TALKING TO AVATARS:
THE COMPUTER AS A TUTOR AND THE INCIDENCE OF
LEARNER’S AGENCY, FEEDBACK, AND GRAMMATICAL FORM IN SLA

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 Linguistics

By

Luis Cerezo Ceballos, M.S.

Washington, DC
December 10, 2010
Increasingly, SLA research is exploring models of hybrid learning where computers are used not only as tools for information and communication purposes but as autonomous electronic tutors. Incipient evidence has shown that e-tutors can facilitate L2 grammar development and even supersede other instructional technologies (e.g., Nagata, 1996; Nutta, 1998; Watts, 1989). However, very few theoretical arguments have been presented to explicate these findings. Quite possibly, the edge of e-tutors may arise from their ability to engage learners in interactive practice (Nagata, 1996), while other multimedia technologies may only, at best, showcase pre-recorded interactions. According to the Interaction Hypothesis (Long, 1996), participation in interactive practice can promote SLA because it allows learners to notice, comprehend, and process input, produce output, form and revise hypotheses in response to feedback, and produce modified output, processes which have been shown to be developmentally helpful (Mackey, 2007). However, the IH does not specify whether or not the effects of these processes are determined by the type of agency of the learner; that is, whether it is necessary for learners to engage proactively in the aforementioned processes, or whether mere exposure to interactive practice by others (by e.g., watching a video) may suffice. On this note, existing empirical research is scant and contradictory, with Mackey (1999)
showing a benefit for practice over exposure to it, Muranoi (2000) showing no significant differences, and Hsieh (2007) yielding mixed results.

Arguably, the relative effects of learner’s agency may be moderated by the type of practice that is being performed or observed. One usual component of practice is corrective feedback, since its presence or absence draws the line between tasks for learning and testing purposes (Loschky & Bley-Vroman, 1993). Hitherto, the facilitative effects of feedback have been largely validated (Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006). However, there is no consensus as to which type of feedback yields the largest effects. These conflicting findings may be explicated by an insufficient isolation of three components of feedback, i.e., positive evidence, negative evidence, and prompting for error repair, so research must address whether they make separate contributions to L2 grammar development (Leeman, 2002; McDonough, 2005). In the realm of CALI this issue remains unaddressed, and additionally, no study to date has explored the potential of oral (as opposed to written) feedback, thus under-exploiting the potential of computers to simulate authentic communication (Warschauer, 2004).

Finally, the effects of both practice and feedback may be mediated by the type of grammatical form under instruction. However, while several studies have corroborated this, there is little research on how the difficulty of a form exerts its mediating role (DeKeyser, 2007; Ellis, 2007).

Using an innovative e-tutor and a pre-test/post-test/delayed post-test design, this study empirically investigated whether intermediate-level university students of L2 Spanish (N=127) learned more grammar when given opportunities to practice or when given audiovisual exposure to others’ practice. Some engaged in computer-simulated
conversations with pre-filmed real people, whereas others observed videos of those interactions. The second variable under investigation was the type of corrective feedback provided, which varied in terms of the explicitness of the negative evidence (whether the error was merely spotted or explained) and the requests for error repair (whether or not students were prompted to reformulate their ill-formed productions). The effects of agency and feedback were investigated on two different grammatical forms varying in complexity, Spanish prepositional relative clauses (easier) and present subjunctive (harder). An exit questionnaire was administered to tap into participants’ perceptions of the learning experience. Results showed that practice had beneficial effects over exposure to it only under more demanding conditions (less explicit feedback) and for one of the targeted structures (prepositional relative clauses). In terms of feedback, prompting yielded no significant effects, while the explicitness of negative evidence did, with more explicit feedback yielding better results for both structures. However, these effects were more evident for the easier form, prepositional relative clauses. In terms of learners’ perceptions, performers were significantly more satisfied than observers, and explicit feedback generated opposite reactions. From a theoretical perspective, these findings contribute to the growing SLA literature on the effects of practice and feedback, while from a pedagogical perspective they prove that computerized oral feedback can work, and help to elucidate precisely when computers may have an edge over non-interactive audiovisual media.
"I am always doing that which I cannot do, in order that I may learn how to do it."

—Pablo Picasso
ACKNOWLEDGMENTS

Tiene que ser en español, las gracias las tengo que dar en español. En primer lugar gracias a mis co-directores, los doctores Ronald P. Leow y Cristina Sanz. Gracias Ron por enseñarme a pensar críticamente, a ser minucioso, a relativizar las cosas, a sonreír ante todo; gracias por tu buen karma, paciencia, cariño y sabiduría —eres un ser de luz—. Gracias Cristina por creer en mí: de no ser por ti simplemente no habría llegado hasta aquí; gracias por tus ideas, por anteponer mis intereses a los tuyos, por tu espíritu de trabajo, fuerza, valentía y sobre todo, por tu sensibilidad y empatía. Gracias asimismo a los miembros de mi tribunal. Al Dr. Robert DeKeyser por los comentarios críticos y por estar siempre disponible en esta vertiginosa carrera de obstáculos, y al Dr. Rusan Chen por el buen humor durante tantas horas entre ANOVAs y Scheffés.

Mi más profundo agradecimiento a todos los que habéis colaborado de manera activa en esta tesis. A mi hermano Antonio, por ayudarme a materializar el software de Talking to Avatars. A mi padre José Luis y mi cuñada Pepa, por meteros con tanta gracia en la piel de mis avatares. A Kasey Dixon, por ayudarme a traducir, a contactar a los participantes y recoger sus datos. A Mika Hama, por ser mi otro yo en el laboratorio. A Rebecca Sachs, por ser más sabia y generosa que Wikipedia. A Grant Armstrong, por ayudarme a encontrar items. A Missy Baralt, por los MLATs y el apoyo logístico. A Jessie Cox, por recodificar mis datos. A Kara Morgan-Short y Jaemyung Goo, por los consejos sobre estadísticas. A Susan Pennestri, por ser la reina de Blackboard. Y a los participantes en mi estudio, por el tiempo dedicado y los mensajes de ánimo.
Gracias también a los profesores que me habéis inspirado y ayudado todos estos años. Especialmente, a la Dra. Gloria Corpas de la Universidad de Málaga, que fue la primera profesora que creyó en mí; al Dr. Héctor Campos y al Dr. Alfonso Morales, por su ayuda y cariño en momentos cruciales; y a María Wear, que reclutó a mis participantes como si el proyecto fuera suyo.

Un fuerte abrazo para todos los amigos que, desde que llegué a Georgetown en 2002, habéis sido mi segunda familia: Carolina León, John Crosby, Ana Carolina Rodríguez, Daniel Giglio, Mika Hama, Rebecca Sachs, Justin Kelly, Patricia Soler, Guillermo Villanueva, Kasey Dixon, Jacopo Meneguzzo, Laura Vázquez, Daniel Loyacano, Nina Moreno, Maite Mascaró, Juan Rojo, Clara Pascual, Antonio Mayorga, Susan Pennestri, Álex Rojas, Gorky Cruz, Paulina Maldonado, Paula Plaza, Missy Baralt, Verónica Colón y Paola Piscioneri, entre otros. Gracias también a mis amigos del otro lado del charco, en especial Federico Gaspari y Maite Molina. Y gracias a Bea, por las sonrisas, la voz, los oídos, el ojo multicolor, la poesía, y por permanecer a mi lado cuando no había luz. No habría podido sin ti, y sigues aquí, en tantas pequeñas cosas.

Por último, gracias, con todo mi corazón, a mi familia, por el apoyo y el amor incondicionales, por alegraros y sufrir conmigo. Esta tesis está dedicada a vosotros. Gracias papá y mamá, porque lo que sabemos Antonio, Maole y yo de honestidad, tesón y creatividad nos los habéis enseñado vosotros. Gracias, Antonio, por volar sin avisar, por escucharme, y por ayudarme a volver a empezar. Gracias Pepa, porque eres pequeña pero muy grande. Gracias Maole, Iago e Iván, por las canciones por Skype. Y gracias también a mis tíos Antonio, Aurora, Marisol y Salvador, y a las que no estáis pero sí estás, mi tía María Luisa y mi abuela Lola. Espero que os gustara la defensa.
DEDICATION

To my family
TABLE OF CONTENTS

CHAPTER 1. STATEMENT OF THE PROBLEM

CHAPTER 2. COMPUTER-ASSISTED LANGUAGE LEARNING

1. CALL

1.1 The “if CALL” question: Can software facilitate L2 development?

1.2 The “which CALL” question: Which software is more beneficial for L2 development?

2. CALI

2.1 The “if CALI” question: Can computerized tutors facilitate L2 development versus other technologies?

3. Mediating factors in L2 instruction

3.1 Type of grammatical form

CHAPTER SUMMARY

CHAPTER 3. THEORETICAL CONSIDERATIONS

1. Learner’s Agency

1.1 Practice

1.1.1 Toward a definition of practice

1.1.2 Incidental vs. deliberate practice

1.1.3 Meaning vs. form-focused practice

1.1.4 Designing deliberate practice

1.1.5 Receptive vs. productive practice

1.2 Practice into practice: Tasks, activities, exercises, and drills

1.2.1 Tasks

1.2.2 Activities, exercises, and drills

1.2 Exposure

Section summary: Learner’s Agency

2. Feedback

2.1 The place of feedback in SLA

2.2 Defining corrective feedback and its components

2.2.1 The explicitness of negative evidence

2.2.1.1 Error signaling

2.2.1.2 Three points in a continuum: recasts, negotiation, explicit feedback

2.2.1.3 Two sides of explicitness: the tutor and the learner

2.2.1.4 An integrative theoretical framework to measure feedback explicitness

2.2.1.5 Re-evaluating the implicitness of recasts

Re-evaluating the implicitness of recasts
CHAPTER 4. REVIEW OF THE LITERATURE .............................................................. 169

1. LEARNER’S AGENCY ....................................................................................... 169
   1.1 Practice ........................................................................................................ 170
      1.1.1 Krashen’s Input Hypothesis ................................................................ 170
      1.1.2 The Canadian French immersion programs ..................................... 171
      1.1.3 White’s Incomprehensible Input Hypothesis ..................................... 172
      1.1.4 Processing Instruction ................................................................. 174
      1.1.5 The Output Hypothesis .................................................................... 177
      1.1.6 The Interaction Hypothesis ......................................................... 186
   1.2 Exposure to practice .................................................................................. 190
      1.2.1 Exposure to practice in oral interaction ........................................... 193
         Classroom contexts ................................................................................. 193
         Laboratory contexts .............................................................................. 197
      1.2.2 Exposure to practice in CALI ....................................................... 200
      Section summary: Learner’s agency ....................................................... 202

2. FEEDBACK ..................................................................................................... 205
   2.1 Feedback in SLA ....................................................................................... 206
   2.2 Feedback in CALI .................................................................................... 213
      2.2.1 Re-organizing feedback types in the CALI literature ................... 217
      2.2.2 Empirical findings on feedback in CALI ....................................... 227
      Section summary: Feedback ................................................................. 232

3. GRAMMATICAL FORM .............................................................................. 236
   3.1 Rules in the input .................................................................................... 237
   3.2 Rules in the feedback ............................................................................ 244
      Summary: Grammatical form ............................................................... 251

4. RESEARCH QUESTIONS ............................................................................... 257

CHAPTER 5. THE EMPIRICAL STUDY ................................................................. 261

1. PARTICIPANTS AND EXPERIMENTAL GROUPS ........................................ 261
2. TARGETED FORMS ..................................................................................... 264
   2.1 Spanish prepositional relative clauses ................................................. 264
   2.2 Spanish present subjunctive in adjectival relative clauses .................. 267
   2.3 Present subjunctive vs. Prepositional phrases in adjectival relative clauses .................................................................................. 273

3. MATERIALS ............................................................................................... 274
   3.1 CALI tutor ............................................................................................... 275
   3.2 Treatment practice ................................................................................ 278
   3.3 Treatment feedback ............................................................................. 287
   3.4 Assessment measures .......................................................................... 293

4. PROCEDURE ............................................................................................... 299

5. CODING AND SCORING ........................................................................... 300

6. PROPOSED STATISTICAL ANALYSIS ...................................................... 304

CHAPTER 6. RESULTS ...................................................................................... 309

1. TEST RELIABILITY ANALYSES ................................................................. 310
2. DESCRIPTIVE STATISTICS ....................................................................... 313
   2.1 L2 accuracy ......................................................................................... 313
2.2 Time on task .......................................................... 315
2.3 Attention to practice .................................................. 317
3. RESEARCH QUESTIONS .................................................. 318
  3.1 Research question #1. Type of learner’s Agency and L2 development ........ 318
      3.1.1 Prepositional relative clauses: Spot & Prompt treatment ................. 324
      3.1.2 Prepositional relative clauses: Explain & Prompt treatment ............. 330
      3.1.3 Prepositional relative clauses: No Feedback treatment ................... 336
      3.1.4 Present subjunctive: Spot & Prompt treatment ............................. 339
      3.1.5 Present subjunctive: Explain & Prompt treatment .......................... 343
      3.1.6 Present subjunctive: No Feedback treatment ............................... 347
      Summary of results for RQ1 .......................................................... 351
  3.2 Research question #2. Type of Feedback and L2 development .................. 355
      3.2.1 Prepositional relative clauses ................................................. 365
      4.2.2 Present subjunctive .......................................................... 376
      Summary of results for RQ2 .......................................................... 383
  3.3 Research question #3. Type of grammatical form and L2 development ......... 385
      Summary of results for RQ3 .......................................................... 402
  3.4 Research question #4. Interactions between Agency, Feedback, Form, and Time ...... 405
4. EXIT QUESTIONNAIRE .................................................. 408

