules and the Right/Wrong groups.

In an attempt to determine if overgeneralization had occurred, the present tense data was analyzed. The predictions of the model with the pre- to immediate posttests data showed a significant loss for both feedback groups. The loss for the Right/Wrong group was significantly different from that of the Rules group and from that of the NF group, as shown in Figure 17, while no significant differences were predicted between the NF and the Rules groups. These results suggest that overgeneralization did occur and it was especially acute with the Right/Wrong group. By the time of delayed posttests, this was overcome by the Right/Wrong group, but not by the Rules group, whose performance did not significantly change. The between-participants comparisons captured this improvement of the Right/Wrong group, that performed significantly better than the Rules group and the NF group, as these latter groups did not significantly change their performance from immediate posttests. Finally, the post hoc analysis with data from pre- to delayed posttests showed that all groups performed better at pretests than at delayed posttests with present tense items, suggesting that the introduction of the new use of subjunctive in temporal clause did confuse them.
Table 16

Summary of the results of the production test

<table>
<thead>
<tr>
<th>Groups</th>
<th>Rules [R]</th>
<th>Right/Wrong [RW]</th>
<th>NF</th>
<th>Group comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre to posttest</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>R=RW</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R, RW&gt; NF</td>
</tr>
<tr>
<td>Post to delayed</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>R=RW</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R, RW&gt; NF</td>
</tr>
<tr>
<td>Pre to delayed</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>R=RW</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R, RW&gt; NF</td>
</tr>
</tbody>
</table>

Note. A= significant positive effect. B= significant negative effect; C= no significant effect

Research question #2

2. A) Is there a minimum threshold regarding the amount of feedback from which learners start to attain accuracy in a more consistent way?

2. B) If so, does the threshold differ depending on the type of feedback?

2. C) Is amount of feedback significantly related to development?

For the model of the second research question, the analysis was conducted with the lme4 package version 1.1-9 (Bates, Maechler, Bolker, & Walker, 2015) in R version 3.2.0 (R Core Team, 2015). Upon a visual examination of the data (that showed the shape of a parabola), a quadratic growth curve model was used to examine the pattern participants showed in regard to the amount of feedback they required during the time course of the treatment. Model testing was

---

6 A parabola is the graph of a quadratic function. It may open upward or downward and may vary in width and steepness.
done with full maximum likelihood and the final model was reported using restricted maximum likelihood estimation. In this model amount of feedback was a dependent continuous variable (range 1-4). Fixed-effects were block centered as a linear term; block centered squared as a quadratic term; and tense (present or subjunctive). The random effects structure of the model included a subject random intercept, and random slopes by subject for the linear and quadratic terms for block. Variation coming from items was not expected since each numbered block had the same items and never items from other blocks. Therefore, item was not included as a term in the model. For interpretation purposes, and in order to investigate participants’ learning process at the middle of the treatment, the model was centered on block 5. The model produced t-values and due to the current debate in calculating p-values in linear multilevel models, only t-values are provided (Baayen et al., 2008). Gelman and Hill (2007) recommend using t-values above 2.0 as being statistically significant and 1.65 marginally significant.

**Amount of feedback during treatment.** With subjunctive as the baseline, the model revealed that there were no differences between groups in the amount of feedback predicted ($\chi^2 = .13, df = 1, p = .72$), so only one line per tense will be shown in the graphs. The means amount of feedback presented in Table 17 are to illustrate the predictions of the model.

---

7 Although the treatment included 3 options, I found that some participants used a maximum of 4 feedback episodes.
Table 17

Mean amount of feedback received during treatment

<table>
<thead>
<tr>
<th>Amount of feedback</th>
<th>Subjunctive</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td>Per trial</td>
</tr>
<tr>
<td>Rules</td>
<td>42.26 (6.88)</td>
<td>1.45</td>
</tr>
<tr>
<td>Right/Wrong</td>
<td>40.50 (7.34)</td>
<td>1.39</td>
</tr>
</tbody>
</table>

*Note: All participants received positive feedback, so the minimum amount is equal to the number of trials (Subjunctive = 29; Present = 12). The maximum amount expected is 87 for subjunctive and 36 for present.*

Results indicated that block linear and block quadratic terms were negative and significant, suggesting the inverse U shape of the model shown in Figure 20.

*Figure 20. Amount of feedback predicted for each tense by block*
In the x axis of Figure 20 blocks are shown. As explained in the materials section, the treatment consisted of blocks with subjunctive items and blocks with present tense items. Each numbered block always had the same items, and the items within each block were randomized. The presentation of the blocks was fixed, such that a maximum of two of the same tense blocks could be in sequence. While block is informative because each block represents that participants had more input from the other form in the previous block, it does not capture the information regarding the number of items included in each block. Figure 21 shows trial numbers, so that I can examine where the peak of the curve (i.e., a possible threshold) is happening in the time course of the practice.

![Graph showing trajectory of feedback predicted for each tense by trial](image)

*Figure 21. Trajectory of the feedback predicted for each tense by trial*

The intercept in Table 18 indicates that subjunctive items were more likely than not to require feedback. Figure 21 graphically shows the main effect for tense \( (b = -.14, SE = .03, t = -4.13) \), meaning that overall participants required less feedback on present tense items than on subjunctive items at block 5. A significant interaction for block and tense would extrapolate this
positive effect that occurred at block 5 to the entire time course of the practice. A marginal positive interaction for block (linear term only) and tense was found, possibly suggesting that participants needed different amounts of feedback for present tense items and subjunctive items during the entire treatment. In an effort to further investigate this marginal effect, I refitted the model at block 8 and no significant differences between tenses were predicted. Taken together with the results from block 5, this suggests that the amount of feedback for present tense and subjunctive were indeed different throughout the experiment. Put another way, this interaction between block and tense seems to indicate a different shape of the curve such that the peak in the amount of feedback with present tense items occurred at a different point (later) in the course of the practice. Moreover, it also suggests that participants required the same amount of feedback with present tense items at the end and at the beginning of the experiment (as can be seen in Figure 21), whereas for subjunctive items participants at the beginning of the experiment appeared to require much more feedback than they did at the end of the experiment. The fact that the effect is only marginal can be explained by the restricted range (1-4) of the dependent variable. While the maximum amount of feedback per trial was expected to be 3 episodes (since there were three options), some participants used 4 feedback episodes.
Table 18

*Results of quadratic multilevel model modeling Amount of feedback*

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>b</th>
<th>SE</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (Subjunctive)</td>
<td>1.47</td>
<td>.50</td>
<td>29.58 *</td>
</tr>
<tr>
<td>Tense</td>
<td>-.14</td>
<td>.03</td>
<td>-4.13*</td>
</tr>
<tr>
<td>Block linear</td>
<td>-.02</td>
<td>.00</td>
<td>-2.97*</td>
</tr>
<tr>
<td>Block linear x tense</td>
<td></td>
<td></td>
<td>1.78^</td>
</tr>
<tr>
<td>Block quadratic</td>
<td>-.005</td>
<td>.00</td>
<td>-2.04*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>SD</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.06</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Block linear</td>
<td>.00</td>
<td>.02</td>
<td>-1.00 -.05</td>
</tr>
<tr>
<td>Block quadratic</td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

*Note: *Significant at $t > 2.00$; ^Marginal at $t > 1.65$

In order to investigate a possible threshold for the amount feedback and when in the practice it happened, the model was graphically examined to determine the peak points and the cliff points of the curves. Peak points indicated the maximum average amount of feedback predicted by the model at the time of maximal need of feedback, while the cliff points indicated the average amount of feedback when a constant decrease of the need of feedback occurred. Table 18 shows the model’s predictions of the needed feedback. I will first explain the predictions with subjunctive items followed by the predictions with present tense items. These explanations are to be read simultaneously with Table 19 and Figure 21.
Table 19

Predictions of the needed feedback

<table>
<thead>
<tr>
<th>Block</th>
<th>Practice trial in treatment (% of total practice)</th>
<th>Subjunctive or Present practice trial (% of specific practice)</th>
<th>Predicted average amount for a trial</th>
<th>Predicted accumulative amount (% of total specific feedback)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjunctive items—Peak 3</td>
<td>13 (31.70%)</td>
<td>11 (37.93%)</td>
<td>1.497</td>
<td>16.33 (18.77%) †</td>
</tr>
<tr>
<td>Subjunctive items—Cliff 5</td>
<td>22 (53.65%)</td>
<td>16 (55.17%)</td>
<td>1.475</td>
<td>23.76 (27.31%) †</td>
</tr>
<tr>
<td>Present tense items—Peak 5</td>
<td>24 (58.53%)</td>
<td>8 (66.66%)</td>
<td>1.332</td>
<td>10.32 (28.66%) ^</td>
</tr>
<tr>
<td>Present tense items—Cliff 7</td>
<td>25 (60.97%)</td>
<td>10 (83.33%)</td>
<td>1.308</td>
<td>12.95 (35.97%) ^</td>
</tr>
</tbody>
</table>

Note: † Indicates % where the maximum number of potential feedback with subjunctive items is 87. ^ Indicates % where the maximum number of potential feedback with present tense items is 36.

By block 3 of practicing with subjunctive items and after having completed 13 trials of practice (the peak), 11 of which were exclusively with subjunctive items, the model predicted that a learner would need a maximum average amount of 1.497 feedback episodes, and an accumulative amount of 16.33 feedback episodes, including all the trials in each block. After that, a constant decrease in the need of feedback would occur by block 5 at subjunctive practice trial 16 (the cliff) with an average amount of 1.475 feedback episodes and an accumulative amount of 23.76. This seems to suggest that after having been practicing for 11 trials with subjunctive items and receiving an accumulative amount of about 16.33 feedback episodes (approximately 19% of the total possible feedback with subjunctive at an average rate of 1.497 feedback episodes per trial), participants started to show a more sustained performance, and after
an accumulative amount of 23.76 feedback episodes in 16 trials (approximately 27% of the total possible feedback with subjunctive at an average rate of 1.475 feedback episodes per trial) their performance was increasingly more accurate as evidenced by the low amount of feedback needed toward the end of the treatment.

To obtain a clear picture of the learning process, data from the present tense items were also examined. With present tense items, the model predicted that the peak of the curve would occur in block 5 (after receiving 8 present tense practice trials) with a maximum average amount of 1.332 feedback episodes, and an accumulative amount of 10.32. After that, a constant decrease in the need of feedback would occur by block 7 at present tense practice trial 10 (the cliff), with an average amount of 1.308 feedback episodes and an accumulative amount of 12.95. This indicates that with present tense items participants started to show a better performance after practicing for 8 trials and receiving an accumulative amount of about 10.32 feedback episodes, approximately 29% of the total possible feedback with present tense items (at an average rate of 1.332 feedback episodes per trial), and that after 10 present tense practice trials (with an accumulative amount of about 12.95, approx. 36% of the total possible feedback episodes with present tense), the need of feedback acutely decreased. The better performance was evidenced by the constant reduced amount of feedback from the cliff point in block 7 until the end of the practice.

**Amount of feedback and development.** In order to investigate whether there was any relationship between the amount of feedback and language development, a series of correlations were run between the amount of feedback data and the accuracy data from the interpretation and the production tests. To see any relationship with the interpretation data, Pearson’s correlations
were run between accuracy data at posttests and amount of feedback with subjunctive items.

Pearson’s correlation coefficients are presented in Table 20.

Table 20

*Results of correlations between feedback and assessment tests*

<table>
<thead>
<tr>
<th></th>
<th>At immediate posttests</th>
<th>At delayed posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rules</td>
<td>Right/Wrong</td>
</tr>
<tr>
<td></td>
<td>( r ) ( p )</td>
<td>( r ) ( p )</td>
</tr>
<tr>
<td>Interpretation:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjunctive items</td>
<td>-.37 &lt; .000*</td>
<td>-.41 &lt; .000*</td>
</tr>
<tr>
<td>Present tense items</td>
<td>-.32 &lt; .000*</td>
<td>-.29 &lt; .000*</td>
</tr>
<tr>
<td>Subjunctive items</td>
<td>.52 &lt; .00*</td>
<td>-.31 &lt; .000*</td>
</tr>
<tr>
<td>Production:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjunctive items</td>
<td>-.07 .15</td>
<td>-.02 .70</td>
</tr>
<tr>
<td>Present tense items</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As Table 20 shows, amount of feedback for subjunctive and present tense items and performance in the interpretation test were negatively correlated for both groups at immediate posttests and delayed posttests, suggesting that the higher the amount of feedback participants received, the lower their accuracy. The same pattern was found with the production data and amount of feedback for subjunctive items also at posttests and delayed posttests. However, results were different for the amount of feedback for present tense items and performance on the production test. While the performance of both groups did not seem to be related to the amount of feedback that they received at immediate posttests, this changed for the Right/Wrong group at
delayed posttests when their performance showed to be negatively correlated to the amount of feedback they received ($r = -0.15, p = 0.005$).

**Summary of the amount of feedback results.** A summary of the results is presented in Table 21. These results indicate that the amount of feedback each group demanded was the same regardless of the type of feedback, either rules or just correct/incorrect. As can be expected, the amount of feedback needed for the subjunctive items was higher than for the present tense items. The trajectory that both groups had during the treatment was that of an inverse U, indicating a higher demand at the beginning of the experiment that progressively decreased as the experiment developed. However, the trajectory of amount of feedback for subjunctive items was different than that of the present tense items, such that the peak in the amount of feedback with subjunctive items occurred at a much earlier time point than that of the present tense items, probably due to the novelty of being exposed to a new targeted structure (i.e., temporal sentences with subjunctive). With regard to a possible threshold in the amount of feedback and practice, results suggest that in a practice with a total of 29 trials with subjunctive, participants started to perform increasingly more accurately after 16 trials, and an accumulative amount of 23.76 feedback episodes (approx. 27% of the total possible feedback with subjunctive). With respect to development and amount of feedback, results indicated that there is a negative relationship, such that the more feedback participants received with subjunctive items, the less accurate they were in both assessment tests.
Table 21  

Summary of the results of the Research Question 2

- Rules and Right/Wrong required the same amount of feedback
- More feedback was needed with subjunctive items.
- With subjunctive, higher demands at the beginning that progressively decreased.
- With subjunctive, the peak occurred earlier in the practice
- With present tense, the peak occurred later in the practice
- After receiving about 27% of the feedback, both groups started to be increasingly more accurate.
- More feedback related to less development.

Research question #3

3. A. Does type of input-based practice (NF vs. NTE) have a differential effect on L2 grammar development, as measured by interpretation and production tests, immediately after the treatment?

3. B. Is this effect maintained after 4 weeks?

For this question, the data from the NF group and the NTE group were compared. I remind the reader that in these two groups participants practiced with input-based practice and no feedback. The main difference between the groups was the nature of the input-based practice. While one carried out task-essential practice, the other carried out a practice that did not direct learners’ attention to the targeted structure.

Results of the interpretation test. Table 22 presents the descriptive statistics with the mean accuracy from the interpretation test. At first glimpse, a striking difference was the means for each group at pretests, indicating that NF was more accurate than NTE before starting the treatment. By the time of the immediate posttests, NF declined in their performance while NTE improved, surpassing the NF group. By delayed posttest NF performed better than they did at
immediate posttests and even better than NTE, whose performance was about the same as in the immediate posttests. However, both groups were always performing below chance at all times, except for the NF group that performed slightly above chance by delayed posttests.

Table 22

*Mean accuracy (SDs) in the interpretation test with subjunctive items*

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback</td>
<td>.21 (.40)</td>
<td>.14 (.35)</td>
<td>.27 (.44)</td>
</tr>
<tr>
<td>NTE</td>
<td>.05 (.23)</td>
<td>.18 (.38)</td>
<td>.17 (.37)</td>
</tr>
</tbody>
</table>

*Note: Highest score possible: 1*

To investigate these patterns statistically, a Piecewise model with the interpretation data from these two groups was built, similar to the previous ones, where the dependent variable was accuracy (0, 1), and the fixed-effects were group (NF [No Feedback] or NTE [Non-Taskessential]); tense (present or subjunctive); and time (pre- to immediate posttests & immediate post- to delayed posttests). The model was leveled with NF and subjunctive as baseline, so comparisons are made against this baseline.