CHAPTER 7. DISCUSSION .................................................. 416
1. RESEARCH QUESTION #1. TYPE OF LEARNER’S AGENCY AND L2 DEVELOPMENT ........ 416
   Agency in Spot & Prompt treatments ............................................. 417
   Agency in Explain & Prompt treatments ......................................... 422
   Agency in No Feedback treatments ............................................. 426
2. RESEARCH QUESTION #2. TYPE OF FEEDBACK AND L2 DEVELOPMENT ............ 434
   Prepositional relative clauses .................................................... 435
   Present subjunctive ............................................................... 447
3. RESEARCH QUESTION #3. TYPE OF GRAMMATICAL FORM AND L2 DEVELOPMENT .... 448
4. RESEARCH QUESTION #4. INTERACTIONS BETWEEN AGENCY, FEEDBACK, FORM, AND TIME ...... 452
5. CONCLUSION AND IMPLICATIONS ........................................ 457
6. LIMITATIONS AND FURTHER RESEARCH ....................................... 459

APPENDIX 3.1 .............................................................. 463
APPENDIX 3.2 .............................................................. 464
APPENDIX 4.1 .............................................................. 468
APPENDIX 4.2 .............................................................. 472
APPENDIX 5.1 A .............................................................. 477
APPENDIX 5.1 B .............................................................. 494
APPENDIX 5.2 A .............................................................. 512
APPENDIX 5.2 B .............................................................. 517
APPENDIX 5.3 A .............................................................. 522
APPENDIX 5.3 B .............................................................. 522
APPENDIX 5.4 .............................................................. 540
APPENDIX 5.5 .............................................................. 555
APPENDIX 5.6 .............................................................. 558
APPENDIX 6.1 .............................................................. 567
APPENDIX 6.2 .............................................................. 575
APPENDIX 6.3 .............................................................. 576
<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4</td>
<td>578</td>
</tr>
<tr>
<td>6.5</td>
<td>579</td>
</tr>
<tr>
<td>6.6</td>
<td>580</td>
</tr>
<tr>
<td>6.7</td>
<td>581</td>
</tr>
<tr>
<td>6.8</td>
<td>582</td>
</tr>
<tr>
<td>6.9</td>
<td>583</td>
</tr>
<tr>
<td>6.10</td>
<td>584</td>
</tr>
<tr>
<td>6.11</td>
<td>585</td>
</tr>
<tr>
<td>6.12</td>
<td>586</td>
</tr>
<tr>
<td>6.13</td>
<td>587</td>
</tr>
<tr>
<td>6.14</td>
<td>588</td>
</tr>
<tr>
<td>6.15</td>
<td>589</td>
</tr>
<tr>
<td>6.16</td>
<td>590</td>
</tr>
<tr>
<td>6.17</td>
<td>591</td>
</tr>
<tr>
<td>6.18</td>
<td>592</td>
</tr>
<tr>
<td>6.19</td>
<td>593</td>
</tr>
<tr>
<td>6.20</td>
<td>594</td>
</tr>
<tr>
<td>6.21</td>
<td>595</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>597</td>
</tr>
</tbody>
</table>
CHAPTER 1.

STATEMENT OF THE PROBLEM

The advent of computers in the past century revolutionized the world as we saw it, redefining the way we process, produce, and disseminate information. A new time, the so-called “information era,” sparked, and the “language industry” was not left untouched. For a few decades now, both research and non-research institutions have been developing a myriad of computer applications for language-related purposes, ranging from translation to language learning and testing. In this context, evaluations are key to make informed decisions about language software development, purchasing, and implementation.

In the realm of Second Language Acquisition (SLA) the facilitative role of computers in L2 development in general, and grammar in particular, has been largely validated (e.g., Chapelle, 1997; Felix, 2005; Grgurovic, 2007; Salaberry, 2001; Zhao, 2003). However, given the diversity of available electronic materials, research must be geared around a clear-cut taxonomy of resources to make well-grounded comparisons and extrapolate findings. A first distinction must be made between CALL (Computer Assisted Language Learning) and CALI (Computer Assisted Language Instruction), depending on whether the computer is used as a tool for information and communication purposes, or as a substitute for other instructional technologies (e.g., instruction with teachers, workbooks, or audiovisual materials). To date, there is a clear imbalance in the amount of research conducted under each strand. Research on CALL, particularly in the realm of CMC
(Computer-Mediated Communication) has grown exponentially, showing that computers can be effectively used to facilitate communication among peers (and teachers) and to promote L2 development at least as equally well as FTF (Face To Face) interaction. However, very few studies have explored CALI, i.e., the potential of computers as tutors in and by themselves.

So far, the incipient literature on CALI has shown that computerized tutors can not only facilitate L2 grammar development but even supersede other instructional technologies based on VCRs, workbooks, and human teachers under certain conditions (e.g., Nagata, 1996; Nutta, 1998; Watts, 1989). However, very few arguments have been presented to explicate these findings based on a theoretical foundation. Quite possibly, as mentioned by Nagata (1996), the edge of CALI against other instructional media may arise from the capability of computers to engage learners in large amounts of interactive practice. Indeed, as shown over three decades of empirical work under the Interaction Approach (Gass & Mackey, 2006), participation in interactive practice can promote SLA significantly better than exposure to L2 input alone (e.g., Norris & Ortega, 2000) because it allows learners to comprehend and notice the input, produce output, and, in response to feedback, revise hypotheses and produce modified output, processes all which have been proved to be developmentally helpful (Long, 1996). Therefore, in light of this it could be argued that computerized tutors can promote L2 development better than other technologies because they can engage the learner in a wider array of processes other than the mere exposure to L2 input.

To the present day, however, the term “exposure” has been understood by most researchers in a rather eschewed way, i.e., the Krashenian way, where it means
comprehending and/or processing positive evidence, i.e., input that is correct in the L2. Yet, as pointed out by Mackey (1999), in many classroom settings and naturalistic environments learners are exposed to more than the input addressed to them, including the output of other learners, the feedback that these receive in return, and their potential output modifications. On this token, the Interaction Hypothesis makes no claims as to the role of the agency of the learner in the communication process, that is, whether acting as an “interactant” or “observer” mediates the effects of interaction and its quintessential components, input, output, and feedback. Filling in this theoretical gap is important for pedagogical reasons both in the classroom and CALI environments. In many classroom contexts, especially since the formulation of the Output and Interaction Hypotheses, teachers are instructed to push students to assume an active role in communication, sometimes even penalizing low participation rates. However, as posed by Muranoi (2000), oftentimes teachers can only give feedback to a limited number of students, and as argued by Ellis, Tanaka, and Yamazaki (1994) there are many students who prefer a “quieter route” and can nevertheless equally benefit from watching the interactional activity of their “more public comrades.” In turn, producing CALI materials is a rather costly endeavor because it requires using computationally complex tools that are able to analyze the student’s response and react accordingly. On the contrary, producing non-interactive materials such as filmed interaction episodes is more affordable, but so far even big-budget audiovisual language series provide a very limited exposure experience. Typically, they only display positive evidence and occasionally metalinguistic information, but very rarely, if ever at all, do they show episodes of feedback in response to L2 errors or output modification. Therefore, research must address whether the
benefits of CALI by itself and against other technologies are due to its capability to engage learners in interactive practice or due to a poor use of less costly, non-interactive media (see, e.g., Stockwell, 2007; C. Williams & Brown, 1991).

Hitherto, empirical literature observing the effects of learner’s agency (i.e., the compared effects of practice vs. exposure to practice) on L2 grammar development is scant and contradictory. To my knowledge, only three studies, two of them in oral interaction (Mackey, 1999; Muranoi, 2000) and one in CALI (Hsieh, 2007), have experimentally isolated two conditions to compare the effects of practice versus exposure to practice in L2 grammar development, showing either a benefit for practice (Mackey, 1999), no significant difference (Muranoi, 2000), or mixed findings (Hsieh, 2007). Clearly, further research in this arena is warranted.

On a different note, observing the effects of learner’s agency in interactive practice requires an understanding of the practice that is being performed or observed. Yet, as DeKeyser (2007: 1) points out, “[t]he concept of second language practice remains remarkably unexamined from a theoretical point of view,” often generating “[m]isgivings and misunderstandings.” Therefore, defining practice is a necessary prerequisite to interpret the existing empirical literature on this construct. Additionally, as pointed out by Hsieh (2007), most studies have failed to isolate practice from corrective feedback, and thus it is impossible to determine whether any observed developmental effects are due to the practice, the feedback, or a combination of both. Consequently, studies addressing the role of practice and/or exposure to practice should incorporate a group without feedback into their research designs.
As just mentioned, the constructs of practice and feedback are closely intertwined. Indeed, as posed by Loschky and Bley-Vroman (1993) the absence or presence of feedback is what typically distinguishes the main purpose of a task, i.e., learning or testing. Arguably, the ability to analyze the learner’s response and react accordingly via different interactional adjustments, such as modified input or corrective feedback, draws the line between computerized tutors and other non-interactive technologies. In recent years, a growing number of empirical studies have examined the role of computerized feedback. However, although its pedagogic value is largely validated (see the recent meta-analyses by Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006), there is no consensus yet as to the relative effects of different types of feedback, or even the same feedback type under different learning conditions. This conflict may be explicated by factors both intrinsic and extrinsic to the construct of feedback.

In relation to intrinsic factors, existing empirical studies show great differences in the way that researchers have defined, provided, or coded feedback types (e.g., Nassaji, 2007: 520; Sheen, 2006: 362). Additionally, researchers have almost exclusively focused on the explicitness of feedback, understood as the degree of detail provided to explain an error, but in so doing they have often neglected the multi-dimensional nature of feedback, where three components can work in concert or independently. These components are the provision of negative evidence (that can be more or less explicit, explaining the nature of the error or not), the provision of positive evidence (Leeman, 2000), and prompting for error repair (Ranta & Lyster, 2007). Consequently, when comparing empirical studies on feedback one must make sure that these specific components are properly isolated in
order to gauge their relative contributions. In the realm of CALI, no study has examined
the role of positive evidence in corrective feedback, so research avenues are open in this
regard. More importantly, all studies to date have conflated negative evidence with
prompting for error repair, so teasing out these two variables is necessary to build upon
the existing literature. Finally, all CALI studies have focused on written feedback
exclusively, thus underexploiting the potential of the computer to simulate authentic
communication (Underwood, 1993; Warschauer, 2004) and hindering comparisons with
the vast body of literature on feedback in oral interaction. Consequently, then, future
research must address the pedagogic potential of oral feedback in CALI.

With regards to extrinsic factors, as pointed out by a number of authors, many studies
have also failed to properly isolate feedback from other pedagogical interventions such as
prior metalinguistic instruction, which makes difficult to interpret their relative
contributions to learning (see Sanz & Morgan-Short, 2005 for more details).
Additionally, a number of possibly intervening variables have been overlooked. These
include the method used to expose participants to the target form, i.e., exposure to input
vs. practice, as well as the practice mode, i.e., input- or output-focused (more in Sanz &
Morgan-Short, 2005); the task-features (Gass, Mackey, & Ross-Feldman, 2005); the
learning context, i.e., classroom vs. laboratory settings (Gass et al., 2005); the learner’s
individual differences, such as memory, aptitude, motivation, and developmental level
(e.g., Ammar, 2008; Ammar & Spada, 2006; Mackey & Philp, 1998), and the type of
targeted form.

Precisely this latter variable, type of linguistic form, with a focus on grammar, has
been increasingly addressed in the feedback literature (e.g., Ayoun, 2001; de Graaff,
Overall, these studies have shown that the effects of feedback can indeed be moderated by the type of linguistic form under instruction. However, there is great disparity in the claims that authors have made. For example, Nagata (1993) argued that the advantages of more explicit types of feedback arise for more complex forms, since for simpler forms minimally explicit feedback may suffice. Conversely, Ammar (2008) and Ammar and Spada (2006), drawing on cognitive psychology (e.g., Reber, 1993), argued that implicit learning is more likely to show advantages for complex structures. Consequently, then, as Ellis (2007: 360) puts it, “[w]hat is needed is further research to help us identify how linguistic factors determine when different kinds of feedback will work for acquisition.” From a theoretical point of view, though, this is not an easy task. Researchers have approached the construct of linguistic form from many perspectives, namely the L1, the L2, and the interlingua, contributing to a plethora of terms, e.g., developmental readiness, difficulty, learnability, fossilization, or ultimate attainment (DeKeyser, 2005). Therefore, examining the literature on type of linguistic form requires a previous understanding of a number of related theoretical constructs.