At pretests both groups were more likely than not to be incorrect with subjunctive items (NF: \( b = -2.16, SE = .54, p < .000 \); NTE: \( b = -3.79, SE = .55, p < .000 \)), which indicates that they did not know the targeted structure before the treatment. The apparent difference between groups at pretests as shown in the descriptive statistics of Table 22 was confirmed by the model analysis. The results of that analysis (see Table 23) showed a significant effect for the NF group, indicating a disadvantage for the NTE group (\( b = -1.63, SE = .75, p = .03 \)). As the graphic representation of the model shows in Figure 20, from pre- to immediate posttests the difference
disappeared: the slope for NF did not change: ($b = -.77$, $SE = .53$, $p = .15$), while the slope for NTE did ($b = 1.14$, $SE = .53$, $p = .03$), thus indicating an improvement. The fact that the slope of one group changed and the slope of the other did not was captured by the model, predicting the slope of the NTE to be significantly different from the one of the NF group ($b = 1.91$, $SE = .74$, $p = .01$). The reversed patterns were found from immediate to delayed posttests, where the model predicted an improvement for the NF group ($b = 1.11$, $SE = .43$, $p = .009$) and no change for the NTE ($b = -.34$, $SE = .45$, $p = .44$). The predictions for group comparisons revealed that the NF group was significantly better than the NTE group at the time of delayed posttests.

Table 23

Results of logistic multilevel model modeling Accuracy in the Interpretation test

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>$b$</th>
<th>$\exp(b)$</th>
<th>$SE$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (NF/subjunctive)</td>
<td>-2.16</td>
<td>.11</td>
<td>.54</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTE</td>
<td>-1.63</td>
<td>.19</td>
<td>.75</td>
<td>.03*</td>
</tr>
<tr>
<td>Pre- to posttest (NF/subjunctive):</td>
<td>- .77</td>
<td>.00</td>
<td>.53</td>
<td>.15</td>
</tr>
<tr>
<td>Pre- to posttest x NTE</td>
<td>1.91</td>
<td>6.75</td>
<td>.74</td>
<td>.01*</td>
</tr>
<tr>
<td>Post- to delayed posttest (NF/subjunctive)</td>
<td>1.11</td>
<td>3.03</td>
<td>.43</td>
<td>.009*</td>
</tr>
<tr>
<td>Post- to delayed posttest x NTE</td>
<td>-1.46</td>
<td>.23</td>
<td>.61</td>
<td>.01*</td>
</tr>
<tr>
<td>Present</td>
<td>5.50</td>
<td>244.69</td>
<td>.72</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Present x NTE</td>
<td>1.37</td>
<td>3.93</td>
<td>.97</td>
<td>.15</td>
</tr>
<tr>
<td>Present x Pre- to posttest</td>
<td>.54</td>
<td>1.71</td>
<td>.48</td>
<td>.25</td>
</tr>
<tr>
<td>Present x Pre- to posttest x NTE</td>
<td>-1.58</td>
<td>.20</td>
<td>.66</td>
<td>.01*</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest</td>
<td>-1.53</td>
<td>.21</td>
<td>.44</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest x NTE</td>
<td>2.82</td>
<td>16.77</td>
<td>.66</td>
<td>&lt;.001*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>subject</td>
<td>3.74</td>
<td>1.93</td>
<td></td>
</tr>
<tr>
<td>Pre- to posttest</td>
<td>subject</td>
<td>2.87</td>
<td>1.69</td>
<td>-.30</td>
</tr>
<tr>
<td>Post- to delayed posttests</td>
<td>subject</td>
<td>1.54</td>
<td>1.24</td>
<td>.13 -.33</td>
</tr>
<tr>
<td>Present</td>
<td>subject</td>
<td>5.88</td>
<td>2.42</td>
<td>-.81 -.02 -.22</td>
</tr>
<tr>
<td>Intercept</td>
<td>version: item</td>
<td>.31</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>item</td>
<td>.00</td>
<td>.00</td>
<td></td>
</tr>
</tbody>
</table>

*Note: *Significant at $p<.05$; ^Marginal at $p<0.1$. Covariates are shaded in gray
As was done with the first research question, a new model with only the data from the pretests and the delayed posttests was built to examine overall development. The results of the post hoc analysis confirmed that no learning had occurred from pre- to delayed posttests for any of the groups (NF: $b = .41, SE = .53, p = .44$; NTE: $b = .81, SE = .59, p = .17$), and no differences between the two groups were predicted either ($b = .40, SE = .78, p = .61$).

In light of these results and upon examination of the predictions of the model with present tense items graphically represented in Figure 2, it did not seem necessary to analyze the present tense data to examine overproduction of the subjunctive.

**Summary of the results of the interpretation test.** Table 24 presents a summary of the results. Altogether the results of the interpretation test indicate that the treatment of the current experiment did not have a strong effect in these two groups. At pretests, there was a significant difference between groups, with an advantage for the NF group that disappeared by the time of immediate posttests. From pre- to immediate posttests no significant effect for the NF group was
predicted while a significant positive effect was predicted by the NTE group. The group comparison indicated a significant positive effect for the NTE group when compared to the NF group. Interestingly, the reverse pattern was found from immediate post- to delayed posttests: the NTE group did not significantly change while the NF group did. A positive effect was found for the NF group when compared to the NTE group. The results of the post hoc analysis with the data from pre- to delayed posttests showed that neither group significantly changed across time nor were they significantly different from each other. All in all, these results seem to indicate that no significant learning occurred in these two groups.

Table 24

*Summary of the results of the interpretation test*

<table>
<thead>
<tr>
<th>Groups</th>
<th>Group comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NF</td>
</tr>
<tr>
<td>Pre to posttest</td>
<td>C</td>
</tr>
<tr>
<td>Post to delayed</td>
<td>A</td>
</tr>
<tr>
<td>Pre to delayed</td>
<td>C</td>
</tr>
</tbody>
</table>

*Note:* A = significant positive effect; C = no significant effect

**Results of the production test.** The production test results were very similar to the results of the interpretation test. Table 25 presents the descriptive statistics with the mean accuracy of the production test, Table 26 shows the predictions of the model, and Figure 21 shows the graphical representation of the predictions by the model. As shown in Table 25 and Figure 21, both groups had quite low and similar scores across time. At pretests the model
predicted that both groups were more likely than not to be incorrect with subjunctive items (NF: \( b = -3.79, SE = .65, p < .000; \) NTE: \( b = -4.00, SE = .64, p < .000 \)), and that they were not different from each other (\( b = -.20, SE = .86, p = .81 \)). From pre- to immediate posttests the slope for the NF group significantly changed in a positive way (\( b = 1.21, SE = .57, p = .03 \)), while the slope for the NTE remained the same (\( b = .61, SE = .59, p = .30 \)), however, no significant differences between the slopes of each group were predicted (\( b = -0.60, SE = .77, p = .43 \)). From immediate post- to delayed posttests no change was predicted for any of the groups (NF: \( b = -0.61, SE = .38, p = .11 \); NTE: \( b = .45, SE = .39, p = .24 \)), but the slope of the NTE group was significantly different (\( b = 1.06, SE = .49, p = .03 \)) from that of the NF group in a positive way.

The post hoc analysis with the predictions from the new model with the data from pre- to delayed posttests to examine change overall revealed a significant effect for the NF group (\( b = 1.31, SE = .54, p = .01 \)) but not for the NTE group (\( b = .29, SE = .96, p = .76 \)). These results indicate that the slope for the NF group was significantly different from 0, suggesting an overall improvement, while the slope for the NTE was not. The predictions for the group comparisons, however, showed no significant differences between groups (\( b = -1.02, SE = .99, p = .30 \)).

As with the interpretation data, considering these results and the predictions of the model with present tense items represented in Figure 22, it was not deemed necessary to analyze the present tense data to examine overproduction of the subjunctive.

Table 25

Mean accuracy (SDs) in the Production test with subjunctive items

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Immediate Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Feedback</td>
<td>.08 (.28)</td>
<td>.11 (.29)</td>
<td>.11 (.30)</td>
</tr>
<tr>
<td>NTE</td>
<td>.07 (.25)</td>
<td>.06 (.23)</td>
<td>.12 (.32)</td>
</tr>
</tbody>
</table>

Note: Higher score possible: 1
Figure 23. Results from the production test
Table 26

Results of logistic multilevel model modeling Accuracy in the Production test

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>$b$</th>
<th>$\exp(b)$</th>
<th>$SE$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (NF/subjunctive)</td>
<td>-3.79</td>
<td>.02</td>
<td>.65</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td><strong>Group:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTE</td>
<td>- .20</td>
<td>.82</td>
<td>.86</td>
<td>.81</td>
</tr>
<tr>
<td>Pre- to posttest (NF/subjunctive):</td>
<td>1.21</td>
<td>3.35</td>
<td>.57</td>
<td>.03*</td>
</tr>
<tr>
<td>Pre- to posttest x NTE</td>
<td>-.60</td>
<td>.55</td>
<td>.77</td>
<td>.43</td>
</tr>
<tr>
<td>Post- to delayed posttest (NF/subjunctive)</td>
<td>-.61</td>
<td>.54</td>
<td>.38</td>
<td>.11</td>
</tr>
<tr>
<td>Post- to delayed posttest x NTE</td>
<td>1.06</td>
<td>2.88</td>
<td>.49</td>
<td>.03*</td>
</tr>
<tr>
<td>Present</td>
<td>7.95</td>
<td>2835.57</td>
<td>.85</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td>Present x NTE</td>
<td>.00</td>
<td>1</td>
<td>1.12</td>
<td>.99</td>
</tr>
<tr>
<td>Present x Pre- to posttest</td>
<td>-2.38</td>
<td>.09</td>
<td>.55</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td>Present x Pre- to posttest x NTE</td>
<td>3.32</td>
<td>27.66</td>
<td>.66</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest</td>
<td>.29</td>
<td>1.33</td>
<td>.44</td>
<td>.50</td>
</tr>
<tr>
<td>Present x Post- to delayed posttest x NTE</td>
<td>-3.56</td>
<td>.03</td>
<td>.73</td>
<td>&lt; .001*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Variance</th>
<th>$SD$</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.36</td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>Pre- to posttest</td>
<td>2.27</td>
<td>1.50</td>
<td>-.84</td>
</tr>
<tr>
<td>Post- to delayed p.</td>
<td>.36</td>
<td>.60</td>
<td>-.99 -.87</td>
</tr>
<tr>
<td>Present</td>
<td>7.03</td>
<td>2.65</td>
<td>-.83 .42 -.76</td>
</tr>
<tr>
<td>Intercept</td>
<td>.22</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.00</td>
<td>.09</td>
<td></td>
</tr>
</tbody>
</table>

Note: *Significant at $p < .05$; ^Marginal at $p < 0.1$. Covariates are shaded in gray

Summary of the results of the production test. Table 27 presents a summary of the results. As with the interpretation data, these results graphically represented in Figure 21 suggest that input-based practice (be it TE or not) and no feedback did not have a strong effect on the production of temporal sentences. From pre- to immediate posttests, the model predicted a significant positive effect for the NF group, no effect for the NTE group, and no significant differences between groups. The predictions of the model from immediate post- to delayed posttests revealed no effect for either group, but a significant positive effect for the NTE group when compared to the NF group. The new post hoc model with only the data from pre- to
delayed posttests showed a significant effect for NF, suggesting an overall improvement for this group, but no effect for the NTE group. However, the group comparisons revealed no significant differences between groups, meaning that both groups performed overall similarly.

Table 27

*Summary of the results of the production test*

<table>
<thead>
<tr>
<th></th>
<th>Groups</th>
<th>Group comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NF</td>
<td>NTE</td>
</tr>
<tr>
<td>Pre to posttest</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Post to delayed</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Pre to delayed</td>
<td>A</td>
<td>C</td>
</tr>
</tbody>
</table>

Note. A= significant positive effect; C= no significant effect
CHAPTER 5: DISCUSSION AND CONCLUSIONS

Discussion

The present dissertation aimed to investigate the effects of type of feedback (i.e., “right/wrong” vs. rule explanation) during input-based practice (task-essential vs. non task-essential). It was of special interest to track the amount of feedback, in order to investigate the trajectory of the feedback during the treatment, specifically whether there was a threshold from which participants started to require fewer feedback episodes to achieve accuracy, and whether different amounts of feedback led to different development. In addition, while task-essential (TE) practice has been identified as a desirable and optimal feature when designing a task (Loschky & Bley-Vroman, 1993), and has been included in several studies (e.g., Cerezo, 2010; Hsieh, 2007; Lado, Bowden, Stafford, & Sanz, 2013; Moreno, 2007; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004; Stafford, Bowden, & Sanz, 2012), the isolated contribution of TE practice has not yet been identified.

In the previous chapter, I presented the results of the analyses used to answer the research questions. In this chapter, I discuss empirical findings concerning each research question. I then draw some conclusions based on the findings and describe the implications derived from this research. Finally, I consider limitations of the study and identify paths for future research.

Research question 1

Does type of feedback (Rules vs. Right/Wrong vs. No Feedback?) have a differential effect on L2 grammar development, as measured by interpretation and production tests, immediately after the treatment? Is this effect maintained after 4 weeks?
Consistent with previous investigations (Lado et al., 2013; Sanz & Morgan-Short, 2004; Stafford et al., 2012), the findings of the current study suggest that task-essential practice coupled with feedback promotes language development at the interpretation level and is sufficient to transfer skills from input-based practice to the ability to produce the targeted structure. Specifically, the results pertaining to the first research question indicated that providing “right/wrong” feedback was more beneficial than providing grammar explanation when interpreting temporal sentences right after treatment, and that this edge was maintained for four weeks. On the other hand, as has been seen in previous studies (e.g., Bowles, 2008, Lado et al., 2013; Rosa & Leow, 2004; Stafford et al., 2012), when transferring skills from input-based practice to production of the targeted structure, participants with metalinguistic information in the form of feedback showed an immediate advantage over those who were only informed of the accuracy of their responses. This was evidenced by the results of the immediate production posttest with the subjunctive items in conjunction with the present tense items. While both groups performed similarly with subjunctive items, with present tense items the group that received “right/wrong” feedback overproduced significantly more the subjunctive in contexts where present tense was required, significantly more than the group that received rule explanation. However, after four weeks participants who received rule explanation did not maintain their advantage and both groups showed evidence of similar performance in production of present tense and subjunctive items. Positive and even better results in retention from less explicit learning conditions are not new in SLA research. Indeed, there are empirical studies (e.g., Bowles, 2008; Hsieh, 2007; Lado et al., 2013; Moreno, 2007; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Stafford et al., 2012) and meta-analyses (e.g., Li, 2010; Mackey & Goo, 2007) that found that groups receiving less explicit type of feedback performed at delayed
posttests as well (e.g., Bowles, 2008; Hsieh, 2007), better (e.g., Cerezo et al., 2016; Moreno, 2007), or did not experience such a great loss as participants in more explicit learning conditions, showing sustained improvements (e.g., Cerezo et al., 2016; Lado et al., 2013, Stafford et al., 2012). At the same time, better performance in production from participants who were exposed to more explicit learning conditions has also been the finding of several studies examining the role of explicitness of the learning conditions (e.g., Bowles, 2008; Lado et al., 2013; Rosa & Leow, 2004; Stafford et al., 2012). Furthermore, while these studies included input-based practice in their treatments, studies that had participants completing output-based practice also found an advantage of more explicit learning conditions in the production tests (e.g., Cerezo, 2010; Nagata, 1993; Nagata & Swisher, 1995).

The findings of the current study, however, differed from those of Camblor (2006), Hsieh (2007), Moreno (2007), and Sanz and Morgan-Short (2004), none of which found differences between participants with different learning conditions. The proficiency level of the participants and the targeted structure of the current study may explain this difference.

As explained by Lado et al. (2013), level of proficiency has been identified as a factor affecting feedback effects (e.g. Iwashita, 2003; Mackey & Philp, 1998). These studies (Camblor, 2006; Hsieh, 2007; Moreno, 2007; and Sanz & Morgan-Short, 2004) included participants with basic proficiency levels, whereas the current study had participants with an intermediate level of Spanish. While a recent study by Morgan-Short et al., (2010) argues that naïve learners may have an initial advantage under explicit learning conditions, previous research shows better results with less explicit type of feedback (i.e., recasts) and more advanced learners (e.g., Ayoun, 2001; Mackey & Philp, 1998). Specifically, the results from Ayoun’s (2001) computerized study indicated that recast was more effective than grammar explanation with all groups and especially
with more advanced learners, who by the nature of their level of proficiency tend to be more experienced learners. It may be that relatively more experienced learners, such as the ones in this dissertation, benefit more from less explicit teaching conditions that leave them to their own devices, which, as argued by Stafford et al. (2012), may have helped them induce patterns and develop intuitions that were better captured in the interpretation test.

In regard to the targeted structures, those studies included targeted structures that involved a non canonical SVO word order (i.e., *gustar* expression and Spanish object clitics), which can be argued to be more salient and thus good candidate for either kind of feedback, since participants could arguably locate the source of the error more easily. The targeted structure in the present dissertation is less salient, and although grammatical explanation was definitely beneficial, the possible deeper processing toward which the participants in the Right/Wrong group were pushed may explain the differences between those studies and the current one.