To sum up, then, the present dissertation seeks to cast some light on whether and how computerized tutors can promote L2 grammar development by taking into consideration the kind of interactive experience they can provide. Specifically, by addressing a gap in the Interaction Hypothesis, i.e., the role of learner’s agency in the interaction process, it investigates if the edge of computerized tutors is determined by their ability to engage learners in an interactive practice experience, or if on the contrary mere exposure to pre-filmed practice performed by other interactors may yield comparable learning.
Additionally, this dissertation is the first study to examine the contribution of computerized oral feedback per se and in terms of two fundamental components, explicitness of negative evidence and prompting for error repair. Finally, the possibly intervening role of the type of grammatical form is examined, in an attempt to uncover how it may moderate the efficacy of practice and feedback.

The structure of this dissertation is as follows. Chapter 2 provides a bird’s eye synthesis of the CALL literature to date, focusing on the under-researched area of CALI and identifying the three variables under investigation here: Learner’s Agency, Feedback, and linguistic Form, with a focus on grammar. Chapter 3 provides a necessary theoretical discussion on these constructs, identifying key issues and proposing classification criteria. Based on these theoretical considerations, Chapter 4 provides a review of the empirical literature on each variable, formulating the research questions. Chapter 5 lays out the research design and methodology of the empirical study. Chapter 6 presents the results of the statistical analyses conducted. Finally, Chapter 7 contains a discussion of the results, the study’s limitations, ideas for further research, and a conclusion.
In the realm of Second Language Acquisition (SLA) computers have often been presented as an aid or even a substitute for traditional learning environments such as classroom- or laboratory-based instruction, for various reasons. For example, “human” teachers, as argued by, e.g., Van den Branden (1997), have different levels of knowledge, expertise, and pedagogic preferences. Some may ignore the full range of pedagogical interventions, such as feedback types, or have an uninformed bias for one over the other, or interact more with certain students, or behave inconsistently. In turn, students present a number of individual differences in terms of cognitive capacity, language experience, motivation, learning style, and personality, and thus may benefit most from customized approaches that take these into consideration. Additionally, the face-to-face relationship between the teacher and the student often translates in increased levels of anxiety. Consequently, in classroom interaction, depending on the specific characteristics of the teacher, the learner, and the environment, students may not receive equal or suitable practice and feedback opportunities. In contrast to classroom instruction, laboratory-based instruction using text, audio, and visual materials may reduce the face-to-face anxiety factor, but the inability of such materials to process the student’s behavior and react to it accordingly reduces the quantity and quality of the interaction opportunities that have been proven to promote SLA (see, e.g., Gass & Mackey, 2006). In this context, computer-based instruction may offer the best of two worlds. Specifically, as Kim, Cowan and Choo (2006) summarize, computers may offer more opportunities for
promoting SLA behavior that enhances learning (Beatty, 2003; Chapelle, 2005), increase the consistency of teacher-delivered pedagogical interventions (Tsutsui, 2004), and reduce anxiety (Torlaković & Deugo, 2004). In this context, empirical investigations involving measures of L2 development are clearly needed in order to determine whether this claimed edge of computers is hype or reality, either independently or against other instructional technologies.

Research on computer-based SLA has been growing at a fast pace over the last decades. As a result, many new labels have sprouted, such as Computer Assisted Language Instruction (CALI), Computer Assisted Language Learning (CALL), intelligent Computer Assisted Language Learning (iCALL), or Computer-Mediated Communication (CMC), to name but a few. All these strands of research study the role of computers in SLA, but from very different angles that crucially need to be distinguished. The present chapter provides a comprehensive status quo of the use of, and research on, computers in SLA. Section 1 analyzes the role of software for language learning in general, where computers are used for a wide variety of purposes, including, for example, as a source of information or a means of communication. This is what researchers usually refer to under the generic terms CALL (Computer Assisted Language Learning) or CASLA (Computer Applications in SLA - Chapelle, 2001). In contrast, section 2 specifically focuses on tutoring applications, where computers seek to assume the role of the teacher. This is often referred to as CALI (Computer Assisted Language Instruction). In both sections, CALL and CALI, the discussion is geared around two pivotal questions: the “if” question (i.e., if software can facilitate L2 development, per se and versus other instructional technologies) and the “which” question (i.e., which software applications
are more beneficial for learning purposes). In dealing with these questions for CALL in sections 1.1 and 1.2, a revision of the history of CALL is presented (1.2.1), followed by a proposal for a taxonomy of CALL resources (1.2.2), and an account of how empirical literature has researched these resources (1.2.3). In this latter section a dearth of research on the particular sub-area of CALI is detected, thus opening section 2. Section 2.1 deals with the “if CALI” question by reviewing the studies that have compared CALI to other instructional technologies, subsequently analyzing the possible causes of its declared success. It is posited here that what distinguishes computerized tutors viz à viz other technologies is their ability to engage learners in large amounts of interactive practice, and thus the empirical findings on participation in practice versus no participation in it (i.e., exposure to practice by others) are briefly introduced. In turn, in answering the “which CALI” question, section 2.2 analyzes the particular features that define the interactive experience that CALI may provide, i.e., the type of practice and feedback, with a special focus on the latter. Next, after dealing with the “if” and “which” questions in CALI, section 3 zooms in on the “what,” specifically, how the linguistic form under instruction may mediate the effects of a learning environment. This chapter closes with a comprehensive summary, thus setting the ground for chapter 3, which provides some necessary theoretical considerations about the constructs investigated in this dissertation: Learner’s Agency (i.e., Practice vs. Exposure to practice), Feedback, and grammatical Form in CALI.
1. CALL

From the early days of the PLATO Project, initiated at the University of Illinois in 1960, to the booming sales of The Rosetta Stone, arguably today’s most popular CALI system, the development of computer-based second language pedagogical materials has grown exponentially, but as argued by Hulstijn (2000: 35), academic research has not always kept up. In this context, a critical look to the past and present history of computer-based SLA is mandatory, in order to determine what has been done and what remains to be done, and most importantly, if and which software can aid learners acquire a second language. The next two sections seek to answer these questions separately.

1.1 The “if CALL” question: Can software facilitate L2 development?

The question of whether software can facilitate L2 development has been a matter of debate since the advent of the first computer-based language learning systems in the 1960s, and the answer has been gradually changing from “no” to “yes” over the years. For example, in the 1980s Higgins and Johns (1984: 86) concluded:

We do not know whether the computer is a suitable supplement or aid to the vast bulk of the work done in the classrooms or by learners on their own. We do not know whether all students should be urged to spend time on CALL activities, or whether it should remain a voluntary activity. We do not know whether CALL has equal value for students at all levels, or whether it favours one particular level.

In contrast, more recent reviews and meta-analyses on CALL research (e.g., Chapelle, 1997; Felix, 2005; Grgurovic, 2007; Salaberry, 2001; Zhao, 2003) have shown that, although many questions remain to be answered, computers can indeed prove beneficial for SLA. Thus, as argued by Zhao’s (2003: 3) meta-analysis of empirical studies
“technology-supported language learning” (where technology refers mainly to computer-based materials – instruction programs, the web, computer-mediated communication, simulation, speech technologies, word processing, e-books, and grammar checkers – but also audio and video) has proved to be “at least as effective as human teachers, if not more so.”

There is, at present, increasing evidence to believe that computers can aid or, in some controlled cases, even replace or supersede human teachers in certain aspects of L2 learning (e.g., Nutta, 1998). However, the term CALL subsumes many different types of computer applications with very possibly different degrees of learning success. It is therefore mandatory to delve deeper into the different subtypes of CALL applications to gauge their relative contributions.

1.2 The “which CALL” question: Which software is more beneficial for L2 development?

The question of which software is more beneficial for L2 development is a rather complex one. In order to answer it, three necessary steps must be taken. First, one should make a journey through the history of CALL to gain an insight into the different types of CALL resources that have been developed. Following that historical account, a classification or taxonomy should be created with the aim of grouping up these resources into a finite number of categories. And third, experimental research should be reviewed in order to determine which CALL categories have been empirically tested and which have not, whether or not they have been found to promote L2 development, and if so, which resources have proven more beneficial than others.
1.2.1 A thumbnail history of CALL

Hitherto, several historical accounts of CALL have been provided, although as posed by Bax (2003: 14) “there exists no in-depth analysis of the history of CALL along the lines of, say Howatt’s history of English Language Teaching (1984).” Depending on their orientation, existing accounts can be divided into two main kinds: comprehensive reviews (lists of resources) and analytic discussions (theoretically-oriented classifications of the resources). According to Delcloque (2000, apud Bax, 2003), the first group (reviews) is represented by Levy (1997) and Ahmad, Corbett, Rodgers, and Sussex (1985) —and very possibly by Chapelle (2001) and Sanders (1995), too— while an example of the second group (discussions) is found in Warschauer and Healey (1998) —as well as later work by Warschauer and colleagues, and Bax (2003). To this regard, CALL researchers are divided as to the particular usefulness of each approach. Thus, Delcloque refers to reviews and discussions as “objective” vs. “subjective” historical accounts, respectively, arguing in favor of the former. From the opposite standpoint, Bax (2003: 14) contends that what the field really needs is more analysis:

If we are to use our past experience not only to know what has happened in CALL but also to understand it, then any history surely needs to be more than a mere recounting of supposed facts concerning software and hardware developments. It needs to offer analysis of what occurred and also why, particularly with regard to the actual use of computers in language teaching and learning.

It is beyond the scope of the present dissertation to produce a catalogue of all the CALL applications that have been created. It is my stand, however, that the field is in need of a clear-cut taxonomy that allows us to organize resources into well-defined
categories, in order to identify which specific subtypes of CALL applications yield better results. To this aim, a historical analysis of CALL developments may be a good start. The following pages summarize the analytic discussion of the history of CALL presented by Warschauer (Kern & Warschauer, 2000; Warschauer, 1996, 2004, 2007; Warschauer & Healey, 1998) and its critical revisitation by Bax (2003). Drawing on this discussion, I will argue that these analyses, whilst contributing to a good overall picture of CALL technology, lack specificity and fail to narrowly discriminate between different types of CALL resources, which will lead to the next section, A usable taxonomy of CALL resources.

Warschauer’s account

Warschauer’s historical account of CALL has undergone various redefinitions in terms of the labeling, origins, and even descriptions of historical stages. However, all his accounts are consistent in distinguishing three main stages, each corresponding to a certain level of technological development, a particular pedagogical approach, and a different perception of the role of the computer in the classroom, along the lines of the trichotomy proposed by Crook (1996), i.e., “computer-as-tutor,” “computer-as-pupil,” and “computer-as-tool.” What follows is a comprehensive summary of Warschauer’s account, primarily based on Warschauer (2004), with additions from Warschauer and Healey (1998) and Kern and Warschauer (2000).

The earliest-appearing batch of electronic language resources is referred to by Warschauer under the umbrella term “Structural CALL.” These resources, which appeared during the 1970s and 1980s, consisted of grammar and vocabulary tutorials,
drill and practice programs, and language testing instruments. According to Warschauer, this “drill-and-kill” approach was the result of both the technological constraints of the moment (computers were costly, hard-to-access machines or mainframes, with limited capabilities) and the adoption of structuralist or behaviorist language teaching approaches (the grammar-translation and audiolingual methods, according to which repeated drilling on the same material is essential to learning). In this paradigm, the computer was viewed as a “mechanical tutor” that generally allows only one acceptable response per item, provides immediate positive or negative feedback on the formal accuracy of the learner’s responses (right vs. wrong), and in so doing never grows tired or judgmental, allowing students to work at an individual pace.

In the 1980s and 1990s, the development of more sophisticated computers (mainstreams were replaced by PCs) and the rejection of purely behavioristic approaches to language learning both propelled CALL into its second generation. “Communicative CALL” replaced the drill-and-kill approach with communicative exercises where the learner was taking part in an interaction, thereby using forms rather than studying them, acquiring grammar implicitly rather than explicitly, and using the target language predominantly or even exclusively. In this paradigm, the computer was seen not so much as a tutor but rather as a peer or a “pupil,” as a party in the interaction. However, from the point of view of the student, these interactions were very limited and primarily input-based. The learner’s output bore little importance, as did the content or subject matter.

In the late 1990s, the popularization of computer networks (especially, the Internet) brought about a new phase in CALL: “integrative CALL.” At that time, many language teachers were abandoning communicative teaching in favor of a more social or socio-
cognitive view, where the goal was not only using language but using language in authentic social contexts. Task-based, project-based, and content-based approaches all sought to integrate learners in authentic environments while promoting practice in every language skill (listening, speaking, reading, and writing). Under this paradigm, the computer was seen neither as a “tutor” nor a “pupil” but a “tool” that students may use to solve real-life problems, one which is fully integrated into the classroom environment. Integrative CALL is thus the basis of what Kern and Warschauer (2000) refer to as Network-Based Language Teaching (NBLT), involving the use of Computer-Mediated Communication (CMC)—both synchronous (i.e., in real time, via e.g., chat software) and asynchronous (i.e., not simultaneous via, e.g., e-mail or electronic forums)—and the use of the Internet for collaborative real-life task-based, project-based, and content-based activities (see the recently coined term PrOCALL, or Project-Oriented CALL, by Debski, 2000).