In fact, depth of processing may explain differences between the group that only had “right/wrong” feedback and the group that also received rule explanation. The fact that participants in the Right/Wrong group were not provided with any linguistic information during the treatment may have pushed them to a deeper and more robust processing, which in turn led to higher performance in the test that had the same format as the treatment. It needs to be noted, however, that since no measures of processing were employed, it remains unclear whether the group receiving only negative evidence had a deeper processing, arguably data-driven processing as suggested by Rosa & O’Neill (1999), than that of the group with negative evidence and rule explanation, that conceivably had concept-driven processing. However, some evidence of this can be found in a study done by Hsieh, Moreno, and Leow (2015). This study, that revised Hsieh
(2008), compared the effects of computerized face-to-face (C-FTF) instruction against computer-assisted instruction (CAI), looking specifically into the cognitive processes employed while interacting with L2 data in both media. On one hand, participants in the C-FTF watched a video where an instructor provided grammatical explanation of the *gustar* expressions and their subcomponents (the object, the subject, and the verb). On the other hand, participants in the CAI group completed a computerized game that incorporated implicit feedback. Using think-aloud protocols, the authors investigated the levels of awareness and the depth of processing of participants while engaged with the L2 data. The researchers found larger amounts of verbal reports in CAI as compared to C-FTF. They also found that an average of 14.6% of the verbal reports generated in the CAI group was at the highest level of processing. Participants in the C-FTF condition did not qualify for the highest level of processing, “since they had received the grammatical rules for the *gustar* structure and, subsequently, they did not need to process the L2 data deeply to arrive at the underlying grammatical rules” (p.144). Interestingly, the results from immediate posttests showed no differences between groups, while gain scores from pre- to delayed posttests suggested an advantage for CAI. The researchers concluded that participants in the CAI group needed to put more effort processing the structure compared to the participants in the C-FTF group, as evidenced by the larger number of verbal reports of the former group, and the presence of the highest levels of processing. They argued that this increased effort might have led to more robust processing of the grammatical information and might have been so strong that it helped to retain it for several weeks after the experimental exposure. Finally, they also claimed that differential depth of processing could plausibly explain the lack of significant differences in performance between groups of the original study, and possibly in other studies such as Sanz and Morgan-Short (2004). Although the present study had no online data that could
show evidence of a more robust processing in the “right/wrong” feedback group, it is possible that this group behaved similarly to the CAI group in Hsieh et al. (2015) since they had similar characteristics (i.e., learner-initiated and task-essential practice, no explicit grammar explanation, and no grammatical rule in the feedback), thus potentially explaining the better performance of the participants in the “right/wrong” feedback group in this dissertation.

Another plausible explanation is a potential cognitive overload with the Rules group. This group received rule explanation that was customized depending on the time reference of the item. Such detailed explanation of the rules could have been taxing for participants that were reading this feedback every time they made a mistake. By taxing learners in this way, rules explanation translated into lower accuracy.

**Research questions 2**

A) Is there a minimum threshold regarding the amount of feedback from which learners start to attain accuracy in a more consistent way? If so, does the threshold differ depending on the type of feedback? B) Is amount of feedback significantly related to development?

Another goal of this dissertation was to investigate feedback during the learning process. To my knowledge, this is the first CALI study to specifically look at the amount of feedback learners require as they interact with the input\(^8\). In addition, it was of interest to investigate whether different types of feedback required different amounts of feedback, and whether different amounts of feedback led to different developmental results. Interestingly, no differences in the amount of feedback that each group required were found, meaning that regardless of whether the group received only “right/wrong” feedback or a combination of “right/wrong” feedback.

---

\(^8\) In a CALI study, Cerezo (2016) investigated the effects of different amounts of practice. By the nature of the design of his study, different amounts of “right/wrong feedback” were provided to the experimental groups.
feedback plus rule explanation, the learners requested the same amount of feedback during the treatment. This contrast with the results of Loewen and Nabei’s (2007) interactional study that found the recast group received many more feedback episodes (18) than the metalinguistic group (5.6) while completing the experimental tasks. The difference between the feedback groups, and the difference between the findings of the present study and their study may result from the nature of the feedback. Loewen and Nabei (2007) investigated the effects of recast, metalinguistic feedback, and clarification requests. It is known that recast frequently go unnoticed, which could explain the higher frequency of recasts as compared to the lower frequency of metalinguistic feedback in their study. It could also explain why their less explicit feedback group (i.e., the recast group) received more feedback episodes than the less explicit feedback group of the present study (i.e., the “right/wrong” feedback group). In an interactional study such as Loewen and Nabei’s (2007) it is not possible to know how many of the 18 recasts were noticed, even at a minimum level, and how many were completely unnoticed. Generally, feedback is provided as a response to non-target like utterances produced by L2 learners, with the goal that the learners will hopefully incorporate that feedback, will test their hypotheses of the language, and will restructure accordingly. In Loewen and Nabei’s study (2007) the fact that instructors provided more recasts could be an indication that that feedback was not being noticed\(^9\) and hence the need to keep providing more. In contrasts, the explicit nature of the metalinguistic feedback that is provided during interaction arguably made this feedback very noticeable and salient, and hence, the lower amount. This contrasts with the characteristics of the feedback of the present study. Due to its design, both types of feedback (“right/wrong” and

\(^9\) It could be an indication that the feedback was not noticed, not understood, or not processed as deeply. But, retrospectively, there is no clear way to identify which one was the reason, or if it was a combination of them.
grammatical rule) had to be noticed\(^\text{10}\) by the learners since it was explicitly shown to them in the screen of their computers, and unless a correct answer was given they were not able to advance to the next question. Thus, the different amounts of feedback in Loewen and Nabei might have been related to the nature of the feedback rather than other possible explanations such as different ways of incorporating or processing different types of feedback.

Participants in the present study needed to respond correctly in order to advance to the next question. Given that one group received the grammatical rule and the other did not, it could be expected that the group not receiving the rule could have needed more feedback episodes than the group that had the rule available to them. However, both groups required the same amount of feedback regardless of the type of feedback. A possible explanation may be the way learners processed the feedback. In fact, it is plausible that the participants in the Right/Wrong group might have processed more deeply than the participants in the Rules group, which would explain (1) why the participants in the Right/Wrong group did not require more feedback episodes even though the rule was not available to them; and (2) why their performance in the interpretation test was statistically better than the performance of the Rules group. As mentioned earlier, the results of previous studies have indicated that participants working in less explicit learning conditions showed more sustained improvement at retention (e.g., Lado et al., 2013; Stafford et al., 2012), and higher performance that was linked to deeper processing and higher levels of awareness at the level of understanding (e.g., Hsieh et al., 2015). As previously mentioned, another plausible explanation is a potential cognitive overload with the Rules group. This group was exposed to linguistic (maybe unfamiliar) explanations that could plausibly have taxed their cognitive resources. Therefore, having the rules available to them did not translate into fewer

\(^{10}\) Note that I argue that the feedback itself was noticed since it was very salient on the screen. I am not making any claims as to the information in the feedback being noticed or further processed.
mistakes and hence fewer feedback episodes. On the contrary, it translated into the same amount of mistakes and feedback episodes as the group that had no explanation available to them.

An interesting finding of the present study pertains to the correlational relationship between amount of feedback and L2 development, being the first study to investigate this. It was found that, regardless of type of feedback, the more feedback learners required the less accurate they were [and conversely, the less feedback learners required the more accurate they were]. This seems to suggest that those participants who were not making many mistakes during the treatment (hence the lesser amount of feedback needed) were “getting it” (i.e., showing signs of learning) and this was revealed later in the results of the posttests. Conversely, those participants who struggled more with the new structure during the treatment, as the practice developed they continued to make errors that led to more feedback episodes. These participants might have arrived to the correct form by process of elimination, trying all the options, which it would suggest that they were not really “getting it” nor understanding it. This lack of understanding became evident with their low performance in the posttests.

There are some possible explanations to account for this. First, it could possibly suggest that the participants who were not “getting it” and required more feedback processed the information less deeply than those who did not require as much feedback. If participants were processing more shallowly while practicing, they may have made more mistakes that led to more feedback episodes. Conversely, participants who were processing more deeply may have been more successful in the practice, leading to fewer feedback episodes. A second explanation may be related to a cognitive overload. Those participants who were making more mistakes during the practice may have had their cognitive capacities taxed by either reading and processing the rules (for those participants in the Rules group) or trying to figure out the underlying rule (for
those participants in the Right/Wrong group). A study of learners’ IDs such as language aptitude, working memory, or anxiety could help reveal why some participants did not have their cognitive capacities as taxed as others. Finally, the high or low level of involvement of the participants while doing the experiment may have played a role in their performance. Those who were less involved were less successful during the treatment which led them to need more feedback episodes. This lesser success in the practice was later observed in the assessment tests, hence, the negative correlation between amount of feedback and performance.

Another goal of the present study was to investigate whether different amounts of feedback led to different results in L2 development. My results indicated that the same amount of feedback received by both groups did lead to different developmental results, with an advantage for the “right/wrong” feedback group. Interestingly, some studies that directly or indirectly looked at amount of feedback found no differences between groups. Cerezo (2016) investigated the effects of different amount of practice that incorporated “right/wrong” feedback. In Ellis et al. (2008) the focused and unfocused groups received different amounts of direct feedback, which can be classified as more explicit type of feedback. In Loewen and Nabei (2007) investigating the effects of recast, metalinguistic feedback, and clarification request, each group received different amounts feedback. Whether receiving the same type of feedback in the

11 In an attempt to determine whether the differences between type and amount of practice in the right/wrong feedback groups from Sanz and Morgan-Short (2004) and Rosa and Leow (2004) could explain the conflicting results of these studies, Cerezo (2016) investigated (a) whether having 4 instead of 2 options as answers in the practice; and (b) whether completing 28 instead of 56 items had differential effects on L2 learning. His results indicated that all groups performed similarly regardless of the amount of practice items completed, or the amount of choices provided. It needs to be noted that Cerezo (2016) called for caution interpreting his results due to the low observed power. To obtain a finer-grained analysis, he computed the standardized mean difference effect sizes, and found that the greater amount of practice completed by the right/wrong feedback group in Sanz and Morgan-Short (2004) (similar to his 56-item group) might explain the better results obtained by this group when compared to the results of the right/wrong feedback group in Rosa and Leow (2004) (p. 115).
experimental groups like in Cerezo (2016) or in Ellis et al., (2008), or whether receiving different types of feedback like Loewen and Nabei (2007), all the experimental groups in these studies performed statistically similarly. The explanation behind may be related to depth of processing. As previously discussed, it may not be a matter of quantity of feedback but a matter of how that feedback is processed. As discussed earlier, the nature of the “right/wrong” feedback may have led the learners to a more robust processing that translated into better performance. Depth of processing data elicitation procedures such as think aloud protocols would allow investigating this. Indeed, as Leow (2015) notes, the depth of processing strand of research has found evidence, from concurrent and non-concurrent data, that greater depths of processing leading to higher levels of awareness appear to be correlated to higher levels of performance (e.g., Cerezo et al., 2016; Hsieh et al., 2015; Morgan-Short, Heil, Botero-Moriarty, & Ebert, 2012). Two of the findings of the present study (i.e., no difference in amount of feedback between the groups, and negative correlations between amount of feedback and performance) seem to be pointing to what Leow’s (2015) model of the L2 learning process in ISLA suggests, namely, that a key factor in the early stages of the learning process is, in addition to attention paid, how incoming L2 information is processed.

To examine the trajectory of the feedback during treatment and to better understand the learning process of the targeted structure, the feedback received with subjunctive items as well as present tense items were examined. For reader’s convenience, the trajectory and predictions of the needed feedback from Figure 21 and Table 19 are again presented below now in Figure 24 and Table 28.
Figure 24. Trajectory of the feedback predicted for each tense by trial

Table 28

Predictions of the needed feedback

<table>
<thead>
<tr>
<th>Block</th>
<th>Subjunctive items—Peak</th>
<th>Subjunctive items—Cliff</th>
<th>Present tense items—Peak</th>
<th>Present tense items—Cliff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Practice trial in treatment (% of total practice)</td>
<td>Subjunctive or Present trial (% of specific practice)</td>
<td>Predicted average amount for a trial</td>
<td>Predicted accumulative amount (% of total specific feedback)</td>
</tr>
<tr>
<td>3</td>
<td>13 (31.70%)</td>
<td>11 (37.93%)</td>
<td>1.497</td>
<td>16.33 (18.77%)</td>
</tr>
<tr>
<td>5</td>
<td>22 (53.65%)</td>
<td>16 (55.17%)</td>
<td>1.475</td>
<td>23.76 (27.31%)</td>
</tr>
<tr>
<td>5</td>
<td>24 (58.53%)</td>
<td>8 (66.66%)</td>
<td>1.332</td>
<td>10.32 (28.66%)</td>
</tr>
<tr>
<td>7</td>
<td>25 (60.97%)</td>
<td>10 (83.33%)</td>
<td>1.308</td>
<td>12.95 (35.97%)</td>
</tr>
</tbody>
</table>

Note: † Indicates % where the maximum number of potential feedback with subjunctive items is 87. ‡ Indicates % where the maximum number of potential feedback with present tense items is 36.
Results indicated that, overall, more feedback was required with subjunctive items than with present tense items, and that this happened at different time points during the time course of the treatment. Thus, the peak of amount of feedback with present tense items occurred later in the treatment, whereas the peak of amount of feedback with subjunctive items occurred quite early in the treatment. When participants were interacting with the subjunctive items, they required higher amounts at the beginning and those amounts progressively decreased over time. Specifically, the highest need of feedback occurred after having completed about 37.93% of the practice with subjunctive items (11 subjunctive practice trials). At that time, participants (regardless of feedback type) required about 18.77% of the total possible feedback episodes with subjunctive items (i.e., 16.33 feedback episodes). Increasingly accurate performance was observed after having completed 55.17% of the subjunctive practice (i.e., 16 subjunctive practice trials), and having received about 27.31% of the total possible feedback episodes with subjunctive (i.e., 23.76 feedback episodes). With present tense items, learners required the same amount at the beginning of the practice as at the end. The highest need of feedback occurred after having completed about 66.66% of the practice with present tense items (8 present tense practice trials). At that time, participants required about 28.66% of the total possible feedback with present tense items (10.32 feedback episodes). A decrease in the need of feedback was observed after completing 83.33% of the practice trials with present tense (10 present tense practice trials), and receiving about 35.97% of the total possible feedback episodes with present tense (12.95 feedback episodes). An interesting fact was observed after approximately 20 practice trials (see Figure 19). The highest need of feedback with present tense items (at trial 24) occurred at the same time as the decrease of need of feedback with subjunctive items (at trial 22), possibly indicating overgeneralization of subjunctive with present tense items.
A similar pattern where fewer mistakes are made as the practice advances is found in Cerezo (2016), who investigated whether having 4 instead of 2 options as answers in the practice affected learning. To this end, he examined where participants clicked during practice. He found that participants that had 4 options very rarely clicked on additional options and they progressively clicked fewer times on those options as the treatment evolved (p.114). The trajectories of the feedback with subjunctive and present tense items are coherent with what can be expected when a new structure is introduced to learners who did not have a grammatical lecture prior to practice. When the structure is first introduced, more information (whether “right/wrong” feedback or rule explanation) is required in an effort to discover/understand the underlying rule, and is necessary to complete the practice. As the practice advanced and this new structure seemed to be competing with an already known structure (i.e., present tense), more feedback was required with the known structure, possibly indicating that hypotheses formation and testing were taking place.

The answer to the question of whether there is a minimum threshold regarding the amount of feedback from which learners start to attain accuracy in a more consistent way seems to be a positive one. In the present study the minimum threshold differed depending on the tense of the item (present tense indicative or present tense subjunctive), and investigating both tenses together offered a better picture of what happened during the learning process. After completing about half of the practice with the new structure (i.e., temporal sentences with subjunctive that refer to the future) and after receiving less than 30% of the total feedback possible, increasingly accurate performance was observed. Hypothesis testing and restructuring of an already known structure (i.e., temporal sentences with present indicative that refer to the present) might have occurred as a result of the introduction of the new structure with which the already known
structure is in paradigmatic relation. While no online data were collected to show direct evidence that hypothesis testing and restructuring took place, the results of the assessment tests that learning happened in both feedback groups are testimony that these two processes must have occurred.