Finally, according to Warschauer and Healey (1998) the history of CALL has yet to see another stage, “intelligent CALL,” where “the idea is to have software that uses the power of the computer to offer easy interaction with the material to be learned, including meaningful feedback and guidance; comprehensible information in multiple media designed to fit the learning style of individual students; and ways for students to carry communication beyond an individual computer screen.” Intelligent CALL programs will thus track learner answers and look for patterns, responding not only with information about the correctness of the learners’ answers but also about the sources of their errors, and offering suggestions for future study. Intelligent CALL software will yield agency to the users, giving them the choice of what to do, when to do it, and for how long, and
limiting itself to offering guidance or making suggestions. The software will fit the medium to the learner, harmonically using different media (graphics, sound, animation, and video clips) to enhance understanding. Finally, it will incorporate high-quality Natural Language Processing (NLP) in order to be able to respond intelligently to the learners’ input. According to Warschauer and Healey, “[w]hile we’re coming closer, advancing in all the areas mentioned above — user interface, learner feedback, integrated media, communication within and outside the classroom, and natural language processing— we’ve still got a very long way to go before CALL can be accurately called ‘intelligent,’” a point reiterated by Kern and Warschauer (2000), who refer to intelligent CALL as a “distant dream.” Table 2.1 below presents a comprehensive summary of the stages in the history of CALL according to Warschauer (2004).

Table 2.1 The three stages of CALL according to Warschauer (2004)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>Mainframe</td>
<td>PCs</td>
<td>Multimedia and Internet</td>
</tr>
<tr>
<td>English-teaching paradigm</td>
<td>Grammar-Translation &amp; Audiolingual</td>
<td>Communicate [sic] language teaching</td>
<td>Content-Based, ESP/EAP</td>
</tr>
<tr>
<td>View of language</td>
<td>Structural (a formal structural system)</td>
<td>Cognitive (a mentally constructed system)</td>
<td>Socio-cognitive (developed in social interaction)</td>
</tr>
<tr>
<td>Principal use of computers</td>
<td>Drill and practice</td>
<td>Communicative exercises</td>
<td>Authentic discourse</td>
</tr>
<tr>
<td>Principal objective</td>
<td>Accuracy</td>
<td>And fluency</td>
<td>And agency</td>
</tr>
</tbody>
</table>
Bax’s account

According to Bax (2003: 19-20), Warschauer’s account deserves credit for being the first attempt to conceptualize the development of CALL; however, his formulation needs amendment and clarification in a number of areas. First, it is unclear whether his stages represent clearly-defined historical periods, given the inconsistent time-frames proposed throughout his work, and given the fact that he himself acknowledges that the three paradigms are currently coexisting in the language classroom. Second, the label “communicative CALL” may be misleading, as the software allegedly used in this phase does not capture the two basic principles of Communicative Language Teaching (CLT), namely that learners learn language in order to communicate, and that they probably learn to communicate best through the process of communication itself (Littlewood, 1981). Third, the distinction between “communicative” and “integrative” CALL in terms of software promoting the use of language vs. the use of language in authentic social contexts is unjustified, since such use had been stressed from the very beginnings of CLT. Finally, it is unclear, or questionable, whether the phases proposed correspond to a list of desiderata of the type of software that was sought at each time, or a description of the software that was actually developed, or a description of how instructors used the software, or how they should be using it.

In his attempt to address these issues, Bax proposed an alternative analysis. Like Warschauer (2004), he presented a three-tiered classification, but he substituted the concept of historical stages by the more general term “approach” — although he later added that they do coincide with general historical periods (that happen to correspond to the same time-frames as those by Warschauer). Second, Bax substituted the labels
“behaviorist/structural” and “communicative” CALL with “restricted” and “open” CALL, respectively, since they prevent conceptual confusion with the language teaching approaches they refer to. Similarly, he also substituted Warschauer’s “integrative” CALL by “integrated” CALL, although he did not provide a rationale for this. Third, Bax used a more fine-grained classification matrix, where the three approaches to CALL are defined and compared in terms of systematic dimensions: type of task, type of student activity, type of feedback, teacher roles, teacher attitudes, position in curriculum, position in lesson, and physical position of the computer. Finally, Bax suggested that his classification is more accurate as a description of what happened in the past and what is happening now, although he specified that the last approach, “integrated” CALL, is currently a work in progress, and that language instructors are still operating within the second approach, “open” CALL. A comprehensive summary of Bax’s (2003) account is reproduced in Table 2.2 below.
<table>
<thead>
<tr>
<th>Content</th>
<th>Type of task</th>
<th>Type of student activity</th>
<th>Type of feedback</th>
<th>Teacher roles</th>
<th>Teacher attitudes</th>
<th>Position in curriculum</th>
<th>Position in lesson</th>
<th>Physical position of computer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Restricted CALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language system</td>
<td>Closed drills</td>
<td>Text reconstruction</td>
<td>Correct/Incorrect</td>
<td>Monitor</td>
<td>Exaggerated fear and/or awe</td>
<td>Not integrated into syllabus-optional extra Technology precedes syllabus and learner needs</td>
<td>Whole CALL lesson</td>
<td>Separate computer lab</td>
</tr>
<tr>
<td></td>
<td>Quizzes</td>
<td>Answering closed questions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Minimal interaction with other students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Open CALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System and skills</td>
<td>Simulations</td>
<td>Interaction with the computer</td>
<td></td>
<td>Monitor/ facilitator</td>
<td>Exaggerated fear and/or awe</td>
<td>Not integrated into syllabus-optional extra Technology precedes syllabus and learner needs</td>
<td>Whole CALL lesson</td>
<td>Separate lab- perhaps devoted to languages</td>
</tr>
<tr>
<td></td>
<td>Games</td>
<td>Occasional interaction with other students</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CMC</td>
<td>Focus of linguistic skills development</td>
<td>Open, flexible</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Integrated CALL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated language skills work Mixed skills and system</td>
<td>CMC</td>
<td>Frequent interaction with other students</td>
<td>Interpreting, evaluating, Commenting, stimulating thought</td>
<td>Facilitator/ Manager</td>
<td>Normal part of teaching-normallised</td>
<td>Tool for learning Normalised Integrated into syllabus, adapted to learners' needs Analysis of needs and context precedes decisions about technology</td>
<td>Smaller part of every lesson</td>
<td>In every classroom, on every desk, in every bag</td>
</tr>
</tbody>
</table>
The usability of Bax’s and Warschauer’s accounts

Overall, the historical accounts by Warschauer (Kern & Warschauer, 2000; Warschauer, 2004; Warschauer & Healey, 1998) and Bax (2003) constitute a remarkable effort to understand the evolution of CALL technology, in terms of historical stages, instructional approaches, or both. Precisely because of their multi-faceted nature, these accounts intertwine a number of variables, such as (a) the type of CALL software and (b) how the software is implemented into the classroom environment. To this regard, in order to create a more objective taxonomy of the CALL applications that are available in the market, it may be advisable to isolate the “computer factor” from the “human factor.” For example, it is easy to imagine different scenarios where the same piece of software could be used “structurally,” “communicatively,” or “integratively,” depending on the teacher’s or learner’s philosophy. Therefore, it is perhaps best to organize computers according to what they are designed to do rather than how they are implemented. Such a taxonomy would allow us to empirically compare the benefits of the different CALL applications that exist, independently of the teaching philosophy that governs their use. This is the goal of the next section.

1.2.2 A usable taxonomy of CALL resources

Quite possibly, a “usable” taxonomy of CALL resources may be best approached by distinguishing the different roles that the computer can play in the language learning process. In the field of education technology, Taylor (1980) distinguished three main roles of the computer, depending on its interaction with the user: tutor, tool, and tutee. To act as a tutor, the computer must be preprogrammed or equipped with software developed
by experts. The computer presents some subject material or input, the student responds, and the computer evaluates the response and determines the next step to be taken. To function as a tool, the computer need only be equipped with some useful capability, allowing the user to perform some kind of operation, such as statistical analysis, super calculation, or word processing. Finally, the computer can be a tutee, responding to the instructions given by the user in a programming language that it understands.

More recently, further research in educational technology has eluded the role of the computer as a tutee, while introducing new roles. Thus, Means (1994) distinguished four different roles of the computer: tutor, exploratory environment, tool, and communication media. Tutoring programs are used to teach students directly by providing information, demonstration, and practice opportunities. Examples of tutor programs are tutorials or practice-based instructional software. Exploratory environments are used to encourage proactive student exploration and discovery learning. Examples include microworlds, simulations, and hypertext- or hypermedia-based learning environments. Tool programs include general-purpose technological tools such as word processors and desktop publishing software, spreadsheets, and data-analysis software. Communication media involve the use of computers as a communication channel, including e-mail, computer-conferencing, and computer-collaborative learning systems.

Moving away from the field of educational technology to the specific realm of SLA, Levy (1997), drawing on Taylor’s (1980) model, proposed a two-tiered classification of CALL resources - the “tutor-tool framework.” In this classification, the computer acts as a tutor when it evaluates the student input in some way and as a tool when it does not. The present study will use an eclectic classification based on Levy (1997) and Means
(1994), thus combining the simplicity of the former and the degree of detail of the latter. In this classification, two main roles of the computer are distinguished, i.e., tutor vs. medium. As a medium, the computer may serve as an exploratory environment for language information retrieval (e.g., via offline vs. on-line databases such as the Internet); a tool for language information processing (e.g., via, e.g., text editors, text concordance software, etc.); and a communication channel for information dissemination, be it synchronous (i.e., chat in its different modalities, text, audio, video, or a combination) or asynchronous (e-mail and electronic forums). Some uses of the computer as a medium in SLA fall under the umbrella term Network-Based Language Teaching (NBLT - Kern & Warschauer, 2000), which subsumes other labels such as Computer-Mediated Communication (CMC) and Project-Oriented CALL (PrOCALL - Debski, 2000). On the other hand, the computer can act as a language tutor and, depending on its ability to understand and respond to the learner’s response, it can lead to “intelligent” vs. “non-intelligent/traditional” Computer Assisted Language Instruction (iCALI vs. CALI). Figure 2.1 below illustrates my taxonomy, which is used as a starting point to present an overview of empirical research on CALL in the next section.
Figure 2.1 A taxonomy of CALL resources based on the role of the computer
1.2.3 Empirical research on CALL

The proposed dichotomy computer as tutor vs. medium can be used to classify different strands of research in CALL. Table 2.3 below presents a classification of empirical research on CALL since 1990. The studies in this classification were extracted from previous reviews and meta-analyses (Debski, 2003; Felix, 2005; Grgurovic, 2007; Hubbard, 2005; Stockwell, 2007; Zhao, 2003), as well as independent searches in two popular databases, Linguistics and Language Behaviour Abstracts (LLBA) and Education Resources Education Center (ERIC). To be included in this classification, a study had to use some measure of L2 development in the form of either learning outcomes (e.g., accuracy, latency/fluency, complexity) or learning processes found or posited to be associated to learning (e.g., noticing and uptake). Studies analyzing discourse, communication strategies, and constructs of the sociocultural framework were not included in this classification. Overall, the referenced studies were published in the following journals, among others: CALICO, Computer Assisted Language Learning, Foreign Language Annals, Hispania, Journal of Educational Computing Research, Language Learning, Language Learning and Technology, Modern Language Journal, Studies in Second Language Acquisition, and System.

Each of the columns in Table 2.3 encompasses different sub-lines of research. For example, the column “Exploratory environments” includes research on the effects of different types of glosses (text, picture, audio, and video) in incidental hypertext reading, research on the effects of different types of subtitling (full, keyword) in film watching, and studies on the effects of different types of instruction, where the computer is just a display of rules, does not interpret the response of the user in any way, and hence does
not produce any feedback. Similarly, the column labeled “Synchronous communication media” contains studies on the different facets of CMC discourse as well as studies on the potential benefits of CMC vs. Face to Face (FtF) interaction in L2 development.

Additionally, as the attentive reader will notice, some studies appear in two columns because they involve or compare different roles of the computer. For example, Nikolova’s (2002) study is placed both under “Exploratory environments” and “Tool programs” because it compares vocabulary learning derived from reading a glossed hypertext (the computer as an exploratory environment) vs. creating a glossed hypertext (the computer as an html editing tool). Similarly, Biesenbach-Lucas, Meloni, and Weasenforth (2000) compare the differential effects of writing a text as an e-mail vs. a text document using a word processor, which is why this study is classified both under “Asynchronous communication media” and “Tool programs,” respectively.
Table 2.3 A classification of empirical research on CALL from 1990 to 2010

<table>
<thead>
<tr>
<th>Tutor</th>
<th>Medium</th>
<th>Communication media</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALI</td>
<td>Exploratory environments</td>
<td>Asynchronous</td>
</tr>
<tr>
<td>iCALI</td>
<td>Tool programs</td>
<td>Synchronous</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medium</th>
<th>Asynchronous</th>
<th>Synchronous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrams (2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Böhlke (2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitze (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hirotani (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kern (1995)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kötter (2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lai &amp; Zhao (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee (2002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matsumura &amp; Hann (2004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morris (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Payne &amp; Whitney (2002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satar (2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shekary &amp; Tahririan (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sotillo (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sullivan &amp; Pratt (1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Compernolle (2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yiilmaz &amp; Granena (2010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zeng &amp; Takatsuka (2009)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As Table 2.3 shows, there is a clear imbalance between experimental research on the use of the computer as a medium vs. tutor, with the former being much more profuse than the latter. This dearth of empirical research on CALI becomes particularly striking when considering the increasing presence of CALI systems in our language laboratories and their booming sales outside academia. A paramount example is Rosetta Stone, which since its foundation in 1992 has become available in 30 languages, is being sold in 150 countries, and according to Inc Magazine (http://www.inc.com) has positioned its developer, Fairfield Language Technologies, within the ranking of the 500 fastest-growing private companies in the United States, with the highest revenue among educational companies in 2006 and 2007 (http://pr.rosettastone.com, accessed 8/22/2009). This research gap (in CALL in general, but even more so in CALI) is, as Hulstijn (2000: 35) puts it, the result of the disinterest of both product developers and academics:

[I]n the area of language teaching, electronic multimedia hardware and software has replaced the traditional audio and video systems in the language laboratory. Hence, educational research comparing the relative benefits of various teaching techniques and methods is bound to investigate the use of these multimedia systems. Unfortunately, there is far too little research of this type. The most likely reason for this dearth of research is that, on the one hand, the companies that develop and produce these systems are hesitant to finance such investigations, and that, on the other hand, foundations for the funding of academic research, often find such “applied” investigations less important than research considered more “fundamental.”

Now, as discussed earlier, existing empirical literature on CALI can be classified into two main blocks, i.e., research on traditional vs. intelligent tutors (CALI vs. iCALI). The difference between the two lies on how the computerized tutor evaluates the learners’ responses. A CALI system simply takes in the response of the learner, compares it against a database of acceptable and unacceptable answers by using a pattern matching
technique, and produces “canned” or pre-fabricated feedback or output, as applicable. In contrast, an iCALI system relies on Natural Language Processing (NLP) tools that are able to parse the students’ utterances, that is, “to encode complex grammatical knowledge such as humans use to assemble sentences, recognize errors and make corrections” (Holland, Maisano, Alderks, & Martin, 1993). Additionally, some recent iCALI systems, often known as “learner modeling” systems, also incorporate algorithms that record and analyze performance and behavioral information about the learners as they progress through the program, and thus these systems are not only able to provide dynamically-generated feedback but are also able to tweak this feedback to replicate pedagogical decisions similar to those made by human teachers (D.-H. Kim et al., 2006).

To date, as Table 2.3 illustrates, very few empirical studies have examined the potential of iCALI for SLA. This is probably due to the fact that iCALI systems are computationally very expensive and typically require the in-tandem collaboration of computational and applied linguists. Further research on this strand is certainly warranted.

Focusing specifically on CALI, existing published empirical studies can be grouped up into five main categories. One category observes whether the use of CALI per se is conducive to L2 learning, without comparing it to other types of tutors or other learning conditions (e.g., Hardison, 2004). In contrast, other studies compare the efficacy of CALI against other learning conditions, e.g., teacher-based instruction, learner-to-learner interaction, or instruction based on other technologies such as audio, video, books, or paper-and-pen (Brett, 1997; Chen et al., 2004; Groot, 2000; Nagata, 1996; Nutta, 1998; Watts, 1989). A third type of studies observe how learners use CALI in naturalistic
settings, observing which features of the tutor are actually accessed by users and which of
them are more facilitative in terms of L2 development (e.g., Hegelheimer & Tower,
2004). Fourth, a number of published studies investigate the effects of different types of
computerized feedback (Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004b;
Sanz & Morgan-Short, 2004). Finally, a fifth line of research investigates the effects of
different types of computerized practice (Morgan-Short & Wood Bowden, 2006; Nagata,
1998a, 1998b), or the separate and combined effects of computerized practice and prior
metalinguistic presentation (Rosa & Leow, 2004b; Sanz & Morgan-Short, 2004).

Clearly, the first and second group of published studies seek to answer the “if CALI”
question, i.e., if CALI can facilitate L2 development, either per se or versus other
technologies. On the other hand, the third, fourth, and fifth groups of studies do not only
indirectly address the “if CALI” question but also cast some light on the “which CALI”
question, i.e., which type of CALI features or systems are more beneficial for L2
development. As posed by Conrad (1996) and Hegelheimer and Tower (2004: 186), since
the 1980s research in CALI has shifted from investigating if this technology is superior to
non-CALI to how CALI can be used effectively in language learning by comparing
different types of CALI programs. Confirming this trend, a growing number of recent
doctoral dissertations have sought to build onto the latter group of studies. Thus, Bowles
(2005), Camblor (2006), Hsieh (2007), Moreno (2007), and Lado (2008) investigated the
effects of different types of computerized feedback. Additionally, Hsieh (2007), Moreno
(2007), and Medina (2008) compared different types of computerized task-based
instruction, respectively looking at the effects of practice vs. exposure to practice, more
vs. less task-essential tasks, and more vs. less complex tasks. Arguably, the field is
catching full speed and a number of investigations is underway, analyzing how the effects of computerized feedback may be moderated by other variables such as amount of exposure to the targeted form and task-induced depth of processing (Cerezo & Leow, 2006), type of targeted linguistic form (Cerezo, 2007), and cognitive capacity (Lado, 2008).

Overall, as mentioned above, most published studies on CALI since the 1980s have focused on comparing different types of CALI rather than comparing it to non-CALI technologies, and in most cases the use of CALI was a means and not a goal; in other words, it was not motivated by a genuine interest in the computer as a learning tool, but rather as a methodological tool to exert greater control over certain pedagogical interventions (e.g., to consistently provide feedback in terms of type and frequency - see, e.g., Hulstijn, 2000; Sanz, 2000 for a discussion on the benefits of the computer for methodological purposes). There is, therefore, a blatant need for further comparative research. As Nagata (1996: 55) puts it, “one now sees few empirical studies on the relative effectiveness of CALI vs. non-CALI. However, when it comes to utilizing computers for second language instruction, the question of whether and when computer programs can be more effective than traditional non-computer instruction is still a basic question to be addressed.” Producing and implementing CALI technology is time-consuming and expensive, and therefore it is crucial to determine whether CALI has an advantage over other traditional instructional technologies, and if so, whether this is due to factors inherent to the computer or a poor use of traditional media (see Stockwell, 2007; C. Williams & Brown, 1991).
In this context, it seems crucial to evaluate the potential of CALI technology by answering both the “if CALI” and “which CALI” questions., i.e., if CALI can lead to significantly higher rates of L2 development when compared to non-CALI technology, and which type of CALI technology yields the best results.

2. CALI

The present section discusses CALI from two different angles. Section 2.1 (the “if CALI” question) summarizes studies comparing CALI vs. non-CALI technology, presenting an overview of the theories that may contribute to explicate the positive results found for computerized tutors. Then, drawing on Nagata (1996), it is posited that what distinguishes computerized tutors against other instructional technologies is their ability to engage learners in interactive practice. This gives way to section 2.1.1, which discusses the literature on learner’s agency, i.e., the compared effects of practice vs. exposure to practice in L2 development. Next, in answering the “which CALI” question, section 2.2 analyzes the particular features that determine the efficacy of computer tutors, focusing specifically on type of computerized feedback in section 2.2.1.

2.1 The “if CALI” question: Can computerized tutors facilitate L2 development versus other technologies?

Recently, CALI systems such as The Rosetta Stone have populated the market and are making their way into the academic and non-academic language learning laboratories. In this context, language program directors considering the option of incorporating this technology into their curricula may want to first gauge their potential benefits versus
existing resources, since either developing CALI technology or purchasing software user licenses is a rather costly endeavor. Recent literature (Camblor, 2006; Hsieh, 2007; Lado, 2008; Moreno, 2007; Morgan-Short & Wood Bowden, 2006; Nagata, 1993, 1998a, 1998b; Rosa & Leow, 2004a, 2004b; Sanz & Morgan-Short, 2004) has shown that CALI systems may promote L2 grammar development, but are they worth the investment? Can they parallel or even supersede other traditional pedagogical technologies?

**CALI can promote L2 development even better than other instructional technologies**

In the field of education technology, there is a growing body of comparative research investigating the effectiveness of computer tutors vs. different forms of classroom instruction to teach an array of subjects in the L1, generally showing a positive effect for computers (see, e.g., McEnery, Baker, & Wilson, 1995, on the positive effects of CALI vs. teacher-centered instruction in the learning of L1 parsing; for non-linguistic subjects see the reviews by Kulik & Kulik, 1987; Lou, Abrami, & d’Apollonia, 2001; McNeil & Nelson, 1991; Niemiec & Walberg, 1987; and Ragan, Boyce, Redwine, Savenye, and McMichael, 1993). In contrast, in the SLA arena most research has focused on comparing the relative efficacy of learner interaction *through* the computer (i.e., computer mediated communication - CMC) against other forms of learner interaction (e.g., face to face - FTF), while little research has been conducted to compare the benefits of the computer as a tutor versus other instructional media. So far, research of this type has mostly explored the comparative effects of multimedia CALI on reading comprehension, and to a lesser extent on listening comprehension (e.g., Brett, 1997), finding either non differential
results or an advantage of computers, but very few studies have done the same for vocabulary or grammar acquisition.

In this dearth of research, perhaps the most representative studies comparing the effects of CALI vs. non-CALI for grammar and vocabulary acquisition are Watts (1989), Nagata (1996), and Nutta (1998). These three studies showed that CALI can outperform VCR-, workbook-, and teacher-centered instruction, respectively, in the acquisition of L2 grammatical structures (and vocabulary, in Watts’s study), and that this advantage manifests itself in production rather than in recognition of the forms.

Probably the earliest empirical attempt to compare CALI vs. non-CALI for grammar and vocabulary instruction is constituted by Watts (1989). In her study, 30 students of beginning French from different adult education centers in the UK were confronted with two different versions of the _Dès le Debut_ video series during two two-hour sessions: a traditional video with exercises in paper-and-pen format and a transposition of the same materials into an interactive video disc. Results of a pretest-posttest experiment showed that the CALI approach yielded better results, with a combined gain of 30-35% on word translation, sentence translation, and listening comprehension, and a notable decrease in pronunciation errors. On the other hand, the VCR and paper-and-pen approach did not seem to yield comparable gains, although no specifics are reported: “[a]lthough some improvement was discerned between the pre- and post-study sessions, the results were not as significant as for the interactive approach and, in particular, only minimal improvement in pronunciation was recorded.” Thus, Watt’s results must be taken with caution in the absence of a solid statistical analysis and various methodological caveats. For example, no specific linguistic forms were investigated, so the relative difficulty of
the items in the assessment tests delivered after each treatment may have played an
intervening role.

Focusing on specific grammatical structures, i.e., the Japanese particles that serve to
mark the grammatical and semantic roles of noun phrases, Nagata (1996) compared the
effectiveness of CALI vs. workbook-mediated instruction. In her study, 26 participants
were exposed to a session of prior grammar explanation and then engaged in one of four
treatment sessions, where they were to read a set of grammar notes and complete a series
of practice exercises with the aid of a CALI system or a workbook and an answer sheet.
Analysis of the achievement scores revealed that the CALI group significantly
outperformed the workbook group on the production posttests (fill-in-the-blank and
sentence production) whereas no significant between-group differences were found in the
comprehension posttests (sentence translation into L1 English). The positive advantage of
the CALI group in production held in a delayed posttest three weeks later, and again in an
oral interview three days after it. Based on these results, Nagata concluded that intelligent
CALI outperforms workbook-mediated instruction for complex structures at the
production level. The lack of significant difference at the comprehension level was
interpreted along Flynn’s (1986) hypothesis that grammar competence is less critical in
comprehension than in production. When interpreting Nagata’s findings, a number of
relevant points must be considered. First, the feedback received by the CALI and non-
CALI (workbook-instruction) groups was qualitatively different: “Nihongo-CALI
provides ongoing feedback in response to the learners’ grammatical errors, while
workbook instruction is simply accompanied by answer sheets without any explanations
on the individual errors” (p. 59). Therefore, it is unclear whether the advantage of the
CALI group was due to the learning environment (CALI vs. non-CALI) or the particular use of that environment (i.e., the degree of explicitness of the feedback provided). Also, timing and amount of modified output on the part of the learner were important factors. The students in the CALI-group received ongoing feedback for their errors and were given up to three opportunities to modify their answer in a more targetlike direction. Conversely, the students in the workbook group completed all exercises and only then were given an answer sheet to correct their answers. This answer sheet, by definition, precluded any possibility for modified output.