**Research question 3**

*Does type of input-based practice (NF vs. NTE) have a differential effect on L2 grammar development, as measured by interpretation and production tests, immediately after the treatment? Is this effect maintained after 4 weeks?*

The last goal of this dissertation was to determine the role of task-essential (TE) practice. Task-essentialness has been identified as the best way to incorporate grammar in language instruction from a focus on form perspective (Loschky & Bley-Vroman, 1993, p. 123). Both paper-and-pencil format studies (e.g. Leow, 1997a, 1997b, 1998a, 1998b; Rosa & O’Neill, 1999) and CALI studies (e.g., Bowles, 2008; Cerezo, 2010; Hsieh, 2007; Lado et al., 2013; Moreno, 2007; Rosa & Leow, 2004; Sanz & Morgan-Short, 2004; Sanz et al., 2014; Stafford et al., 2010; Stafford et al., 2012) have included TE practice in their designs with positive results, which had led some authors to attribute task-essentialness as playing an important role in L2 learning. However, only Moreno (2007) set out with the goal to investigate the isolated role of TE practice. While her results indicated that TE practice did not have a significant role in L2 development, some methodological limitations such as small sample sizes and conflation of task-essentialness and feedback render the results inconclusive. In the present study the data from the interpretation and production tests from the last research question do not show evidence of a strong advantage of TE practice, when compared to input-based practice that is not specifically
designed to draw learners’ attention to the relevant structure. This is exemplified in the TE versus non-TE comparisons that showed similar results of the interpretation and production tests from the two groups that had no feedback and that completed input-based practice (whether TE or not).

A study that had groups completing input-based practice without pedagogical feedback, just like the NF and the NTE groups of the current study, is Henshaw (2012). The Henshaw study, following the tenets of PI, had three groups receiving explicit information about the subjunctive in temporal clauses and completing some input-based practice. Participants were assigned to one of the following groups: (1) REF, with referential activities (that are TE); (2) AFF, with affective activities (non TE); and (3) RA, a combination of both (TE & non TE). All groups received some kind of feedback, but only the groups completing referential activities (REF and RA) were provided with pedagogical feedback on the relevant structure, while the AFF group received affective feedback and no information on the target form. Results on recognition and interpretation tests did not show group differences, thus indicating that type of practice did not have a differential effect on performance. Specifically relevant to my study are the results obtained with the AFF group that did not receive any pedagogical feedback on the form (only affective feedback) and still experienced learning. In the present study, the NTE group that also completed input-based practice without feedback did not obtain such positive results. One plausible explanation for this discrepancy may be related to the amount of exposure received in the AFF group. By the design of Henshaw’s study, participants in the AFF group received more instances of positive evidence since in the treatment they always had to choose between two grammatically correct options. On the contrary, in the current dissertation all groups, including the NTE group, received the same amount of exposure of the targeted
structure. The extra amount of exemplars the AFF group had could partially explain the better results of this group as compared to the NTE group.

The feedback versus no feedback comparisons of the present study seems to highlight that, although TE without feedback may be somewhat facilitative of L2 development based on the immediate results from the interpretation test, the role feedback plays is superior. While TE may draw learners’ attention to the target L2 data, it is the feedback that allows them to confirm or disconfirm their original hypotheses (Loschky & Bley-Vroman, 1993) and it is this process that allows them to internalize some robust L2 information (Leow, 2015).

Conclusions

The present dissertation aimed to investigate the effects of type of feedback (i.e., “right/wrong” vs. rule explanation) during input-based practice (task-essential vs. non task-essential), with a targeted structure that has received much attention in SLA research (cf. Benati, Lee, & McNulty, 2010; Henshaw, 2012; Kanwit & Geeslin, 2014; Lee & McNulty, 2013). Of special interest was to investigate the amount of feedback that was provided as learners were interacting with the L2 input in the treatment. The goal was to determine whether different types of feedback generated different amounts of feedback; whether the amount of feedback was related to development; and whether there was a threshold in the amount of feedback from which L2 learners started to require fewer feedback episodes to achieve accuracy. Finally, the acclaimed benefits of task-essential practice (conflated with feedback) were tested to determine their validity.

The research was carried out with 71 native-English speakers learning Spanish in a private Northeastern university in the United States. All participants were at the beginning of their fourth semester in Spanish and were randomly assigned to one of four groups that had
different characteristics regarding provision of feedback, type of feedback, and type of input-based practice. Learners completed a treatment on Spanish temporal adverbial clauses. Participants completed assessment tests prior to the treatment (30 min), immediately following the treatment, and again four weeks later. The amount of feedback from the treatment was tracked and analyzed. Multilevel models and correlational analyses were employed to analyze the data. The results and the pertinent conclusions are summarized below.

First, with regard to the efficiency of different types of feedback, results showed that providing “right/wrong” feedback was more beneficial than providing the grammatical rule when interpreting temporal sentences, and this edge was maintained for four weeks after treatment. Examining participants’ production of critical items (subjunctive) and their counterparts (present tense), results showed similar patterns in both groups with subjunctive items, but an initial advantage for the Rule group and a disadvantage for the Right/Wrong group with the present tense items at immediate posttests, possibly indicating overproduction of the subjunctive. However, the advantage of the Rules group was not maintained, the Right/Wrong group overcame its erroneous overproduction of subjunctive, and after four weeks both groups showed evidence of similar performance in production of present tense and subjunctive items.

Although this finding is consistent with previous research that showed evidence of better, more sustained, or lesser loss in performance with the less explicit learning condition (e.g., Bowles, 2008; Hsieh, 2007; Hsieh et al., 2015; Lado et al., 2013; Moreno, 2007; Stafford et al., 2012), the present study contributes to the debate of the effects of type of feedback that differ in the degree of the explicitness, and extends our knowledge in that it provides empirical evidence of the positive effects of merely providing learners with “right/wrong” feedback with a targeted structure that is under-researched. While providing grammatical rule has long been proven to
have positive effects, providing learners with minimal amounts of information of the L2 in the form of “right/wrong” feedback that leaves them to their own devices may be more conducive to greater cognitive efforts, which in turn may lead to more robust processing and subsequent internalization and retention of the relevant structure (Hsieh et al., 2015).

Second, and related to the previous finding, it was found that different types of feedback did not require different amounts of feedback, and that the same amount of feedback led to L2 development regardless of the type of feedback. Moreover, with the same amount of feedback, the Right/Wrong group obtained better results. This is interesting since the Rules group received grammatical explanation while the Right/Wrong group did not. These findings possibly suggest that the “right/wrong” feedback group had to process more deeply since they needed to arrive at the underlying grammatical rule by themselves, and this deeper processing was strong enough that the targeted structure was retained four weeks after treatment.

The third finding is related to L2 development and amount of feedback. The present study is the first to examine this relationship, something that SLA researchers (e.g., Mackey, Oliver, & Leeman, 2003) have called to investigate. More specifically, the findings of the current study contribute to answering two of the questions proposed by Mackey et al. (2003) regarding the relationship between amount of feedback and development, and the minimum level of feedback for development. My results indicated that more feedback did not lead to more development. On the contrary, it was found that feedback was negatively correlated to development. A priori this may sound counterintuitive, but if we consider how the feedback is processed rather than how much is provided, and that more feedback could be a result of shallow processing rather than deeper or better processing, this finding is not unreasonable. While it can be logical to suspect that the more feedback the better, the finding that the more feedback the less
development pointed to the possibility that quantity of feedback is not as important as *how* the information is processed, something that other researchers have recently pointed out (e.g., Hsieh et al., 2015; Leow, 2015; Leow & Mercer, 2015). This underscores the need for researchers to directly investigate *how* the information regarding the L2 (be it feedback, rule explanation, or mere L2 input) is processed online. Only by examining how learners interact and process the L2 will we be able to understand the real intricacies of instructed second language learning.

A fourth finding of the current study suggested that massive amounts of feedback are not necessary but that a minimum amount of feedback integrated into practice that allows for error correction should be provided. In the current study practicing with about half of the 29 critical items and receiving less than 30% of the total feedback possible (almost 24 feedback episodes) was enough to attain accuracy that progressively improved as the practice evolved. This underscores the importance of providing learners with opportunities to repair their errors. Some studies and pedagogical interventions provide feedback, but fail to provide prompts and opportunities to put into practice the hypotheses learners form while interacting with the L2 and while receiving feedback regarding their interactions, which is crucial for language development. The analysis of the amount of feedback that learners required as they interacted with the L2 provided a window into the learning process of the targeted structure. In this study the learning process had the shape of an inverted U (see Figure 18) that showed a gradual decline in the need for feedback, which can be interpreted as a gradual increase on accuracy. That is, the learners were increasingly choosing the correct answer as the treatment evolved.

Finally, the results of the assessment tests regarding the role of task-essentialness did not provide evidence of a strong advantage of TE practice. The comparisons between the two groups that did not received feedback but differed in the type of input-based practice (TE vs. Non TE)
revealed that both groups performed similarly. It remains unclear, and needs to be empirically addressed, whether completing input-based practice that is not specifically designed to draw learners’ attention to the relevant structure, but that does provide feedback, yields the same benefits as with TE practice and feedback. Comparing the performance of the two feedback task-essential groups versus the no feedback task-essential group leads me to conclude that although TE may be beneficial for L2 development, the role of feedback is superior. Previous research that attributed task-essentialness as playing an important role in L2 learning conflated TE with feedback (e.g., Leow, 1998a, Sanz & Morgan-Short, 2004). The findings of the current research revealed that while L2 development may be promoted by the combination of TE practice and feedback (e.g., as previously reported on Hsieh, 2007; Moreno, 2007), TE practice alone does not.

In light of the findings of this dissertation, and as Leow (2015) suggests, it seems advisable for language practitioners and researchers to provide L2 learners with TE practice that directs learners’ attention to the relevant structure and incorporates right/wrong feedback. The feedback will allow them to form, reinforce, and test their hypotheses about the language, and the nature of the feedback will arguably lead them to a deep processing and internalization of the targeted information. A third feature should be providing prompts and creating opportunities for error correction, like in the present dissertation, which will allow the learners to put into practice and test their hypotheses, and will encourage deeper processing while they interact with the L2 data (Cerezo et al., 2016; Leow, 2015).
Limitations and Future Research

The present study, as with all empirical studies, presents some limitations that need to be considered in order to accurately interpret and situate the findings. In this section, I briefly discuss its limitations and consider new paths for future investigations.

First, the empirical findings of previous studies such as Hsieh et al. (2015), and the results of the present study seem to point to depth of processing as a key factor to explicate the lack of differences in amount of feedback of both feedback groups, and the better results yielded by the “right/wrong” feedback group. However, this dissertation could not offer empirical data to directly address this. Future research should probe into whether different types of feedback are processed in different ways, and data elicitation procedures such as think aloud protocols would allow for investigating this. External and internal variables such as the degree of the difficulty of the targeted structures (external), learners’ working memory capacity and aptitude, or learning styles (internal) may mediate the way different types of feedback are processed. The ISLA research, and specifically the feedback strand of research, can benefit from investigations addressing the role of such variables when learners process feedback that differs in the degree of its explicitness, which will contribute to a better understanding about how feedback works and its role in L2 learning. The results of such investigations will also have important implications for language instructors.

Second, the predictions of the mixed-effects model for the interpretation test indicated that the slope for the Rules group was not different than the slope for the Right/Wrong group from immediate post- to delayed posttests, meaning that the significant difference observed between the groups from pretest to immediate posttests was maintained. However, upon further inspection of the estimates in the model, it appeared that there was a suggested difference
between feedback groups at pretests and maybe, unlike the predictions of the model, no difference at delayed posttests. This would mean that perhaps there was not enough power in the model to observe a significant effect from immediate post- to delayed posttests that would suggest that the groups were not different at delayed posttests. This possible lack of power in the interaction between group and time from immediate to delayed posttests was probably due to the number of participants. While the sample size of the present dissertation was not small, it was not very large either. A larger sample size would have allowed for more robust results in the interaction from immediate to delayed posttests that would have provided more generalizable findings.

Third, the better performance of the Rules group in production may be partially due to the nature of the controlled written production test. It has been pointed out before that production tests such as the one employed in this dissertation “may have favored the use of metalinguistic knowledge” (Lado et al., 2013, p. 19), making it more difficult to capture the performance of the group not receiving metalinguistic information derived from less explicit learning conditions. Future research should include assessment tools that do not force the learners to think in linguistic terms, with the goal to elicit a more accurate sample of their interlanguage.

Fourth, no reliable results regarding possible change from pre- to immediate posttests and delayed posttests of the NF group with the present tense items in production could be reported. When the model built for the production data was refit with both the present tense items and the NF group as baseline in order to determine this group’s change, the model did not converge and consequently no reliable results could be reported. As a reminder, the analysis of the present tense items was done in order to determine whether overproduction of the subjunctive occurred. In spite of not having reliable results, the visual inspection of the data in Figure 17 suggests that
participants in the NF group were more accurate at pretests than at immediate posttests and delayed posttests, possibly indicating that overproduction did occur.

Fifth, the design of the interpretation task in the treatment involved several options as answers from which participants had to choose. If they chose a wrong answer, they received feedback and had to try again until they obtained the correct answer. This was done so that participants could form, test, and put into practice their hypothesis of the language as they interacted with the L2 data, and to investigate how much feedback was needed for that purpose. Due to the limited range of options in the treatment (3 options) the participants could have arrived at the correct answer not by processing the data, but through a process of elimination. Ideally, technology-based studies should provide feedback that directly addresses each learner’s individual errors and provides not a limited amount of opportunities for error correction but unlimited opportunities. In spite of the apparent benefits of such designs, the technical difficulties in creating such tasks are not to be underestimated.

Sixth, the wording in the feedback that explained the rule and that included linguistic terms might have been quite technical, and participants might have had a difficult time understanding the explanation. Future research should make an effort to use plain language that is not linguistically charged and that explains grammatical rules in the simplest way.

Seventh, from the data gathered in the exit questionnaires, it became apparent that some participants were confused with the use of indicative present tense to express futurity, in spite of this issue being addressed in the review section prior to the experiment. Future research interested in studying futurity in temporal clauses should include future tense to avoid any potential confusion.
Finally, while it was not the goal of the present study to see the effects of feedback in input-based tasks that are not TE, the inclusion of a group with such features (i.e., non-TE practice and feedback) would have shed more light into the role of feedback and practice. As mentioned in the previous section, it needs to be empirically addressed whether input-based practice that is not specifically designed to draw learners’ attention to the relevant structure but that includes feedback, yields the same results as TE practice and feedback, potentially lending stronger weight to the role of feedback.
### Appendix A: Rules in the input

<table>
<thead>
<tr>
<th>Study</th>
<th>Prior Rule presentation</th>
<th>Type of feedback</th>
<th>Language &amp; Targeted forms</th>
<th>Groups-conditions</th>
<th>Type of Practice</th>
<th>Results</th>
</tr>
</thead>
</table>