Switching the scope of the comparison from poorly interactive workbooks to the very interactive humans, Nutta (1998) compared the effects of a CALI system vs. teacher-directed instruction for the acquisition of two different grammatical structures, English past and conditional tenses. In her study, 53 participants were exposed to an hour of instruction per day for seven days. Students in the teacher-directed group “used the Focus on Grammar (1994) textbook series and engaged in a variety of activities emphasizing interactive, meaningful, and creative expression”, whereas students in the CALI group followed a sequential checklist developed by the researcher to navigate through a multimedia CALI system, ELLIS Mastery, completing a series of activities with computerized feedback. The results of this study for the past tense structure showed a significant advantage of the CALI group over the teacher-directed group in production posttests (fill-in-the-blank and open-ended sentence completion), while no differences were observed for recognition of the structure via a multiple-choice task. These results were partially maintained after 2 weeks, with the computer group outperforming the teacher-directed group in open-ended sentence production. The results for the other target
structure, English conditional, were less differential, with similar gains for both groups on all tests, except for delayed open-ended sentence production, where again the CALI group outperformed the teacher directed group. Interestingly, Nutta’s results show that computer-based grammar practice can work at least as effectively as, or even more effectively than, teacher-directed practice. However, very few details are provided about the kind and quality of the interactive experience of the students in the teacher-directed group. Consequently, since the activities and materials used by the CALI system and the human teachers were not the same, one could argue that the poorer results of the teacher-directed students were due to poor teacher performance, among other factors.

To sum up, then, incipient research suggests that CALI can work at least equally well, and even better than, other instructional technologies using VCRs, workbooks, and human teachers. However, given the aforementioned methodological caveats and the overall paucity of studies, more research is warranted in order to assess whether the edge of computers versus other technologies is due to a poor use of such technologies (see, e.g., Stockwell, 2007; C. Williams & Brown, 1991) or factors that are inherent to the computer. On this token, the studies reviewed above provide little detail as to the theoretical underpinnings that may contribute to explicate their findings. That is the intended goal of the next subsections.

**Why can computers promote L2 learning?**

At least three different theoretical frameworks can be proposed to account for the edge of CALI systems versus other pedagogical technologies: theories on the effects of code type on information processing (e.g., Paivio’s Dual Coding Theory and Mayer’s Theory of
Multimedia Learning), theories on learners’ attention and motivation, and the Interaction Approach (Gass & Mackey, 2006).

**CALI can effectively combine multiple media**

Arguably, computers have an extraordinary ability to display different types of information (verbal or non-verbal) through a variety of modes (visual and aural). Thus, text, pictures, videos, and sound can be effectively combined through the computer to create multimedia learning materials that may enhance information processing and learning according to at least two theories of cognitive psychology, Paivio’s (1969; 1986) Dual Coding Theory and Mayer’s (2001) Theory of Multimedia Learning. In a nutshell, these theories contend that visual and verbal information are processed differently and separately, along distinct channels, and that they can facilitate processing and learning of information if displayed concurrently through a multimedia environment as long as they do not compete with each other overloading working memory. For example, a television documentary that shows images of plant and animal life in a rain forest while also simultaneously providing narration that describes the animal life could potentially enhance learning because the visual and verbal information do not compete with each other. Thus, the theories by Paivio and Mayer could contribute to explicate why CALI technology can boost information processing and learning over workbook instruction, since computers are more versatile than books in displaying simultaneously multimedia information. However, this is not the case in Nagata (1996), because her Nihongo-CALI system displayed only verbal information in the written mode, and yet it promoted higher learning gains than the workbook approach. Furthermore, these theories would not
explain why CALI may outperform VCR-based instructional materials (Watts, 1989) or teacher-centered instruction (Nutta, 1998), since both movies and teachers can concurrently display visual and verbal information, written and aural, in effective ways.

**CALI can raise learners’ attention and motivation**

From a different standpoint, it may be then the case that the personalized experience that the computer can provide helps students focus their attention on L2 input. According to Watts (1989: 19-20), one of the advantages of the computer-based interactive approach over the VCR approach is that it eliminated student passivity toward the television. For example, Watts attributed the edge of CALI for pronunciation to the fact that the vast majority of the students (29 out of 30) actually spoke aloud when asked to do so, as opposed to a minority (4-7 students out of 30) in the VCR approach. Interestingly, when asked about their reluctance to follow instructions, the students declared that they felt stupid talking to a TV and that “[i]t doesn’t give you the same confidence that a computer gives you.” This is interesting because the software used by Watts had no ability to evaluate the students’ oral output in any way, and yet an overwhelming majority of students behaved as if it did. When asked about their opinions in a post-experiment structured interview, most students preferred the interactive approach because they deemed it to be more enjoyable and conducive to learning, both for them and future users. Furthermore, some students who enjoyed both approaches found their attention wandering when using the VCR and worksheet approach, which is borne out by the fact that fewer than half of the worksheets were fully completed at the end of the sessions. This suggests that computers have a greater ability than other media (such as books or
non-interactive videos) to direct students’ attention to L2 input, perhaps because they increase students’ motivation to learn. This point is typically corroborated by comparative studies where participants are surveyed on their attitudes towards CALI vs. non-CALI. For example, in Nagata’s (1996: 66) study, participants in the CALI group manifested a significantly greater satisfaction in their ratings of the format, interest, and content of the practice exercises, “in spite of the fact that both groups received the same exercises in the same format.” Overall, the CALI group found that their learning experience had been more engaging, as shown by their greater enthusiasm in rating a survey item reading “I want to practice Japanese by using this type of exercise on a regular basis.” To this regard, further empirical research that compares CALI vs. non-CALI technology from an attentional and/or motivational perspective is certainly warranted.

**CALI can promote large amounts of interactive practice**

Even if, as suggested by previous literature, CALI is confirmed to raise students’ motivation and attention to L2 input, there still remains the question of why this occurs. As pointed out by Nagata (1996: 66), one possible reason for the differences observed between CALI and workbook-based instruction might be the latter’s “lack of interactive features.” From the late 1960s to the mid 1980s, a number of empirical comparative studies reviewed by Pederson (1987) showed no significant advantages of the computer vs. other L2 learning technologies, such as audiovisual lab-based instruction (Morrison & Adams, 1968), teacher-centered instruction (Adams & Rosenbaum, 1969; Barrutia, 1970; R. Williams, 1980), or paper-and-pencil instruction (Mellgren, 1983). Although this
could at first be interpreted as counterevidence for the benefits of CALI-based instruction presented earlier, it should be noted that none of the computer systems used in these studies conformed to the definition of computer tutors (see, e.g., Levy, 1997, Means, 1994) since they were not capable to evaluate the student input in any way and respond to it accordingly. Rather, what these studies did was use computers as media, and more explicitly, as the specific subtype that Means (1994) referred to as exploratory environments, since they were a mere transposition of printed materials into a screen. As argued by Nagata (1996: 54), “if CALI programs simply transfer printed textbooks to the computer screen, the computer is nothing more than an electronic page-turner, and we cannot expect better results from CALI.” It appears then, that the advantage of the computer tutor versus other technologies lies in its capability to interact with learners, or to present them with large amounts of interactive practice. Interestingly though, very few empirical studies on CALI have been premised upon the theoretical framework of the Interaction Hypothesis (Long, 1981; 1996, and elsewhere), recently rechristened as the Interaction Approach (Gass & Mackey, 2006). As Chapelle (1997: 22) puts it:

CALL [and CALI] researchers can turn to the work of interactionist SLA researchers who operate under the assumption that the L2 is acquired through learners’ interaction in the target language because it provides opportunities for learners to: (a) comprehend message meaning, which is believed to be necessary for learners to acquire the L2 forms that encode the message; (b) produce modified output [in response to a particular type of feedback], which requires their development of specifics of morphology and syntax; and (c) attend to language form, which helps to develop their linguistic systems.

To wrap up, then, (1) one possible factor contributing to the benefits of CALI over other instructional technologies may be the ability of computer tutors to provide learners with interactive practice; and (2) as posited by the supporters of the Interaction Approach, interaction is connected to L2 development because “it connects input, internal learner
capacities, particularly selective attention, and output in productive ways” (Long, 1996: 451-2).

Now, in order to do research on CALI under the Interaction Approach, it is necessary to understand the nature of the interactive learning experience that computer tutors can facilitate. As Nutta (1998) notes, here researchers are divided in two fronts. Some advocate for a “drill-and-kill” approach, where computers would serve to teach grammar, thus freeing up more classroom time for real communication (e.g., Gilby, 1996; Hoffman, 1996). Others, however, have set more ambitious goals for computer tutors, asserting that they should attempt to replicate what occurs in the classroom in order to provide opportunities for meaningful communication (e.g., Garrett, 1991; Lavine, 1992; Quinn, 1990; Underwood, 1993). Regardless the nature of computer-human interaction in CALI, the underlying workflow is still the same. First and foremost, the learner must be engaged in some sort of practice. This practice, which can be more or less structured (from solving a task to taking part in casual conversation), may require the learner to recognize, interpret, comprehend, or produce a message. Whatever the nature of the practice, when the learner produces a response, the computer must then analyze it, either superficially, by using string-matching techniques that compare the student’s response against a database of possible answers (CALI), or more in depth, by using parsing and knowledge representation techniques (iCALI). Next, the computer must react accordingly, by providing feedback (which may reformulate the learner’s response in a more target-like way, or may invite the learner to reformulate the response, or both), and/or by moving forward in the dialogue. In conclusion, in CALI, like in non-CALI environments, (1) there is no interaction without practice; and (2) the nature of the interactive learning
experience is determined by the type of practice used and the type of reaction to learners’ behavior, where feedback has a central role.

The two points just raised are directly linked to the “if” and “which” questions in CALI. If the edge of computer tutors viz à viz other instructional technologies lies in their ability to engage learners in interactive practice, then empirical research must address this by comparing whether or not CALI systems that engage learners in interactive practice outperform CALI systems that only act as electronic page turners. This comparison is referred to in this dissertation as the role of “learner’s agency” in interaction.

Additionally, further research must explore which systems yield the best results by examining whether the nature of the practice (either the type of practice used or the type of other practice components such as feedback) also makes a difference. These are the goals of the next sections.

2.1.1 Learner’s agency

As posed by a number of authors (Nagata, 1996; C. Williams & Brown, 1991), the edge of computer tutors versus other instructional technologies (e.g., workbook-based instruction, VCR-based instruction) may lie in the interactivity that CALI systems propel. However, this claim remains to be validated empirically. In order to do so, and to guarantee that the edge of CALI over other technologies is indeed due to its potential for interactivity and not a poor use of traditional technologies (see, e.g., Stockwell, 2007; C. Williams & Brown, 1991), all variables but interactive practice must be isolated. Specifically, future research must compare the learning gains of students who interact with two different versions of the same computerized materials. In one version, the CALI
system acts as such, providing interactive practice and engaging the learner in it. In the other, the system acts as a VCR or an electronic page-turner that showcases the interaction sessions of practicing students. Hitherto, in the incipient body of CALI literature, only Hsieh’s (2007) doctoral dissertation has specifically done this for L2 grammar development. In her study, which targeted Spanish gustar constructions, half of the participants completed an input-based task with different types of written feedback, while others observed their performance thanks to video capture software. Results showed that there were no significant differences between groups for either written recognition or production, while interactants outperformed observers in oral production. Outside the realm of CALI, empirical research on practice versus exposure to practice is also incipient. Strictly speaking, only Mackey (1999) and Muranoi (2000) have compared two experimental conditions (practice vs. exposure to practice) to address the very ecological fact that both in the classroom and in naturalistic contexts learners may not only act as interactors but also as observers, merely monitoring what others say and the responses they get in return. The results of these two experiments are mixed. Mackey obtained a positive effect of practice on L2 development, with practicing participants outperforming exposure participants. Conversely, Muranoi did not obtain significantly different learning gains between the two groups.

Prior to Mackey’s and Muranoi’s research, a number of observational studies reviewed by Ellis (1988) focused on the impact of voluntary in-class participation on L2 development, comparing the learning gains of more vs. less participating students. This construct, however, must be examined with caution. While some authors refer to it as “practice” (Ellis, 1988; Seliger, 1977), it should probably be more conservatively referred
to as “participation” (R. Day, 1984) since both constructs are actually different. There are many variables that can determine the rate of active participation of a student: motivation, personality, language experience, etc. As a result, this early research may show us whether there is a correlation between voluntary in-class participation (and hence more output practice) and L2 development, but it cannot show us whether students, regardless of their personal drive to participate, may differently benefit from being pushed to practice or observe the practice of others. With this precision in mind, it is still hard to extrapolate any general findings from this block of early studies: three studies showed positive effects for participation (Ellis & Rathbone, 1987; Naiman, Fröhlich, Systern, & Todesco, 1978; Seliger, 1977), two yielded no significant differences between participation and observation (R. Day, 1984; Ely, 1986), and one showed negative effects of participation (Ellis, 1984).

On a different note, very few studies have successfully isolated the practice or participation variables. As pointed out by Hsieh (2007), most studies have conflated these constructs with corrective feedback, and thus it is impossible to determine whether any observed developmental effects were due to the practice/participation, the feedback, or a combination thereof. To this regard, studies addressing the role of practice and exposure to practice should incorporate a control group with no feedback into their research designs.

To wrap up, then (1) empirical literature on the effects of practice versus exposure to practice on L2 grammar development is notably scant, with only one study in CALI (Hsieh, 2007) and two in oral interaction (Mackey, 1999; Muranoi, 2000); (2) practice should not be confused with the related construct of voluntary participation; (3) research
on both practice and voluntary participation shows mixed results; and (4) practice and participation have often been conflated with another variable, feedback. For all these reasons, more research is warranted to elucidate the effects of practice vs. exposure to practice with and without feedback on L2 development.

2.2 The “which CALI” question: Which computerized tutors yield the best results?

Answering the “which CALI” question requires investigating the contributions of the different components that make up an interactive learning experience, not in terms of their presence or absence but rather in terms of their type. Arguably, two are the main components of an interactive learning experience: (1) practice, which as pointed out by DeKeyser (2007) is typically thought of as a pre-requisite for interaction, and (2) feedback, which, as argued by Loschky and Bley-Vroman (1993) typically distinguishes practice for mere testing purposes from practice for learning purposes. As discussed below, research on both areas is clearly warranted. This dissertation deals exclusively with feedback, but in order to grasp a better understanding of the status quo in CALI a brief summary of the findings on both constructs is separately provided in sections 2.2.1 and 2.2.2 below.

2.2.1 Type of practice

Hitherto, the incipient literature on the effects of type of practice in CALI can be grouped up into two blocks: (1) studies focusing on the skills that the practice elicits on the part of the learner, i.e., comprehension, interpretation, recognition, or production (see the published studies by Morgan-Short & Wood Bowden, 2006; Nagata, 1998a, 1998b), and
(2) studies analyzing the features of the pedagogical activity through which practice is
embodied, e.g., task-essentialness, task-complexity (see the doctoral dissertations by
Medina, 2008; Moreno, 2007).

Specifically on the issue of elicited skills, the pedagogical effects of input- vs. output-
focused practice has been studied by Nagata (1998a) for Japanese honorifics, Nagata
for Spanish preverbal direct object pronouns. In Nagata (1998a; 1998b), five types of
exercises were used to elicit interpretation or production at the word-, sentence-, or
paragraph-level. Input-focused practice involved selecting, from three choices, the correct
English interpretation for the Japanese targeted form. In contrast, output-focused practice
involved using Japanese in translation, editing, or retelling activities. Similarly, in
Morgan-Short and Wood Bowden (2006) referential and affective activities were
included to elicit interpretation or production in the aural and written modes, although the
focus was on the word and sentence levels, rather than the paragraph level. Taken
together, these studies found that both input- and output-focused CALI led to comparable
gains in the interpretation of the target structures at play, while for production output-
focused practice yielded the best results, either only in the short-term (Morgan-Short &
Wood Bowden, 2006) or even weeks later as well (Nagata, 1998a, 1998b). Furthermore,
as showed by Nagata (Nagata, 1998a) and Nagata (1998b), the superiority of output-
focused instruction manifested itself for both complex and easy structures, respectively.

When comparing these findings to non-CALI literature, the following conclusions can be
drawn: (1) These findings lend support to the growing number of empirical studies
showing the facilitative role of output in L2 grammar development (e.g., Gass & Varonis,
They contradict a number of comparative studies, mostly in the processing instruction strand (PI), which found that the two types of practice do not lead to significantly different gains in production, whereas input-focused PI yields better performance in interpretation (Benati, 2001, 2005; Cadierno, 1995; Farley, 2001a, 2001b; VanPatten & Cadierno, 1993a, 1993b). And (3) they contradict skill acquisition theory (Anderson, 1982, 1987, 1993), according to which both input- and output-focused practice develop corresponding comprehension and production skills (DeKeyser, 1997, 2001; DeKeyser & Sokalski, 1996).

Moving away from elicited skills to task features, Medina’s (2008) doctoral dissertation compared the pedagogical effects of task complexity in input-focused tasks. In her experiment, participants were asked to read a text containing exemplars of the targeted structure, Spanish imperfect subjunctive. Participants in the “more complex task” condition were asked to spot the structures and select them in the text. Conversely, students in the “less complex task” condition were to re-highlight the forms that were already underlined for them. The results showed that task complexity had a significant effect on the written production of the target structure, with participants in the more complex task group achieving significantly better scores. Comparison of these findings to non-CALI related literature yields the following two conclusions: (1) Medina’s findings confirm most empirical research on the Cognition Hypothesis, according to which more complex tasks promote higher levels of accuracy (Iwashita, McNamara, & Elder, 2001; Kuiken, Mos, & Vedder, 2005; Kuiken & Vedder, 2007; Robinson, Ting, & Urwin,
and (2) the observed lack of effects of task complexity for recognition confirms Flynn’s (1986) hypothesis that comprehension precedes production, so the effects of more complex tasks are specially noticeable in production measures.

Finally, also delving into task features, Moreno’s (2007) doctoral dissertation investigated the effects of task-essentialness in input-focused practice. In her experiment, which targeted Spanish pre-verbal direct object pronouns, students were presented with a picture cue and were then asked to compose or select a sentence describing it. In the more task-essential condition, students were required to compose the sentence by navigating a tree that successively branched out with a set of object pronouns. In contrast, students in the less task-essential condition were to complete an interpretation task by selecting one out of two fully-composed sentences. However, in this latter case, attention to the targeted form was not a requirement, as the two sentences only differed in terms of their subjects rather than their objects. Moreno’s results did not show any significant differences for type of task, and to the present date her findings cannot be contrasted given the singularity of her study in the SLA literature.

In sum, the issue of type of practice in the CALI literature, either in terms of the skills elicited or the features of the task used, is only now starting to be empirically addressed. The incipient research on skill development unanimously suggests that output-focused practice yields better results. However, many research avenues are open. First, future research must corroborate or discard these findings. More interestingly, different subtypes of input- and output-focused practice should be compared. For example, in both Nagata (1998a; 1998b) and Morgan-Short and Wood Bowden (2006) input-focused practice was instantiated as interpretation practice, where learners were given a number
of options to select from to determine the meaning of a target structure. Thus, future designs could, for example, observe the relative contributions of interpretation vs. comprehension practice, where learners are left to their own devices and are not given options to select from. The same applies to output-focused practice, and how less vs. more controlled production at different levels (words, sentences, paragraphs) relatively affects learning. In turn, out of the two dissertations that have investigated the role of task features (Medina, 2008; Moreno, 2007), increased task complexity has been shown to have a positive effect on L2 development, while task-essentialness has not been proven to play a role. Clearly, however, the marked paucity of studies in this field of inquiry warrants further investigation to elucidate the role of these and other task features.

As opposed to the scant literature on type of practice, there is a considerably larger body of CALI research investigating the effects of type of feedback, the other component of interactive learning. Yet, the existing findings are far from conclusive. The next section delves deeper into this issue.

### 2.2.2 Type of feedback

Notably, the vast majority of studies that have addressed the “which CALI” question have focused on evaluating the contributions of different types of computerized feedback on L2 development. While these studies all agree in that feedback, regardless the type, has a positive effect on L2 grammar development, the jury is still out as to the relative effectiveness of different types of feedback against each other. Some studies have shown a significant advantage of more “explicit” types of feedback, where learners are given an explanation of the sources of their errors (see the published studies by Nagata, 1993,
Nagata & Swisher, 1995, and Rosa & Leow, 2004; and the unpublished doctoral dissertations by Bowles, 2005, and Lado, 2008). In contrast, a comparable number of studies found no significant differences between more and less informative or explicit types, where learners are merely informed that their utterances are ill-formed (see the published studies by Sanz, 2004 and Sanz & Morgan-Short, 2004, and the unpublished doctoral dissertations by Camblor, 2006; Hsieh, 2007, and Moreno, 2007). Finally, one doctoral dissertation showed significantly greater gains of less over more explicit feedback in oral delayed measures (Moreno, 2007).

In a first attempt to explicate the conflicting findings between the two seminal studies in the field (Rosa & Leow, 2004b; Sanz & Morgan-Short, 2004), Sanz and Morgan-Short (2004) proposed a number of variables that may mediate the efficacy of feedback, many of which are already under investigation. For example, Cerezo and Leow (2006) examined the role of amount of exposure to the targeted form and the depth of processing induced by the task, an issue which has also been recently investigated by Medina (2008) from the perspective of task complexity. In turn, Moreno (2007) analyzed the role of task-essentialness. Finally, Cerezo (2007) investigated the role of type of targeted structure.

The studies mentioned above are certainly igniting an interesting debate on the extrinsic factors that may mediate the efficacy of feedback; crucially, however, the conflicting findings in the literature may also result from a poor definition of the intrinsic nature of feedback. Specifically, as posed by a number of recent studies (Leeman, 2003; McDonough, 2005; Ranta & Lyster, 2007), comparing feedback types solely based on the explicitness of their negative evidence is a too reductive way of tackling the construct. In
addition to negative evidence, feedback moves may provide positive evidence (i.e., models of what the correct utterance should be) and prompting for error repair (i.e., learners may be encouraged to fix their own mistakes). However, no single study in CALI has explored the contributions of positive evidence and all of them have failed to isolate negative evidence from prompting for error repair, so it is impossible to determine the relative weight of these components in L2 development. Outside CALI, SLA research on this arena is also incipient, but evidence from some pioneering studies suggests that under some circumstances positive evidence and output modification in response to prompts may outweigh the provision of negative evidence (Leeman, 2003; Loewen, 2005; McDonough, 2005). Clearly, then, further research on this arena is certainly warranted. Additionally, no CALI studies thus far have utilized feedback in the oral mode, thus underexploiting the potential of the computer and hindering comparisons with the vast literature on feedback in oral interaction. Therefore, the evaluation of CALI tutors providing oral feedback is in order.

In sum, then, further research on computerized feedback should (1) compare the individual and combined contributions of negative evidence (and different degrees of explicitness), positive evidence, and prompting for repair on L2 development; (2) explore the potential of computerized feedback in the oral mode; and (3) investigate the effects of the variables that may mediate the efficacy of feedback types.

3. Mediating factors in L2 instruction

Arguably, the effects of pedagogical interventions, CALI- or non-CALI-based, are not exclusively determined by how something is taught but also by what is being taught and
to whom. In other words, in addition to controllable, manipulable variables such as provision of practice and feedback and type thereof, the effects of pedagogical interventions are also impacted by non-manipulable variables such as that which is being taught (e.g., the targeted linguistic form) and the individual differences of the recipients of the instruction (i.e., age, intelligence, learning (dis)ability, language aptitude, working memory, motivation, anxiety, emotion —see Robinson, 2002, for a complete overview of all these variables).

Possibly one of the main advantages of the interactive learning experience that computers can facilitate is that of accommodating to the learner’s individual profile. Whereas human teachers in regular classroom settings have to strive to carefully design and implement lesson plans and teaching strategies that include a wide spectrum of learners and do not disfavor any particular group, computers can easily customize their instruction thanks to the machine-to-student one-to-one ratio which is characteristic of CALI. On this token, future CALI systems (those anticipated by Warschauer & Healey’s 1998 ultimately attainable “intelligent CALL” phase) should preliminarily test their learners on a battery of cognitive and motivational measures to subsequently provide a personalized learning experience. For example, based on incipient evidence suggesting that working memory positively correlates to noticing of implicit forms of feedback (see Mackey, Philp, Egi, A., & Tatsumi, 2002, on recasts), CALI systems could adjust the explicitness of corrective feedback depending on each individual learner’s working memory scores.

To this regard, research on variables such as learners’ individual differences is of extreme value for the development of theoretically grounded and empirically tested CALI
systems. Nevertheless, as mentioned earlier, it should not be overlooked that, as shown by a growing body of research, the efficacy of the variables that define pedagogical interventions (e.g., presence of practice and feedback, and type thereof) is affected by the target form under instruction. This is precisely the focus of the following section.

3.1 Type of grammatical form

Experimental research in SLA is typically geared around a specific type of linguistic form. Most published studies include a section specifically devoted to the targeted structure used in the experiment, and many limitations sections ask readers not to overgeneralize their findings, calling for replication studies with other forms. However, the role of type of linguistic form remains largely unexplored in the SLA literature (DeKeyser, 2005: 11).

In recent years, several SLA studies specifically focused on pedagogical interventions, both preemptive and reactive, have considered the differential effects of type of linguistic form. In particular, at least four different strands of research have directly or indirectly addressed the construct, i.e., research on input processing (e.g., Greenslade, Bouden, & Sanz, 1999; VanPatten, 1989, 1990; Wong, 2001), literature on input enhancement (e.g., Leow, 1993; Leow, Egi, Nuevo, & Tsai, 2003; Shook, 1994), studies on the role of written output, by itself or as opposed to input-based learning approaches (DeKeyser & Sokalski, 1996; Izumi et al., 1999), and literature on the role of feedback in oral interaction (Ellis, 2007; Iwashita, 2003; Long et al., 1998; Nagata, 1993).