1) "Intelligent" performed significantly better w/ complex forms. 2) No difference for the simple form.
<table>
<thead>
<tr>
<th>Study</th>
<th>Prior Rule presentation</th>
<th>Type of feedback</th>
<th>Language &amp; Targeted forms</th>
<th>Groups-conditions</th>
<th>Type of Practice</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>“Implicit”: error signaling</td>
<td>Contrary-to-fact conditionals</td>
<td>(2) EPIFE [+TE practice, +grammar, + implicit feedback]);</td>
<td></td>
<td>2) “Explicit” condition better than “implicit” condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(3) EFE [+TE practice, - grammar, + explicit feedback]);</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4) EPE [+TE practice, +grammar, - feedback]);</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(5) IFE [+TE practice, - grammar, + implicit feedback]);</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(6) Control: [-TE practice, - grammar, - feedback]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Prior Rule presentation</td>
<td>Type of feedback</td>
<td>Language &amp; Targeted forms</td>
<td>Groups-conditions</td>
<td>Type of Practice</td>
<td>Results</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Sanz & Morgan-Short (2004)   | Yes—pre-practice grammar lesson | “[+F]”: grammar explanation “[-F]”: error signaling | Spanish Preverbal DO pronoun-OVS | 1) [+E, +F] grammar explanation and explicit feedback  
2) [+E, -F] grammar explanation and no explicit feedback  
3) [-E, +F] no grammar explanation and explicit feedback,  
4) [-E, -F] no grammar explanation and no explicit feedback. | Input-focused, TE-computerized practice | 1) All groups improved significantly and similarly.  
2) Their results suggested that explicit information (prior to practice or during practice) was not necessarily conducive to language development. |
<table>
<thead>
<tr>
<th>Study</th>
<th>Prior Rule presentation</th>
<th>Type of feedback</th>
<th>Language &amp; Targeted forms</th>
<th>Groups-conditions</th>
<th>Type of Practice</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stafford, Bowden, &amp; Sanz (2012)</td>
<td>Yes—pre-practice grammar lesson</td>
<td>“More explicit”: grammar explanation “Less explicit”: error signaling</td>
<td>Latin Assignment of thematic functions</td>
<td>1) [+ grammar, + explicit feedback] 2) [- grammar, + explicit feedback] 3) [+ grammar, - explicit feedback] 4) [-grammar, - explicit feedback]</td>
<td>Input-focused, TE-computerized practice</td>
<td>1) [-grammar, -explicit feedback] w/ TE practice is enough for improvement at interpretation level. 2) + explicit feedback was necessary for improvement at production level. 3) + explicit grammar only effective w/ explicit feedback.</td>
</tr>
</tbody>
</table>
Appendix B: Rules only in the feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Prior Rule presentation</th>
<th>Type of feedback</th>
<th>Language &amp; Targeted forms</th>
<th>Groups-conditions</th>
<th>Type of Practice</th>
<th>Results</th>
</tr>
</thead>
</table>
| Bowles (2008)    | No                      | “Explicit”: grammar explanation                      | Spanish                   | 1) [+ metaling verbal + explicit feedback];          | Input-focused, TE-computerized maze, sentence translation | 1) [+ explicit feedback] better at posttests.  
N = 194            |                         | “Implicit”: (inherent in the practice—wall closing the maze): error signaling | gustar                    | 2) [+ metaling verbal - explicit feedback];          | 2) Both groups statically similar at delayed post-tests. | 3) [- metaling verbal + explicit feedback];  
<pre><code>                            |                         |                                                        |                           | 3) [- metaling verbal + explicit feedback];          | 4) [- metaling verbal - explicit feedback]                                                                                             | 4) [- metaling verbal - explicit feedback]                                                                                             |
</code></pre>
<table>
<thead>
<tr>
<th>Study</th>
<th>Prior Rule presentation</th>
<th>Type of feedback</th>
<th>Language &amp; Targeted forms</th>
<th>Groups-conditions</th>
<th>Type of Practice</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camblor (2006) N = 77</td>
<td>No</td>
<td>“Explicit”: error signaling &amp; error location “Implicit”: (inherent in the task—textual enhancement): error signaling Interactive feedback</td>
<td>Spanish Noun-adjective gender &amp; number agreement</td>
<td>1) Implicit feedback: error signaled through textual enhancement 2) Explicit feedback: error signaled through textual enhancement. Error 3) Interactive feedback: a problem-solving exercise to guide to correct answer 4) Control: no feedback</td>
<td>Output-focused, computerized, sentence production task</td>
<td>1) Feedback provision resulted beneficial. 2) No conclusions can be drawn as for the beneficial effects of one type of feedback over another</td>
</tr>
<tr>
<td>Study</td>
<td>Prior Rule presentation</td>
<td>Type of feedback</td>
<td>Language &amp; Targeted forms</td>
<td>Groups-conditions</td>
<td>Type of Practice</td>
<td>Results</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Spot”: Error signaling</td>
<td></td>
<td>2) “Explain &amp; Prompt”: [Grammar explanation, prompt]</td>
<td></td>
<td>2) “Explain” groups better results at immediate posttests, but better effects washed out by delayed posttest.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Continue”: no error repair by learner</td>
<td></td>
<td>3) “Spot &amp; Continue”: [error signaling, no prompt]</td>
<td></td>
<td>3) Prompts had no significant effects.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Prompt”: error repair by learner</td>
<td></td>
<td>4) “Spot &amp; Prompt”: [error signaling, prompt]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“NF”: No feedback</td>
<td></td>
<td>5) “NF”: [no feedback]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Prior Rule presentation</td>
<td>Type of feedback</td>
<td>Language &amp; Targeted forms</td>
<td>Groups-conditions</td>
<td>Type of Practice</td>
<td>Results</td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------</td>
<td>------------------</td>
<td>---------------------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Hsieh (2007)</td>
<td>No</td>
<td>“Explicit”: grammar explanation</td>
<td>Spanish gusta</td>
<td>1) PRONLY [practice only, no feedback]</td>
<td>Input-focused, TE-computerized practice</td>
<td>1) Feedback groups significantly better than control (no feedback)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Implicit”: (inherent in the practice—a wall closing the maze): error signaling</td>
<td>2) PRIFB [practice, implicit feedback]</td>
<td>2) No significant differences between “explicit” and “implicit” feedback.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3) PREFB [practice, explicit feedback]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4) EXPONLY [exposure to practice, no feedback]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5) EXPIFB [exposure to practice, implicit feedback]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6) EXPEFB [exposure to practice, explicit feedback]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Prior Rule presentation</td>
<td>Type of feedback</td>
<td>Language &amp; Targeted forms</td>
<td>Groups-conditions</td>
<td>Type of Practice</td>
<td>Results</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Lado, Bowden, Stafford, & Sanz (2013) | No                      | "NE+MI": grammar explanation, "NE-MI": error signaling                          | Latin assignment of thematic functions         | 1) NE+ MI [negative evidence + explicit metaling feedback]                         | Input-focused, TE-computerized practice     | 1) Both groups improved in accuracy and RTs on immediate posttests.  
2) NE+ MI better than NE-MI at immediate posttests  
3) At delayed posttests, differences dissipated. |
| N = 58                       |                         |                                                                                  |                                                | 2) NE-MI [negative evidence - explicit metaling feedback]                         |                                               |                                                                                                                                 |
2) At delayed posttests: significantly greater gains of less explicit feedback;  
3) Explicit feedback group: significant greater loss between immediate and delayed posttests. Implicit feedback group did not. |
| N = 59                       |                         |                                                                                  |                                                | 2) [+TE, -EF]: TE practice, implicit feedback.                                     |                                               |                                                                                                                                 |
|                              |                         |                                                                                  |                                                | 3) [-TE, +EF]: non-TE practice, explicit feedback.                                  |                                               |                                                                                                                                 |
|                              |                         |                                                                                  |                                                | 4) [-TE, -EF]: non-TE practice, implicit feedback.                                  |                                               |                                                                                                                                 |
### Appendix C: Amount of feedback studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ Type of feedback</th>
<th>Targeted forms</th>
<th>Practice/Assessment tests</th>
<th>Conditions</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ellis, Sheen, Murakami, &amp; Takashima</td>
<td>Quasi-experimental study</td>
<td>English Articles</td>
<td>Task(s): 3 read &amp; rewrite stories Assessment: 1) Error correction test 2) Free writing</td>
<td>1) Focused [F]: Received direct CF only on articles 2) Unfocused [U]: Received direct CF on articles + more errors 3) Control: No CF</td>
<td>1) Both groups improved significantly from pre to post on Error Test &amp; Free Writing. 2) [U] maintained accuracy from immediate post to delayed posttest. -F, U &gt; Control -CF equally effective for both F and U groups.</td>
</tr>
<tr>
<td>(2008) N = 35</td>
<td>Type of feedback: “direct CF”: error signaling &amp; provision of right form</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Havranek</td>
<td>Exploratory study</td>
<td>English</td>
<td>Task(s): 6-8 whole class periods (w/ gap texts &amp; questions) Assessment: Class-specific language tests w/ items that received feedback &amp; an error correction section</td>
<td>“Self”: received CF “Peer”: audit peers’ CF</td>
<td>1) 50% of learners being corrected profited from feedback. 2) Repeated corrections on the same item did not contribute much to learning: a) hardly any difference in accuracy of items corrected 3 times b) a decrease of the accuracy of items corrected more than 3 times.</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/ Type of feedback</td>
<td>Targeted forms</td>
<td>Practice/Assessment tests</td>
<td>Conditions</td>
<td>Results</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>------------</td>
<td>---------</td>
</tr>
<tr>
<td>Loewen &amp; Nabei (2007) N = 66</td>
<td>Exploratory study (with regard to amount of feedback) Quasi experimental (with regard to type of feedback)</td>
<td>English Question formation</td>
<td>Task(s): 1) Spot-the-difference 2) Guess-the-storyline task. Assessment: 1) Timed GJT 2) Untimed GJT 3) Oral production task</td>
<td>1) Recast 2) Clarification 3) Metalinguistic 4) No CF</td>
<td>1) Frequency of feedback types: Recast (18), metalinguistic (5.6), Clarification (18) 2) Untimed GJT: No sig increase in scores for no feedback groups and no differences between groups. 3) Timed GJT: gain scores from pre to posttests for feedback groups but no differences between groups. 4) All feedback groups performed similarly in spite of receiving different amounts of feedback.</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/ Type of feedback</td>
<td>Targeted forms</td>
<td>Practice/Assessment tests</td>
<td>Conditions</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mackey, Oliver, &amp; Leeman (2003) N = 96</td>
<td>Experimental study</td>
<td>English</td>
<td>Task(s):</td>
<td>1) Adult NNs-NNs</td>
<td>1) Adult NSs provided more feedback than adult NNSs.</td>
</tr>
<tr>
<td></td>
<td>Recast</td>
<td>Grammar errors</td>
<td></td>
<td>2) Adult NNs-Ns</td>
<td>2) No sig diff in the amount provided in adults (NNs &amp; NSs) and child dyads (NNs &amp; NSs).</td>
</tr>
<tr>
<td></td>
<td>Confirmation checks</td>
<td></td>
<td></td>
<td>3) Child NNs-NNs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clarification requests</td>
<td></td>
<td></td>
<td>4) Child NNs-Ns</td>
<td></td>
</tr>
</tbody>
</table>

1) One-way information gap task
2) Two-way information gap task
In the following slides you will review the conjugation of some Spanish tenses.

You will be asked to write down some Spanish verbs. Please write your answers in the white sheet provided.

After answering the questions, you will check your answers with the solutions provided in this presentation.

If they are wrong, please write the correct form next to your answer.

The answer sheets will be collected at the end of the review.
You know this tense already, so we will quickly refresh how it's formed.

Remember how to form the preterit?

Verb stem + endings

<table>
<thead>
<tr>
<th>VERB STEMS</th>
<th>-AR VERB ENDINGS</th>
<th>-ER/-IR VERB ENDINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hablar</td>
<td>-é</td>
<td>-í</td>
</tr>
<tr>
<td></td>
<td>-aste</td>
<td>-iste</td>
</tr>
<tr>
<td></td>
<td>-ó</td>
<td>-ió</td>
</tr>
<tr>
<td>Comer</td>
<td>-amos</td>
<td>-imos</td>
</tr>
<tr>
<td></td>
<td>-asteis</td>
<td>-isteis</td>
</tr>
<tr>
<td>Vivir</td>
<td>-aron</td>
<td>-ieron</td>
</tr>
</tbody>
</table>
So now, please write down the preterit form of the following verbs:

- Yo, comer ➔ comí
- Él, saltar ➔ saltó
- Tú, asistir ➔ asististe

Don't forget the accent marks!

Preterit

Ok, now check your answers

Preterit-answers
Preterit

- How did it go? Feel free to go back using the Page Up/Down keys if you want to review the forms.

- Now, do you remember some irregular verbs in preterit? Please use your sheet to write some down.

- After that, let’s move on to another tense.

PRESENT SUBJUNTIVE

This is a newer tense, so we will take some more time to review it.
How do you form the present subjunctive?

Verb stem + endings

<table>
<thead>
<tr>
<th>VERB STEMS</th>
<th>-AR VERB ENDINGS</th>
<th>-ER/-IR VERB ENDINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hablar&gt;Habl-</td>
<td>-e</td>
<td>-a</td>
</tr>
<tr>
<td></td>
<td>-es</td>
<td>-as</td>
</tr>
<tr>
<td></td>
<td>-e</td>
<td>-a</td>
</tr>
<tr>
<td>Comer&gt;Com-</td>
<td>-emos</td>
<td>-amos</td>
</tr>
<tr>
<td></td>
<td>-éis</td>
<td>-áis</td>
</tr>
<tr>
<td>Vivir&gt;Viv-</td>
<td>-en</td>
<td>-an</td>
</tr>
</tbody>
</table>

Did you notice something? Yes, the endings are the same as in the Present Indicative.

However, in subjunctive the endings for -ar verbs and -er/-ir verbs are swapped when compared to the endings of these verbs in Present indicative.


If you want to take a look at the endings again, please use the Page Up/Down keys to review them.
Present Subjunctive

Now, can you write down the following forms of these -AR verbs? As an example: Yo, tomar → yo tome

- Yo, trabajar →
- Él, estudiar →
- Yo, acabar →

Ok, now check your answers

Present Subjunctive-answers

- Yo, trabajar → yo trabaje
- Él, estudiar → él estudie
- Yo, acabar → yo acabe
How did it go? Feel free to go back using the Page Up/Down keys if you want to review the forms.

Now, let’s see –ER verbs.

Can you write down the following forms of these –ER verbs? As an example: Yo, comer ➔ yo coma

Yo, beber ➔

Él, leer ➔

Yo, responder ➔

Ok, now check your answers
Present Subjunctive-answers

- Yo, beber ➔ beba
- Él, leer ➔ lea
- Yo, responder ➔ responda

Present Subjunctive

- How did it go? Feel free to go back using the Page Up/Down keys if you want to review the forms.
- Now, let’s see –IR verbs.

Present Subjunctive

Can you write down the following forms of these –IR verbs? As an example: Yo, vivir ➔ yo viva

- Yo, escribir ➔
- Él, decidir ➔
- Yo, asistir ➔
Ok, now check your answers

Present Subjunctive-answers

- Yo, escribir → escriba
- Él, decidir → decida
- Yo, asistir → asista

Present Subjunctive

- How did it go? Feel free to go back if you want to review the forms.
- Now, do you remember some irregular verbs in Present Indicative? Well, those are also irregular in Present Subjunctive.
- Let’s see some irregulars.
Present Subjunctive-irregular verbs

Irregular verbs in Present Indicative

- **-go verbs in the ‘yo’ form. E.g. Tener > Tengo**
  - These verbs are also irregular in Present Subjunctive.
  - To form the Present Subjunctive of these verbs, you need to use the stem from the ‘yo’ form of Present Indicative, and then plug the endings for Present Subjunctive.
  - E.g. Teng-o > Teng- +
    - **-a >** Tenga
    - **-as** Tengas
    - **-a** Tenga
    - **-amos** Tengamos
    - **-áis** Tengáis
    - **-an** Tengan

- **E>IE** E.g. Querer > Quiero
- **O>UE** E.g. Volver > Vuelvo
- **E>I** E.g. Pedir > Pido
  - These verbs are also irregular in Present Subjunctive.
  - Remember that not all the forms are irregular.

<table>
<thead>
<tr>
<th>E.g. Querer</th>
<th>E.g. Volver</th>
<th>E.g. Pedir</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiera</td>
<td>Vuelva</td>
<td>Pida</td>
</tr>
<tr>
<td>Quieras</td>
<td>Vuelvas</td>
<td>Pidas</td>
</tr>
<tr>
<td>Quiera</td>
<td>Vuela</td>
<td>Pida</td>
</tr>
<tr>
<td>Queramos</td>
<td>Volvamos</td>
<td>Pidamos</td>
</tr>
<tr>
<td>Queréis</td>
<td>Volváis</td>
<td>Pidáis</td>
</tr>
<tr>
<td>Quieran</td>
<td>Vuelvan</td>
<td>Pidan</td>
</tr>
</tbody>
</table>
Present Subjunctive-irregular verbs

Now, let’s see how you conjugate the following irregular verbs in Present Subjunctive

- Yo, hacer
- Él, pensar
- Yo, salir
- Ellos, tener
- Tú, querer
- Ella, poder
- Yo, vestirse

Present Subjunctive

Ok, now check your answers
Present Subjunctive-irregular verbs

Answers

- Yo, hacer ➔ haga
- Él, pensar ➔ piense
- Yo, salir ➔ salga
- Ellos, tener ➔ tengan
- Tú, querer ➔ quieras
- Ella, poder ➔ pueda
- Yo, vestirse ➔ me vista

Present Subjunctive

- How did it go? Feel free to go back using the Page Up/Down keys if you want to review the forms.
Recognizing the tenses

Now we will see how easy it’s for you to recognize the tenses present indicative, preterit, and present subjunctive

Indicative, subjunctive, or preterit?

- Indicate the tense of each verb and its infinitive

<table>
<thead>
<tr>
<th>BLOQUE 1</th>
<th>BLOQUE 2</th>
<th>BLOQUE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conozca</td>
<td>Dí</td>
<td>Llueve</td>
</tr>
<tr>
<td>Tenga</td>
<td>Dé</td>
<td>Llueva</td>
</tr>
<tr>
<td>Llegue</td>
<td>Me convierta</td>
<td>Se celebre</td>
</tr>
<tr>
<td>Sepa</td>
<td>Me gradúo</td>
<td>Se celebró</td>
</tr>
<tr>
<td>Sé</td>
<td>Me gradúe</td>
<td></td>
</tr>
<tr>
<td>Supe</td>
<td>Fui</td>
<td>Lleve</td>
</tr>
<tr>
<td>Conocí</td>
<td>Vaya</td>
<td>Llevó</td>
</tr>
<tr>
<td>Me case</td>
<td>Haya</td>
<td>Hace</td>
</tr>
<tr>
<td></td>
<td>Cumpla</td>
<td>Haga</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sea</td>
</tr>
</tbody>
</table>
Ok, now check your answers

BLOQUE 1
Conozca—Present subjunctive; conocer: to meet for first time/to know
Tenga—Present subjunctive; tener: to have
Llegue—Present subjunctive; llegar: to arrive
Sepa—Present subjunctive; saber: to know
Sé—Present indicative; saber: to know
Supe—Preterit; saber: to know
Conocí—Preterit; conocer: to meet for first time/to know
Me case—Present subjunctive; casarse: to get married
BLOQUE 2
Di—Preterit; dar: to give
Dé—Present subjunctive; dar: to give
Me convierta—Present subjunctive; convertirse: to become
Me gradúo—Present indicative; graduarse: to graduate
Me gradúe—Present subjunctive; graduarse: to graduate
Fui—Preterit; ir or ser: to go or to be
Vaya—Present subjunctive; ir: to go
Haya—Present subjunctive; haber: there to be
Cumpla—Present subjunctive; cumplir: to turn years

BLOQUE 3
Llueve—Present indicative; llover: to rain
Llueva—Present subjunctive; llover: to rain
Se celebre—Present subjunctive; celebrarse: to celebrate
Se celebró—Preterit; celebrarse: to celebrate
Lleve—Present subjunctive; llevar: to take
Llevó—Preterit; llevar: to take
Hace—Present indicative; hacer: to do
Haga—Present subjunctive; hacer: to do
Sea—Present subjunctive; ser: to be
Tense usage

In general:
- Preterit tense is used for actions that occurred in the past.
- Present tense is used for habitual actions.
- Future tense for is used for actions that will occur in the future.

HOWEVER, sometimes present tense can also have a future meaning. See the following examples:

Ejemplo 1:
A: Esta noche cenamos en casa de Ángela. (We’re having dinner tonight at Angela’s)
B: Vale. ¡Yo llevo las bebidas! (Ok. I’ll bring the drinks).

Ejemplo 2:
En noviembre voy a Cancún de vacaciones con mi familia. (In November I’m going to Cancun on vacation with my family)
Now you will see a few pictures. Pay attention to the pictures and the sentence below. Please decide whether the action expressed in the sentence occurs in the present, occurred in the past, or will occur in the future.

What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to…

a) The present
b) The past
c) The future

En el 2016 pienso volar alrededor del mundo.
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

Juego al baloncesto después de clase.
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to…

a) The present
b) The past
c) The future

Ana vendía galletas de las Girl Scouts.

What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to…

a) The present
b) The past
c) The future

Me voy de D.C. para ir a la escuela graduada.
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

What point in time does the action expressed in the sentence refer to?

Le regalé flores a mi novia por San Valentín.

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

Saco a mi perro a pasear.

Sacar a pasear: to walk the dog

Ok, now check your answers
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

En el 2016 pienso volar alrededor del mundo.

What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

En noviembre quemo las hojas de los árboles.
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present  

b) The past  

c) The future  

Juego al baloncesto después de clase.

What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present  

b) The past  

c) The future  

Ana vendía galletas de las Girl Scouts.
What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

Me voy de D.C. para ir a la escuela graduada.

What point in time does the action expressed in the sentence refer to?

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

Estudio después de clase.
What point in time does the action expressed in the sentence refer to?

Le regalé flores a mi novia por San Valentín.

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future

What point in time does the action expressed in the sentence refer to?

Lunes Martes Miércoles Jueves

Saco a mi perro a pasear.

Sacar a pasear: to walk the dog

The action expressed in the sentence refers to...

a) The present
b) The past
c) The future
Did you notice that…?

- When the sentence referred to the present, there were several pictures to indicate an habitual action?
- When the sentence referred to the future, there was a bubble above to indicate a future thought?
- When the sentence referred to the past, there was a bubble below to indicate a memory?

*If you didn’t notice this, please go back and check the pictures*

Values of **Present Indicative**

Notice that **Present Indicative** can have 2 values/meanings:

1) To express something that happens habitually. (present value)
2) To express something that will happen in the future. (future value)
Present indicative (present value)

To express something that happens habitually:

Ej.: **Me levanto** todos los días a las 8 de la mañana.

Me levanto todos los días a las 8 de la mañana.

Present indicative (present value)

To express something that happens habitually:

Ej.: **Juego al baloncesto después de clase**.
Present indicative (**future** value)

To express something that will happen in the near future.

Ej.: *Esta noche yo hago la cena.*

Present indicative (**future** value)

To express something that will happen in the near future.

Ej.: *En noviembre quemo las hojas de los árboles.*
You’re done with this review! Thanks!
Now, please call the researcher
Appendix E: Review session answer sheet

REVIEW SESSION-ANSWER SHEETS

PARTICIPANT NUMBER:

1. PRETERIT

A) (slide 4)

☐ Yo, comer ➔

☐ Él, saltar ➔

☐ Tú, asistir ➔

B) Irregular verbs:

2. PRESENT SUBJUNCTIVE

A) -AR verbs (slide 11)

☐ Yo, trabajar ➔

☐ Él, estudiar ➔

☐ Yo, acabar ➔

B) -ER verbs (slide 15)

☐ Yo, beber ➔

☐ Él, leer ➔

☐ Yo, responder ➔

C) -IR verbs (slide 19)

☐ Yo, escribir ➔
- Él, decidir
- Yo, asistir

**D) IRREGULAR VERBS (slide 26)**
- Yo, hacer
- Él, pensar
- Yo, salir
- Ellos, tener
- Tú, querer
- Ella, poder
- Yo, vestirse

**3. ¿INDICATIVO, SUBJUNTIVO O PRETÉRITO? Indicate the tense and the infinitive of each verb.**

**BLOQUE 1 (slide 32/34)**
- Conozca
- Tenga
- Llegue
- Sepa
- Sé
- Supe
- Conocí
- Me case

**BLOQUE 2 (slide 32/35)**
• Di
• Dé
• Me convierta
• Me gradúo
• Me gradúe
• Fui
• Vaya
• Haya
• Cumpla

BLOQUE 3 (slide 32/36)

• Llueve
• Llueva
• Se celebre
• Se celebró
• Lleve
• Llevó
• Hace
• Haga
• Sea

4. PICTURES

Slide 40—En el 2016 pienso volar alrededor del mundo: __________

Slide 41—En noviembre quemo las hojas de los árboles: __________

Slide 42—Juego al balonces después de clase: __________
Slide 43—Ana vendía galletas de las Girl Scouts: ________________

Slide 44—Me voy de DC para ir a la escuela graduada: ________________

Slide 45—Estudio después de clase: ________________

Slide 46—Le regale flores a mi novia por San Valentín: ________________

Slide 47—Saco a mi perro a pasear: ________________
Appendix F: Stimuli of the treatment

TE condition

You will be presented with sets of pictures and different sentences.

Please select the sentence that best fits what is conveyed in the picture by pressing a), b), or c) on the keyboard.

Please answer as ACCURATE and as QUICK as possible.

Press the SPACE key to start.

Present tense items

<table>
<thead>
<tr>
<th>Lunes</th>
<th>Martes</th>
<th>Miércoles</th>
<th>Jueves</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td>Viernes</td>
<td>Sábado</td>
<td>Domingo</td>
<td></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>

a. Desayuno cuando me levante.
b. Desayuno cuando me levanto.
c. Desayuné cuando me levanté.
a. Me acuesto cuando termine mi tarea.
b. Me acosté cuando terminé mi tarea.
c. Me acuesto cuando termino mi tarea.

Cada verano

a. Puse el aire acondicionado cuando hacía calor.
b. Pongo el aire acondicionado cuando hace calor.
c. Pongo el aire acondicionado cuando haga calor.

*poner: to turn on*
a. Cuando hice ejercicio me duché.
b. Cuando haga ejercicio me ducho.
c. Cuando hago ejercicio me ducho.

a. Cuando tengo sueño me voy a la cama.
b. Cuando tuve sueño me fui a la cama.
c. Cuando tenga sueño me voy a la cama.
a. Cuando el primero de cada mes llega pago mi alquiler.
b. Cuando el primero de cada mes llegue pago mi alquiler.
c. Cuando el primero de cada mes llegaba pagaba mi alquiler.

a. María practicaba rafting cuando iba al río.
b. María practica rafting cuando va al río.
c. María practica rafting cuando vaya al río.
Cada fin de semana

a. Bailo con mi novio cuando vaya a la discoteca.

b. Bailaba con mi novio cuando iba a la discoteca.

c. Bailo con mi novio cuando voy a la discoteca.

a. Cuando entrenaba cenaba pasta.

b. Cuando entrene ceno pasta.

c. Cuando entreno ceno pasta.

entrenar: to train     cenar: to have dinner
a. Cuando me duchaba cantaba ópera.

b. Cuando me duche canto ópera.

c. Cuando me ducho canto ópera.

Cada semana

a. Cuando no tengo comida hago la compra.

b. Cuando no tenga comida hago la compra.

c. Cuando no tenía comida hacía la compra.

hacer la compra: to do grocery shopping
Subjunctive items

a. Cuando me levanto salgo a correr.
b. Cuando me levantaba salía a correr.
c. Cuando me levante salgo a correr.

a. Me compro una casa bonita cuando tenga mucho dinero.
b. Me compro una casa bonita cuando tengo mucho dinero.
c. Me compré una casa bonita cuando tuve mucho dinero.
Mayo 2016

a. Hago un viaje cuando termine el curso.
b. Hice un viaje cuando terminé el curso.
c. Hago un viaje cuando termino el curso.

Octubre 2015

2015

2018

a. Me casé cuando conocí a la chica de mis sueños.
b. Me caso cuando conozco a la chica de mis sueños.
c. Me caso cuando conozca a la chica de mis sueños.
a. Cuando mi hija cumplió 16 años, yo le compré un coche.

b. Cuando mi hija cumpla 16 años, yo le compro un coche.

c. Cuando mi hija cumple 16 años, yo le compro un coche.

a. Compro las entradas cuando llego al teatro.

b. Compro las entradas cuando llegue al teatro.

c. Compré las entradas cuando llegué al teatro.
a. Voy a Francia cuando aprendo francés.
b. Fui a Francia cuando aprendí francés.
c. Voy a Francia cuando aprenda francés.

a. Apago el despertador cuando suene por la mañana.
b. Apagué el despertador cuando sonó por la mañana.
c. Apago el despertador cuando suena por la mañana.

sonar: to sound
a. Preparo la cena cuando llegue a casa.
b. Preparé la cena cuando llegué a casa.
c. Preparo la cena cuando llego a casa.

a. Hablé con mi familia cuando llegué a casa.
b. Hablo con mi familia cuando llego a casa.
c. Hablo con mi familia cuando llegue a casa.
a. Cuando me gradué viajé a Nueva York.
b. Cuando me gradúe viajo a Nueva York.
c. Cuando me gradúo viajo a Nueva York.

a. Cuando llego a casa cocino algo.
b. Cuando llegue a casa cocino algo.
c. Cuando llegué a casa cociné algo.
a. Cuando hablo con mi profesor empiezo el ensayo.
b. Cuando hablé con mi profesor empecé el ensayo.
c. Cuando hable con mi profesor empiezo el ensayo.

a. Cuando sepa los resultados de los exámenes llamo a mis padres.
b. Cuando sé los resultados de los exámenes llamo a mis padres
   c. Cuando supe los resultados de los exámenes llamé a mis padres.
a. Cuando el 4 de julio llegue, veo los fuegos artificiales.

b. Cuando el 4 de julio llegó, vi los fuegos artificiales.

c. Cuando el 4 de julio llega, veo los fuegos artificiales.

---

Diciembre 2015

fuegos artificiales: fireworks

---

a. Cuando fue la fecha límite devolví los libros a la biblioteca.

b. Cuando es la fecha límite devuelvo los libros a la biblioteca.

c. Cuando sea la fecha límite devuelvo los libros a la biblioteca.

fecha límite: due date    devolver: to return
a. Cuando mi profesor me dio una tarea yo la hice.
b. Cuando mi profesor me dé una tarea yo la hago.
c. Cuando mi profesor me da una tarea yo la hago.

a. Cuando acabo mis clases hago ejercicio en el gimnasio.
b. Cuando acabe mis clases hago ejercicio en el gimnasio.
c. Cuando acabé mis clases hice ejercicio en el gimnasio.
a. Cuando tengo dinero extra me compro ropa nueva.

b. Cuando tenía dinero extra me compraba ropa nueva.

c. Cuando tenga dinero extra me compro ropa nueva.

Diciembre 2015

a. Me disfrazo cuando se celebre Halloween.

b. Me disfrecé cuando se celebró Halloween.

c. Me disfrazo cuando se celebra Halloween.
a. Mi padre me regaló una motocicleta cuando saqué buenas notas.
b. Mi padre me regala una motocicleta cuando saque buenas notas.
c. Mi padre me regala una motocicleta cuando saco buenas notas.

regalar: to make a gift

a. Fui de visita cuando mi abuelo cumplió 90 años.
b. Voy de visita cuando mi abuelo cumpla 90 años.
c. Voy de visita cuando mi abuelo cumple 90 años.

cumplir: to turn (years)
a. Cuando vaya a la playa me compro protección solar.
b. Cuando voy a la playa me compro protección solar.
c. Cuando fui a la playa me compré protección solar.
Noviembre 2015

a. Vuelo en globo cuando me case.
b. Volé en globo cuando me casé.
c. Vuelo en globo cuando me caso.

Octubre 2015

casarse: to get married   globo: balloon

2016

a. Me voy a Hawaii cuando me convierto en campeón profesional de surf.
b. Me fui a Hawaii cuando me convertí en campeón profesional de surf.
c. Me voy a Hawaii cuando me convierta en campeón profesional de surf.

convertirse: to become   campeón: champion

2015
a. Mi amigo Alex se mudó cuando empezó la universidad en NY.
b. Mi amigo Alex se muda cuando empieza la universidad en NY.
c. Mi amigo Alex se muda cuando empiece la universidad en NY.

mudarse: to move

a. Esquío cuando haya mucha nieve en diciembre.
b. Esquío cuando hay mucha nieve en diciembre.
c. Esquéi cuando había mucha nieve en diciembre.

esquiar: to ski
2016

a. Me corté mucho el pelo cuando entré en el ejército.

b. Me corto mucho el pelo cuando entro en el ejército.

c. Me corto mucho el pelo cuando entre en el ejército.

cortarse el pelo: to get a haircut    ejército: military

2015

a. Cuando nació mi hijo compré un carrito.

b. Cuando nace mi hijo compro un carrito.

c. Cuando nazca mi hijo compro un carrito.

carro: stroller    nacer: to be born

Octubre 2015

Diciembre 2015
Past tense items

a. Me compro un coche nuevo cuando obtenga mi licencia de conducir.
b. Me compro un coche nuevo cuando obtengo mi licencia de conducir.
c. Me compré un coche nuevo cuando obtuve mi licencia de conducir.

a. Me compruego una mochila azul cuando vaya a la escuela primaria.
b. Me comprqué una mochila azul cuando iba a la escuela primaria.
c. Me compruego una mochila azul cuando voy a la escuela primaria.
a. Vendía limonada en la calle cuando era niño.
b. Vendo limonada en la calle cuando soy niño.
c. Vendo limonada en la calle cuando sea niño.

a. Cuando reconocí al actor le pedí un autógrafo.
b. Cuando reconozca al actor le pido un autógrafo.
c. Cuando reconozco al actor le pido un autógrafo.
a. Cuando tengo vacaciones busco un trabajo para el verano.
b. Cuando llegaba a la playa construía un castillo de arena.
c. Cuando llegue a la playa construyo un castillo de arena.

2. a. Cuando tengo vacaciones busco un trabajo para el verano.
b. Cuando tenga vacaciones busco un trabajo para el verano.
c. Cuando tuve vacaciones busqué un trabajo para el verano.

Junio 2015

Septiembre 2015

Se busca ayuda para el verano (222) 222-2002
a. Cuando estaba en una banda de rock me dejé el pelo largo.
b. Cuando estoy en una banda de rock me dejo el pelo largo.
c. Cuando esté en una banda de rock me dejo el pelo largo.

d. Hago muchas acrobacias cuando soy animadora.
b. Hago muchas acrobacias cuando sea animadora.
c. Hacía muchas acrobacias cuando era animadora.
Non TE condition

You will be presented with some pictures and a sentence.

Please select the picture that best describes what is said in the sentence by pressing a), b), or c) on the keyboard.

Please answer as ACCURATE and as QUICK as possible.

After you submit your answer, YOU CANNOT GO BACK.

Press the SPACE key to start.

Present tense items

Desayuno cuando me levanto.
Me acuesto cuando termino mi tarea.

Pongo el aire acondicionado cuando hace calor.
Cuando hago ejercicio me ducho.

Cuando tengo sueño me voy a la cama.
Cuando el primero de cada mes llega, pago mi alquiler.

María practica rafting cuando va al río.
Bailo con mi novio cuando voy a la discoteca.

Cuando entreno cenó pasta.
Cuando me ducho canto ópera.

hacer la compra: to do grocery shopping

Cuando no tengo comida hago la compra.
Cuando me levanto salgo a correr.

Subjunctive items

Me compro una casa bonita cuando tenga mucho dinero.
Hago un viaje en coche cuando termine el curso.

Me caso cuando conozca a la chica de mis sueños.
Cuando mi hija cumpla 16 años, yo le compro un coche.

Compro las entradas cuando llegue al teatro.
Voy a Francia cuando aprenda francés.

2017

2017

2017

A

B

C

Apago el despertador cuando suene por la mañana

Martes, 8:00am

Martes, 8:00am

Martes, 8:00am

A

B

C

Lunes, 10:00pm

sonar: to sound
Preparo la cena cuando llegue a casa.

Hablo con mi familia cuando llegue a casa.
Cuando me gradúe, voy a Nueva York.

Cuando llegue a casa, cocino algo.
Cuando hable con mi profesor, empiezo el ensayo.

Cuando sepa los resultados de los exámenes, llamo a mis padres.
Cuando el 4 de julio llegue, veo los fuegos artificiales.

Cuando sea la fecha límite, devuelvo los libros a la biblioteca.
Cuando mi profesor me dé una tarea, yo la hago.

Cuando acabe mis clases, hago ejercicio.
Cuando tenga dinero extra me compro ropa.

Me disfrazo cuando se celebre Halloween.

disfrazarse: to dress up
Mi padre me regala una motocicleta cuando saque buenas notas.

Voy de visita cuando mi abuelo cumpla 90 años.
Cuando llegue el año nuevo abro una botella de champán.

Cuando vaya a la playa me compro protección solar.
Vuelo en globo cuando me case.

agosto
globo: balloon
casarse: to get married

Me voy a Hawaii cuando me convierta en campeón profesional de surf.

convertirse: to become
campeón: champion
Mi amigo Alex se muda cuando empiece la universidad en NY.

Mi amigo Alex se muda cuando empiece la universidad en NY.

Esquío cuando haya mucha nieve en diciembre.

Esquío cuando haya mucha nieve en diciembre.
Me corto mucho el pelo cuando entre en el ejército.

A

B 2016

C 2016

ejército: military
cortarse el pelo: to get a haircut

2015

Cuando nazca mi hijo compro un carrito.

A Diciembre

B Diciembre

C Diciembre

carrito: stroller
nacer: to be born

Octubre
Past tense items

Me compré un coche nuevo cuando obtuve mi licencia de conducir.

Tenía una mochila azul cuando iba a la escuela

obtener: to get, to obtain
Vendía limonada en la calle cuando era niño.

Cuando reconocí al actor, le pedí un autógrafo.
Cuando llegaba a la playa, construía un castillo de arena.

Cuando tuve vacaciones busqué un trabajo para el verano.
Cuando estaba en una banda de rock me dejé el pelo largo.

Hacía muchas acrobacias cuando era animadora.
Appendix G: Stimuli of the interpretation test

Interpretation test-Version 1
Present tense items

Vivo en Inglaterra, por eso...

Veo la tele...

Tomo café...

a. ...cuando llueve fuera.
b. ...cuando lloverá fuera.
c. ...cuando llueva fuera.
d. ...cuando llovió fuera.

a. ...cuando esté cansada.
b. ...cuando estaba cansada.
c. ...cuando estoy cansada.
d. ...cuando estaré cansada.
Cada mes

Lavo el coche...

a. ...cuando estará sucio.
b. ...cuando esté sucio.
c. ...cuando estaba sucio.
d. ...cuando está sucio.

Enciendo la radio...

a. ...cuando preparo el desayuno.
b. ...cuando prepararé el desayuno.
c. ...cuando preparé el desayuno.
d. ...cuando prepare el desayuno.
a. ...cuando mi madre me lleva a la escuela.

Juego con mi teléfono...

b. ...cuando mi madre me lleve a la escuela.

c. ...cuando mi madre me llevaba a la escuela.

d. ...cuando mi madre me llevará a la escuela.

Lunes Martes Miércoles

Jueves Viernes

a. Cuando yo le dé un hueso...

b. Cuando yo le doy un hueso...

...mi perro salta feliz

c. Cuando yo le daba un hueso...

d. Cuando yo le daré un hueso...

Hueso: bone
a. Cuando salga de entrenar...

b. Cuando salgo de entrenar... ...como fruta.

c. Cuando saldré de entrenar...

d. Cuando salía de entrenar...

Entrenar: to train

a. Cuando salía para la escuela...

b. Cuando saldré para la escuela... ...me llevo una manzana.

c. Cuando salgo para la escuela...

d. Cuando salga para la escuela...
a. Cuando mi clase de laboratorio terminará...

b. Cuando mi clase de laboratorio terminaba...  ...leó el periódico.

c. Cuando me monte en el metro...

d. Cuando me monto en el metro...  

Montarse: to get on

a. Cuando mi clase de laboratorio terminará...

b. Cuando mi clase de laboratorio terminaba...  ...almuerzo en la cafetería.

c. Cuando mi clase de laboratorio termina...

d. Cuando mi clase de laboratorio termine...

Almorzar: to have lunch
a. ...cuando acabaré de comer.

b. ...cuando acababa de comer.

c. ...cuando acabe de comer.

da. ...cuando acabo de comer.

Hago mi tarea...

Vivo en California, por eso...

a. ...cuando voy a la playa.

b. ...cuando iba a la playa.

c. ...cuando vaya a la playa.

da. ...cuando iré a la playa.
Apago la alarma...

a. ...cuando llego a casa.

b. ...cuando llegué a casa.

c. ...cuando llegue a casa.

d. ...cuando llegaré a casa.

*apagar: to turn off*

La temperatura baja...

a. ...cuando el sol se ponía.

b. ...cuando el sol se pone.

c. ...cuando el sol se ponga.

d. ...cuando el sol se pondrá.

*Ponerse: to set*
a. Cuando me levantaba...

b. Cuando me levanto...  ...hago la cama.

c. Cuando me levantaré...

d. Cuando me levante...

**Cada día empiezo a trabajar a las 9am, por eso...**

a. Cuando darán las 8 de la mañana...

b. Cuando dan las 8 de la mañana...  ...sucita mi despertador.

c. Cuando den las 8 de la mañana...

d. Cuando daban las 8 de la mañana...
Cada día no tengo tiempo en casa, por eso...

a. Cuando llegue al trabajo...
b. Cuando llegaba al trabajo... ...yo desayuno.
c. Cuando llegaré al trabajo...
d. Cuando llego al trabajo...

a. Cuando tengo hambre...
b. Cuando tendré hambre... ...preparo comida.
c. Cuando tenía hambre...
d. Cuando tenga hambre...
a. Cuando termine de lavarme los dientes...

b. Cuando terminaré de lavarme los dientes...

c. Cuando terminaba de lavarme los dientes...  ...me acuesto.

d. Cuando termino de lavarme los dientes...

**Subjunctive items**

a. ...cualdo tomo las medicinas.

b. ...cualdo tome las medicinas.

c. ...cualdo tomé las medicinas.

d. ...cualdo tomaré las medicinas.
a. ...cuando voy a Japón.

b. ...cuando vaya a Japón.

c. ...cuando fui a Japón.

d. ...cuando iré a Japón.

Pruebo el sake...

a. ...cuando el precio del billete de avión baje.

b. ...cuando el precio del billete de avión baja.

c. ...cuando el precio del billete de avión bajó.

d. ...cuando el precio del billete de avión bajará.

Viajo a Asia...
a. Cuando me compré un barco...  
  b. Cuando me compraré un barco...  
    ...navego en el mar.  
  c. Cuando me compre un barco...  
  d. Cuando me compro un barco...  
  
Navegar: to sail

a. ...cuando esté en Las Vegas.  
  b. ...cuando estoy en Las Vegas.  
  c. ...cuando estuve en Las Vegas.  
  d. ...cuando estaré en Las Vegas.
a. ...cuando obtuve mi diploma del MFA.
b. ...cuando obtendré mi diploma del MFA.
c. ...cuando obtenga mi diploma del MFA.
d. ...cuando obtengo mi diploma del MFA.

Escribo un libro...

a. Cuando me compré un barco...
b. Cuando me compraré un barco...
c. Cuando me compré un barco...
d. Cuando me compro un barco...

...navego en el mar.

Navigar: to sail
A. ...cuando me convertiré en chef.
B. ...cuando me convertí en chef.
C. ...cuando me convierto en chef.
D. ...cuando me convierta en chef.

**convertirse: to become**

Abro un restaurante...

A. Cuando subiré al Himalaya...
B. Cuando subí al Himalaya...
C. Cuando suba al Himalaya...
D. Cuando subo al Himalaya...

**subir: to climb  bandera: flag**

...pongo una bandera de mi país.
a. Cuando paso el examen de bomberos...

b. Cuando pasé el examen de bomberos... me hago bombero.

c. Cuando pase el examen de bomberos...

d. Cuando pasaré el examen de bomberos... bombero: firefighter

a. Cuando visito la Antártida...

b. Cuando visité la Antártida... fotografío a un pingüino.

c. Cuando visitaré la Antártida...

d. Cuando visite la Antártida...
Cuando termine de comprar.  
Cuando termine de comprar.  
Cuando terminaré de comprar.

Cuando me gradué.  
Cuando me gradúo.  
Cuando me graduaré.

Traje: suit
Cuando me haré policía...

...cojo a un ladrón.

Cuando me hago policía...

Cuando me haga policía...

Cuando me hice policía...

coger: to catch

Cuando esté de vacaciones en la playa...

...buceo en el mar.

Cuando estaba de vacaciones en la playa...

Cuando estaré de vacaciones en la playa...

Cuando estoy de vacaciones en la playa...

bucear: scuba diving
Voy a Italia...

2015

2016

a. ...cuando hablo italiano.
b. ...cuando hablará italiano.
c. ...cuando hablé italiano.
d. ...cuando hable italiano.

diciembre 2015

a. Cuando conozca al presidente...
b. Cuando conoceré al presidente....
c. Cuando conocí al presidente...
d. Cuando conozco al presidente...

...llamo a mis padres.

octubre 2015
a. Cuando llega el otoño...

b. Cuando llegará el otoño...

c. Cuando llegue el otoño...

d. Cuando llegó el otoño...

...me caso.

a. Cuando tendré la receta...

b. Cuando tenga la receta...

...cocino un pavo.

c. Cuando tengo la receta...

d. Cuando tenía la receta...
a. Cuando termine con mis estudios...

b. Cuando terminaré con mis estudios...  ...voy de vacaciones

c. Cuando termino con mis estudios...

d. Cuando terminé con mis estudios...

a. ...cuando se celebrará el Prom.

b. ...cuando se celebra el Prom.

c. ...cuando se celebró el Prom.

d. ...cuando se celebre el Prom.
Past tense items

a. Cuando se me cura el brazo...
b. Cuando se me cure el brazo... ...juego al béisbol
c. Cuando se me curará el brazo...
d. Cuando se me curó el brazo...

curarse: to heal

...juego al béisbol

Montaba en bicicleta a la escuela...

a. ...cuando mi madre me regala una bicicleta.
b. ...cuando mi madre me regale una bicicleta.
c. ...cuando mi madre me regaló una bicicleta.
d. ...cuando mi madre me regalará una bicicleta.

regalar: to give as a present

Regaló una bicicleta en el año 2000...

Montaba en bicicleta a la escuela en el año 2015...

montar: to ride

Regaló una bicicleta en el año 2000...

Regaló una bicicleta en el año 2015...
a. Cuando el profesor no mira...
b. Cuando el profesor no mire...
c. Cuando el profesor no mirará...  ...yo dibujaba en el cuaderno.
d. Cuando el profesor no miraba...

dibujar: to draw

a. Cuando llegue a casa...
b. Cuando llego a casa...  ...jugaba a la pelota con mi hermano.
c. Cuando llegaba a casa...
d. Cuando llegaré a casa...
Interpretation test-Version 2
Present tense items

Enero  Febrero  Marzo  Vivo en Inglaterra, por eso...
Octubre  Noviembre  Diciembre

a. ...cómo hace mal tiempo.
b. ...cómo hará mal tiempo.
c. ...cómo haga mal tiempo.
d. ...cómo hacía mal tiempo.

Me quedo en casa...
a. ...cuando tenga sueño.

b. ...cuando tenía sueño.

c. ...cuando tengo sueño.

d. ...cuando tendré sueño.

Bebo café...

cada mes

Enero Febrero Marzo

Octubre Noviembre Diciembre

ensuciarse: to get dirty

Limpio el coche...

a. ...cuando se ensuciará.

b. ...cuando se ensucie.

c. ...cuando se ensució.

d. ...cuando se ensucia.
a. ... cuándo desayuno.

Pongo la radio...

b. ... cuándo desayunaré.

c. ... cuándo desayuné.

d. ... cuándo desayune.

desayunar: to have breakfast

Uso mi teléfono...

a. ... cuando voy a la escuela.

b. ... cuando vaya a la escuela.

c. ... cuando iba a la escuela.

d. ... cuando iré a la escuela.
a. Cuando mi perro reciba un hueso...

b. Cuando mi perro recibe un hueso...

...él salta feliz

c. Cuando mi perro recibía un hueso...

d. Cuando mi perro recibirá un hueso...

Hueso: bone

a. Cuando termine mi partido de tenis...

b. Cuando termino mi partido de tenis...

...me como una manzana.

c. Cuando terminaré mi partido de tenis...

d. Cuando terminaba mi partido de tenis...

partidón: match
a. Cuando iba para la escuela...

b. Cuando iré para la escuela... ...me tomo una manzana.

c. Cuando voy para la escuela...

d. Cuando vaya para la escuela...

a. Cuando iba en el metro...

b. Cuando iré en el metro... ...hojeo el periódico.

c. Cuando vaya en el metro...

d. Cuando voy en el metro... *hojear: to flick through/to glance through*
a. Cuando mi clase de laboratorio acabará...

b. Cuando mi clase de laboratorio acababa...

c. Cuando mi clase de laboratorio acaba...

d. Cuando mi clase de laboratorio acabe...

...como en la cafetería.

...cuando terminaré la cena.

...cuando terminaba la cena.

...cuando termine la cena.

...cuando termino la cena.

Acabo mi tarea...
Vivo en California, por eso...

Surfeo...

a. ...cuando estoy en la playa.
b. ...cuando estaba en la playa.
c. ...cuando esté en la playa.
d. ...cuando estaré en la playa.

Surfar: to surf

Pongo la alarma...

a. ...cuando salgo de casa.
b. ...cuando salí de casa.
c. ...cuando salga de casa.
d. ...cuando saldré de casa.

Poner: to turn on
Hace más frío...

Esconderse: to hide

...cuando el sol se escondía.
...cuando el sol se esconde.
...cuando el sol se esconda.
...cuando el sol se escondrá.

arreglar: to tidy up/ to order
a. Cuando serán las 8 de la mañana...
b. Cuando son las 8 de la mañana...
  ...mi despertador me despierta.
c. Cuando sean las 8 de la mañana...
d. Cuando eran las 8 de la mañana...

a. Cuando esté en mi oficina por la mañana...
b. Cuando estaba en mi oficina por la mañana...
  ...yo como algo rápido.
c. Cuando estaré en mi oficina por la mañana...
d. Cuando estoy en mi oficina por la mañana...
a. Cuando me da hambre...

b. Cuando me dará hambre...

c. Cuando me daba hambre...

...hago algo de comer.

d. Cuando me dé hambre...

dar (a alguien) hambre: to get hungry

---

a. Cuando acabe de lavarme los dientes...

b. Cuando acabará de lavarme los dientes...

c. Cuando acababa de lavarme los dientes...

...me voy a la cama.

d. Cuando acabo de lavarme los dientes...
Subjunctive items

a. ...cuando recibo las medicinas.
b. ...cuando reciba las medicinas.
c. ...cuando recibí las medicinas.
d. ...cuando recibiré las medicinas.

Mejorarse: to get better/to improve

a. ...cuando llego a Japón.
b. ...cuando llegue a Japón.
c. ...cuando llegué a Japón.
d. ...cuando llegaré a Japón.

Mejorar: to get better/to improve

probar: to try

Pruebo el sake...

octubre 2015

diciembre 2015
a. ...cuando el billete de avión sea más barato.
b. ...cuando el billete de avión es más barato.
c. ...cuando el billete de avión era más barato.
d. ...cuando el billete de avión será más barato.

Vuelo a Asia...

octubre 2015

a. ...cuando llegue a Las Vegas.
b. ...cuando llego a Las Vegas.
c. ...cuando llegué a Las Vegas.
d. ...cuando llegaré a Las Vegas.

Llamo a mi hermana...

diciembre 2015

diciembre 2015

Volar: to fly
a. ...cuando recibí mi diploma del MFA.

b. ...cuando recibiré mi diploma del MFA.

c. ...cuando reciba mi diploma del MFA.

d. ...cuando recibo mi diploma del MFA.

Completo mi novela...

2015

2017

a. Cuando conseguí un barco...

b. Cuando conseguiré un barco...

...me voy de viaje por el mar.

c. Cuando consiga un barco...

d. Cuando consigo un barco...

Conseguir: to get
a. ...cuando recibiré mi diploma de chef.

b. ...cuando recibí mi diploma de chef.

C. ...cuando recibo mi diploma de chef.

d. ...cuando reciba mi diploma de chef.

*poner: to open	 propio: own*

...cuando llegaré a la cima de la montaña...

b. Cuando llegué a la cima de la montaña...

c. Cuando llegue a la cima de la montaña...

d. Cuando llego a la cima de la montaña...

* mayo 2016

cima: peak*
a. Cuando termino en la academia...

b. Cuando terminé en la academia... ...me hago bombero.

c. Cuando termine en la academia...

d. Cuando terminaré en la academia...

bombero: fire fighter

a. Cuando llego a la Antártida...

b. Cuando llegué a la Antártida... ...saco una foto a un pingüino.

c. Cuando llegaré a la Antártida...

d. Cuando llegue a la Antártida...
a. ...cuando acabé de comprar.
b. ...cuando acabo de comprar.
c. ...cuando acabe de comprar.
d. ...cuando acabaré de comprar.

a. ...cuando terminé mis estudios.
b. ...cuando termino mis estudios.
c. ...cuando terminaré mis estudios.
d. ...cuando termine mis estudios.

traje: suit
2016

a. Cuando me convertiré en policía...

b. Cuando me convierto en policía...

c. Cuando me convierta en policía...

d. Cuando me convertí en policía...

\textit{convertirse: to become} \quad \textit{detener: to detain/to arrest}

octubre 2015

a. Cuando vaya a la playa...

b. Cuando iba a la playa...

c. Cuando iré a la playa...

d. Cuando voy a la playa...

\textit{bucear: scuba diving}

mayo 2016

...detengo a un ladrón.

...buceo en el mar.
2015

Voy a Italia...

2016

a. ...cuando soy bilingüe.

b. ...cuando seré bilingüe.

c. ...cuando era bilingüe.

d. ...cuando sea bilingüe.

diciembre 2015

a. Cuando hable con el presidente...

b. Cuando hablaré con el presidente....

c. Cuando hablé con el presidente...

d. Cuando hablo con el presidente...

octubre 2015

...telefoneo a mis padres.
 Cuando es otoño...
 Cuando será otoño...
 Cuando sea otoño...
 Cuando era otoño...

 Cuando encontré la receta...
 Cuando encuentre la receta...
 Cuando encuentro la receta...
 Cuando encontraré la receta...
¡FELIZ VERANO!!

mayo 2016

a. Cuando sea fin de curso...

b. Cuando será fin de curso...

c. Cuando es fin de curso...

d. Cuando fue fin de curso...

...me voy de viaje.

octubre 2015

a. ...cuando llegará el Prom.

b. ...cuando llega el Prom.

c. ...cuando llegó el Prom.

d. ...cuando llegue el Prom.

Le pido salir a María...

pedir salir: to ask (somebody) out
Past tense items

a. Cuando tengo el brazo bien...
b. Cuando tenga el brazo bien...   \[\text{...vuelvo a jugar al béisbol}\]
c. Cuando tendré el brazo bien...
d. Cuando tuve el brazo bien...

volver: to come back

...cuando mi madre me compra una bicicleta.
...cuando mi madre me compre una bicicleta.
...cuando mi madre me compró una bicicleta.
...cuando mi madre me comprará una bicicleta.
a. Cuando el profesor escribe en la pizarra...

b. Cuando el profesor escriba en la pizarra...

c. Cuando el profesor escribirá en la pizarra... …yo hacía dibujos.

d. Cuando el profesor escribía en la pizarra...

dibujos: to draw

a. Cuando vuelva de la escuela...

b. Cuando vuelvo de la escuela... …pasaba tiempo con mi hermano.

c. Cuando volvía de la escuela...

d. Cuando volveré de la escuela...
Estuve con un canguro...

a. ...cuando vuelo a Australia.
b. ...cuando volé a Australia.
c. ...cuando vuele a Australia.
d. ...cuando volaré a Australia.
Appendix H: Stimuli of the production test

Production test-Version 1

Present tense items

1. Cada invierno, me pongo un abrigo cuando _________________________(hacer) frío.
2. Cada día me ducho cuando ___________________ (yo, levantarse).
3. Cada día apago mi alarma cuando __________________ (sonar).
4. Cada martes, voy al gimnasio cuando ___________________ (yo, hacer) mi tarea.
5. Cada día hago ejercicios de relajación cuando ___________________ (yo, llegar) a casa.
6. Voy al supermercado todas las semanas y cuando ___________________ (yo, acabar) la compra en el supermercado, pago con mi tarjeta.
7. Cada semana cuando ___________________ (yo, terminar) mi trabajo voy de compras.
8. Cada semana cuando ___________________ (yo, tener) tiempo, hablo con mis padres por teléfono.
9. Todos los días cuando ___________________ (llegar) la mañana, desayuno.
10. Cada noche cuando ___________________ (dar) la una de la mañana, me acuesto.
11. Cada día me lavo los dientes cuando ___________________ (yo, despertarse) por la mañana.
12. Voy al gimnasio regularmente por eso me canso mucho cuando ___________________ (yo, realizar) ejercicios aeróbicos.
13. Cada día duermo una siesta cuando ___________________ (yo, almorzar).
14. Todas las semanas empiezo mi día con energía cuando ___________________ (yo, despertarse).
15. Tengo que comunicarme con mis profesores todos los días, por eso cuando ___________________ (yo, escribir) un email, repaso mi ortografía.
16. Habitualmente cuando ___________________ (yo, querer) algo dulce, compro algo en Georgetown Cupcake.
17. Cuando ___________________ (yo, escuchar) música, me gusta escucharla alta.
18. Cuando ___________________ (yo, leer) un buen libro, suelo escribir anotaciones. [suelo: soler: to usually do something]
19. Cada semana cuando ___________________(yo, limpiar) mi habitación encuentro cosas perdidas.
20. Cada noche cuando ___________________(yo, dormir) 8 horas por la mañana me siento relajada.

**Subjunctive items**

21. Mi cumpleaños es en mayo, por eso voy a un bar cuando ___________________(yo, cumplir) 21 años.
22. Voy a ser abogado y ganar mucho dinero, por eso estoy seguro de que me compro un Ferrari cuando ___________________(yo, tener) suficiente dinero.
23. Tengo una reunión con mi profesor a las 5:00pm. Cuando ___________________(yo, terminar ) de hablar con mi profesor llamo a mis padres.
24. En las próximas vacaciones de acción de gracias quiero darle una sorpresa a mi madre, por eso cuando ___________________(yo, aprender) bien la receta del pavo, lo cocino.
25. El año que viene empiezo una nueva vida en otra ciudad cuando ___________________(yo, terminar) la universidad.
26. Todavía no conozco a mi nuevo sobrino, pero la semana que viene voy a conocerlo. Cuando lo ___________________(yo, conocer), empiezo un álbum de fotos para él.
27. Cuando ___________________( yo, graduarse) me caso con mi novio.
29. Mi ordenador es muy viejo y funciona muy mal. Cuando ___________________(romperse), me compro otro.
30. Estoy muy cansada. Cuando ___________________(yo, terminar) el ensayo me acuesto.
31. Esta noche tengo una cena con amigos. Cuando ___________________(yo, tener) todos los ingredientes la preparo.
32. Me duele mucho la espalda. Cuando ___________________(yo, nadar) un poco en la piscina seguro que me encuentro mejor.
33. He perdido mi cartera. Cuando la ___________________(yo, encontrar) te llamo por teléfono para contártelo.
34. Tengo mucha tarea ahora, pero cuando _______ (yo, acabar) salgo a dar un paseo.
35. Mañana cuando _______ (yo, sacar) el billete me voy a Montevideo.
36. La semana que viene cuando _______ (yo, ir) a casa de mis padres, veo a toda mi familia.
37. ¡María, no puedo hablar contigo ahora porque estoy en la ducha! Cuando _______ (yo, salir) de la ducha te llamo y te cuento los planes para esta noche.
38. Me duele la barriga. Cuando _______ (yo, levantarse) mañana, llamo al médico para pedir una cita.
39. Este verano cuando _______ (yo, volver) a mi casa, me busco un trabajo.
40. Para el año que viene cuando el invierno _______ (llegar), me voy a Florida.

**Past tense ítems**

41. Me compré un apartamento cuando _______ (yo, tener) 25 años.
42. Iba al centro comercial cuando _______ (yo, ser) adolescente.
43. Cuando _______ (yo, terminar) con el primer proyecto, empecé con el segundo.
44. Cuando _______ (yo, saber) las noticias, llamé a mi madre para contárselo.

**Production test-Version 2**

**Present tense ítems**

1. Por las mañanas me despierto cuando el camión de la basura _______ (pasar)…¡a las 6:30! :(
2. Por las tardes almuerzo cuando _______ (yo, salir) de Yates.
3. Por las noches veo la tele cuando _______ (yo, volver) de mi trabajo.
4. En verano me gusta ir a la playa cuando _______ (hacer) calor.
5. En invierno me gusta esquiar cuando _______ (nevar).
6. En otoño me gusta quedarme en casa cuando _______ (llover) mucho.
7. En primavera me gusta pasear por Rock Creek Park cuando el clima _______ (estar) bueno.
8. Entre semana duermo la siesta cuando ____________________ (yo, regresar) de mi clase de español intensivo.

9. Por las mañanas me ducho cuando ____________________ (yo, levantarse) para empezar bien el día.

10. Por las mañanas me gusta hacer mi cama cuando ____________________ (yo, levantarse), así mi habitación está ordenada.

11. Suelo comer chocolate cuando ____________________ (yo, estar) triste. [suelo: soler: to do something regularly]

12. Pedro suele salir a correr cuando ____________________ (tener) mucha tensión.

13. Carlos semanalmente compra flores cuando ____________________ (él, visitar) a su abuela.

14. Por las noches uso un cepillo eléctrico cuando ____________________ (yo, lavarse) los dientes.

15. Los fines de semana me quedo en casa cuando mis amigos ____________________ (jugar) al frisbee. No me gusta jugar al frisbee.

16. Los fines de semana me quedo en casa cuando ____________________ (llover). No me gusta la lluvia.

17. Los viernes voy al cine cuando mis profesores me ____________________ (poner) poca tarea.

18. Los sábados voy al gimnasio cuando ____________________ (yo, estar) muy estresado.

19. Los días entre semana duermo una siesta cuando ____________________ (yo, tener) menos tarea.

20. Los domingos me relajo cuando ____________________ (yo, completar) mi tarea.

**Subjunctive items**

21. Quiero ir al concierto de Bruno Mars en diciembre. Cuando ____________________ (ellos, poner) las entradas a la venta, las compro. [poner entradas a la venta: to put tickets on sale].

22. Para mi clase de historia tengo que hacer una presentación. Cuando ____________________ (yo, acabar) la tarea de Gobierno, empiezo a hacer mi presentación.

23. Voy a ir al cine con mi novia esta noche. Cuando ____________________ (yo, hablar) con ella, decidimos qué película queremos ver.
24. No tengo planes definitivos para este fin de semana. Pero cuando los ________________ (yo, tener), llamo a mi amiga Kara para que venga conmigo.

25. Este semestre pienso mejorar mi dieta. Cuando ________________ (yo, recibir) mi paga, voy a Whole Foods a comprar comida saludable! [paga: allowance/money]

26. Este invierno quiero patinar sobre hielo. Cuando el lago cerca de mi casa ________________ (congelarse), voy con mis amigos! [patinar: to skate; lago: lake; congelarse: to freeze]

27. ¡Me encanta Juego de Tronos y van a sacar una película este inverno! Cuando ________________ (salir) pienso ir a verla. [Juego de Tronos: Game of Thrones]

28. Planeo disfrazarme de zombie en Halloween, pero todavía no sé a qué fiesta voy. Cuando lo ________________ (yo, saber) se lo digo a mi novio (su disfraz es de Rick Grimes :)

29. Este Acción de Gracias cocino yo, pero mis primos también!! Cuando los ________________ (yo, llamar) nos repartimos los platos para no repetir!! [repartir: to divide]

30. ¡Mamá, estoy en clase y no puedo hablar por teléfono ahora! Cuando ________________ (yo, salir) de clase te llamo.

31. Estoy estudiando mandarín este semestre y me gusta mucho, pero todavía tengo mucho que estudiar. Cuando ________________ (yo, aprender) mandarín, quiero buscar trabajo en la embajada china.

32. Estoy estudiando medicina y voy a terminar pronto. Cuando ________________ (yo, convertirse) en médico, abro una consulta privada.

33. Las elecciones 2016 están próximas. Cuando ________________ (llegar) el día de votar, yo voto a…(¡secreto! ¡Es privado!).

34. Mi hermana está a punto de terminar la universidad. Cuando ________________ (ser) la graduación, voy a Washington para verla. (estar a punto de: to be about to).

35. Llevo casi 4 años en Georgetown, mi graduación está ya cerca y tengo muchos planes. Cuando ________________ (yo, graduarse), quiero trabajar en un banco!

36. ¿A qué hora llegas a casa esta noche? No lo sé. Cojo el autobús a las 7, cuando ________________ (yo, salir) de la biblioteca después de la reunión con mi grupo de estudio… posiblemente a las 7:30.
37. ¿Vienes a dar un paseo? Ahora no, gracias. Cuando _________(yo, leer) el último capítulo del libro de Stephen Colbert. ¡Es tan divertido!
38. Cuando en diciembre _____________________(llegar) Navidad, pienso dormir mucho y ver muchas películas. ¡Quiero unas vacaciones!
39. Marta, te llamo cuando _________________________(tú, regresar) de tus vacaciones.
    ¡¡Tengo muchas cosas que contarte…¡¡he conocido a un chico guapísimo!!
40. Alex, voy a tu cuarto cuando ____________________ (yo, comprar) mi computadora nueva
    y así me enseñas la nueva aplicación.

Past tense items
41. Comí mucha pasta cuando ___________ (yo, ir) a Italia.
42. Mi hermano se puso enfermo cuando _________________ (comer) comida en mal estado.
43. Mi amigo lloró mucho cuando _________________ (ver) la película La lista de Schindler.
44. Mis padres subieron a la Torre Eiffel cuando _________________ (estar) en París.
REFERENCES


Cerezo, L. (2012). Beyond hybrid learning: A synthesis of research on e-tutors under the lens of


and research synthesis.


Hsieh, H. C. (2008). "The Effects of Type of Instruction and Type of Post-Exposure Task on L2 Development."


Lado, B. (2008). The role of bilingualism, type of feedback and cognitive capacity in the


Skehan (Eds.), *Individual differences in L2 learning* (pp. 181-209). Amsterdam: John Benjamins.


Ortega, L. (2007). Meaningful L2 practice in foreign language classrooms: A cognitive-
interactionist SLA perspective. In R. M. DeKeyser (Ed.), Practice in a second language:
Perspectives from applied linguistics and cognitive psychology (pp. 180-207). New York:
Cambridge University Press.


matter?* Unpublished doctoral dissertation, Georgetown University, Washington, DC


Second Language Acquisition, 6, 186-214.

*Applied Linguistics*, 10, 217–244.

Py, Bernard. (1999). "Enseignement, apprentissage et simplification de la langue". *In De la
didactique des langues à la didactique du plurilinguisme : Hommage à Louise Dabène,

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral
language abilities: The awareness-practice-feedback sequence. In R. DeKeyser (Ed.),
*Practice in a second language: Perspectives from applied linguistics and cognitive
psychology* (pp. 141-160). Cambridge: Cambridge University Press.


Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on