Of special interest to this dissertation is this latter strand of research, which investigates the efficacy of different types of feedback depending on the targeted
grammatical form under instruction. Overall, the studies by Nagata (1993), Long, et al. (1998), Iwashita (2003) and Ellis (2007) point to the following two conclusions: (1) type of linguistic form plays a role in the type of feedback which is naturally produced, a point which is widely accepted in descriptive studies. For example, Lyster (1998a) found that “type of response move varied according to the linguistic focus of the learner’s errors. Grammatical and phonological errors received more recasts, whereas lexical errors received more negotiation of form (i.e., elicitation).” And (2) type of linguistic form may determine the effectiveness of feedback. This is in line with Schachter (1991) and Nicholas, Lightbown, and Spada (2001), according to whom “the efficacy of recasts might be influenced by the targeted linguistic feature, the amount of attention brought to bear through the task and context, and the developmental readiness of the learner, among other factors”. Despite the general consensus shown by these empirical studies, their results should be taken with caution in light of some of their limitations, including, in some cases, a very low number of participants per cell, posttests based on too few items, and insufficient control of prior knowledge. Additionally, different claims have been made to explain how the difficulty of a target form may mediate the effects of a feedback type. For example, Nagata (1993) argues that the advantages of more explicit types of feedback arise for more complex forms, while Ammar (2008) and Ammar and Spada (2006) argue that implicit feedback is more likely to show advantages for complex structures. Consequently, then, further research is needed to not only confirm that language form plays a role in L2 acquisition, but also to elucidate how it does so in terms of its difficulty.
Chapter summary

Existing research on the use of computer technology for SLA has extensively focused on the role of the computer as a medium, where it may serve, e.g., as an exploratory environment to provide access to input, or as a means of communication that enables interaction among learners; however, very few studies have examined the role of computers as tutors, despite their booming sales and their increased presence in language laboratories (Hulstijn, 2000: 35). In this context, empirical research on computer tutors is crucial to determine whether they parallel or supersede traditional technologies and are thus worth the investment, and if so, research should clarify which of their features make the difference.

An incipient number of studies (Nagata, 1996; Nutta, 1998; Watts, 1989) have shown that computerized tutors can yield greater gains in L2 grammar development than other forms of traditional instruction using workbooks, VCR, and human teachers. However, these studies have not addressed the factors that contribute to this. Quite possibly, what distinguishes computers is their ability to engage learners in interactive practice (Nagata, 1996), whereas other non-interactive technologies can only provide exposure to input, or at their very best, display pre-recorded interactions. To date, one doctoral dissertation (Hsieh, 2007) has examined the effects of computerized practice versus exposure to it in L2 grammar development, with mixed results. Both conditions experienced statistically similar gains in written recognition and production of Spanish gustar constructions, while for oral production practice was more beneficial. Outside the realm of computerized instruction, the scant literature on learner’s agency, i.e., the compared effects of practice vs. exposure to practice (or the related construct of in-class participation) in L2 grammar
development does not clarify the picture, with some studies showing positive effects (Ellis & Rathbone, 1987; Mackey, 1999; Naiman et al., 1978; Seliger, 1977) and others showing no significant difference (R. Day, 1984; Ely, 1986; Muranoi, 2000). Additionally, as posed by Hsieh (2007) most studies have failed to isolate practice from feedback, so it is impossible to determine the relative contributions of these two variables.

Arguably, feedback in response to learners’ errors is the second component that makes up an interactive learning experience. While the existing CALI literature agrees in that feedback, regardless the type, is beneficial for L2 learning, the jury is still out as to the relative effectiveness of different types of feedback against each other. Some studies showed a significant advantage of more “explicit” types of feedback, where learners were given an explanation of the sources of their errors (Bowles, 2005; Lado, 2008; Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004b). In contrast, a comparable number of studies found no significant differences between more and less informative or explicit types, where learners were merely informed that their utterances were ill-formed (Camblor, 2006; Hsieh, 2007; Moreno, 2007; Sanz, 2004; Sanz & Morgan-Short, 2004). Finally, one doctoral dissertation showed significantly greater gains of less over more explicit feedback in oral delayed measures (Moreno, 2007). Quite possibly, the lack of consensus here may be due to factors both intrinsic and extrinsic to feedback. On the one hand, existing studies have failed to isolate the basic components of feedback moves, thus contributing to hardly comparable feedback operationalizations. These components are (1) negative evidence (that may present itself in different degrees of explicitness), (2)
positive evidence (or a model illustrating an L2 solution to a problem), and (3) prompting for error repair (i.e., requesting learners to fix their own mistakes).

On the other hand, a number of variables that may mediate the efficacy of feedback types have been overlooked. One prominent candidate is type of linguistic form. So far, several studies in oral interaction (Ellis, 2007; Iwashita, 2003; Long et al., 1998) have provided evidence showing that type of form may have an impact on the efficacy of different types of feedback. Hitherto, though, no published CALI study has addressed this issue, and the specifics on how the difficulty of a form mediates the efficacy of a feedback type are largely unexplored.

To wrap up then, future research should address whether the learning contributions of CALI result from the ability of computers to engage learners in interactive practice, or conversely, effective use of non-interactive media showcasing interactions may be equally effective. Also, empirical approaches to feedback should tease out the relative contributions of its three components, i.e., negative evidence (with different degrees of explicitness), positive evidence, and prompting for error repair. Finally, the role of the difficulty of the type of linguistic form under instruction should be elucidated.

From a theoretical point of view, all the constructs mentioned above, Learner’s Agency, Feedback, and grammatical Form, are often the source of various misgivings and misunderstandings. Addressing this gap, Chapter 3 provides a necessary theoretical discussion on these constructs, identifying key issues and proposing classification criteria.
CHAPTER 3.
THEORETICAL CONSIDERATIONS

In the field of SLA, theoretical approximations to the constructs of Learner’s Agency (Practice vs. Exposure to practice), Feedback, and linguistic Form have been scant or contradictory, and therefore a number of gaps and misunderstandings must be addressed. For example, with regards to *practice*, DeKeyser (2007: 1) stated:

> Practice gets a raw deal in the field of applied linguistics. Most lay-people simply assume that practice is a necessary condition for language learning without giving the concept much further thought, but many applied linguists eschew the term practice. For some, the word conjures up images of mind-numbing drills in the sweatshops of foreign language learning, while for others it means fun and games to appease students on Friday afternoons […] The concept of second language practice remains remarkably unexamined from a theoretical point of view. Misgivings and misunderstandings about practice abound and are often rooted in even deeper misunderstandings about what it is that language learners are supposed to learn.

Similarly, the term *exposure* has traditionally been conceived of in a very restricted way, i.e., the Krashenian sense, where it refers to the learner’s reception of positive evidence or models of correct input in the L2. However, as posed by Mackey (1999: 559) in many classrooms and naturalistic contexts learners are often exposed to the interactions of others, thus receiving exposure not only to positive evidence but also the output of other peers, the corrective feedback they get in return, and the subsequent input and/or output modifications. Yet, very little effort has been devoted to analyze the psycholinguistic processes that this exposure may trigger.

With regards to *feedback*, authors have often manipulated very narrow operational definitions, focusing exclusively on the explicitness of the negative evidence that feedback provides and ignoring other important components such as positive evidence
and prompting for error repair (Leeman, 2000; McDonough, 2005; Ranta & Lyster, 2007). Moreover, even the term “explicitness” has been differently understood in the literature. For some authors it refers to how overtly the tutor signals the occurrence of an error, while for others it means how much detail is provided to explain the reason of the error (see, e.g., Leeman, 2000).

Finally, as far as linguistic form is concerned, the picture is also pretty blurry. Some linguists have tackled the construct from the perspective of the L1, proposing taxonomies of forms based on their “order of acquisition;” others have focused on the L2 (in relation to the L1), and thus have attempted to categorize grammatical structures in terms of their relative “difficulty,” while a third group of researchers have looked at the interlingua, introducing notions such as “learnability,” “fossilization,” and “ultimate attainment.” From the perspective of instructed SLA, perhaps the most interesting notion is that of difficulty, but as pointed out by DeKeyser (2005: 2), “[e]ven a cursory glance at some well-known discussions of what is difficult in L2 acquisition shows how tricky this concept is,” to the point that many “researchers, recognizing the difficulty of defining difficulty, have avoided a theoretical conceptualization altogether” (p. 4).

On this note, the goal of this chapter is to present some basic theoretical considerations on the constructs of Practice (and Exposure to practice), Feedback, and linguistic Form, identifying key issues and proposing classification criteria that will be implemented in the review of empirical literature provided in Chapter 4.
1. Learner’s Agency

This section is divided into two parts. Section 1.1 tackles the construct of Practice, while Section 1.2 focuses on Exposure to practice. A comprehensive summary of both sections is provided in section 2.3.

1.1 Practice

The present section consists of two parts. Subsection 1.1.1 provides a definition of practice in general, focusing on different aspects of its multi-layered nature. Next, subsection 1.1.2 specifically focuses on the different ways in which practice may materialize in a learning environment, distinguishing among often-confounded constructs such as tasks, activities, exercises, and drills.

1.1.1 Toward a definition of practice

In the next sections the construct of practice is contemplated from various angles. Section 1.1.1.1 analyzes how the construct has been defined in the fields of cognitive psychology, educational psychology, and applied linguistics. Next, section 1.1.1.2 draws a necessary line between incidental and deliberate practice. Zooming in on the notion of deliberate practice, section 1.1.1.3 discusses its implementation in the classroom environment, distinguishing between meaning- and form-focused practice. After that, section 1.1.1.4 briefly discusses the different practice features that must be attended to when designing examples of deliberate practice. Finally, section 1.1.1.5 specifically explores the notion of practice mode, distinguishing between receptive and productive practice and their
different variants. A comprehensive summary of all sections is provided at the end in section 1.1.1.6.

1.1.1.1 Definitions in cognitive psychology, educational psychology, and SLA

Searching for a definition of *practice*, DeKeyser (2007: 2-8) presented a brief overview of the notion in three different arenas —cognitive psychology, educational psychology, and applied linguistics. In his review of the cognitive psychology literature, DeKeyser pinpointed the roots of the notion of practice in the study of skill acquisition, where practice is agreed to gradually reduce reaction times and error rates on a given task. This opens up two questions: first, ‘what is a task?’ and second, ‘why or how does practice promote better reaction time and error rates?’ The first question has a direct impact on the definition of practice. Arguably, tasks consist of many components, so it is important to determine how narrowly components can be defined, and also, whether practice of such components separately is better or worse than practice of the whole task. In turn, the second question taps into the mental processes that are posited to be triggered throughout practice, which according to skill acquisition theory involve the conversion of declarative knowledge (i.e., knowledge of rules) to procedural knowledge (i.e., behavioral routines) and its ulterior conversion into automatized knowledge (i.e., final stage of fully spontaneous, effortless, errorless use of a rule that was initially acquired declaratively, often without being aware of it anymore). Answering this question is also important for the instantiation of practice, as research suggests that practice may differently affect the knowledge conversion process depending on the specific skill that is being practiced (L2 comprehension or production). Hitherto, according to DeKeyser, research contends that,
whereas the common part of declarative knowledge can be reinforced to some extent by practice in either skill, procedural knowledge is highly skill-specific (it cannot easily be transferred from one skill to another) and thus separate practice in each skill is required.

In his revision of the notion of practice in educational psychology, DeKeyser observed how educators, like cognitive psychologists, have long envisaged practice as a requirement for learning and retention. He then rescued the term “deliberate practice,” which according to Bransford, Brown, and Cocking (1999: 166) necessarily involves the presence of a tutor, human or computer-based, to help learners rehearse appropriate practices aimed at reaching real-world performance criteria. Finally, he noted the precisions made by authors like Legge (1986) and Haskell (2001) between the terms practice and training, where the latter is respectively conceived of as either “a scheme to assist learning” or a sequence of “drilling or teaching recipes” (p. 5-6).

Finally, in the field of applied linguistics theoretical approximations to the notion of practice have often been avoided (cf., however, recent work by Sanz & Morgan-Short, 2005). According to DeKeyser, one of the first authors to provide an explicit definition was Ellis (1993: 109), for whom 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.

This definition, according to DeKeyser, poses a number of problems. First, it places a strong emphasis on production over comprehension (“…producing targeted structures”), which clashes with much theorizing in applied linguistics, especially VanPatten’s theory of input processing, according to which interpretation practice can also lead to
proceduralization. And secondly, this definition, by placing a strong emphasis on implicit knowledge (“…to develop fully proceduralized implicit knowledge”), may be (mis)interpreted as being in line with the Chomskyan idea that practice is aimed at improving performance rather than teaching competence, which is something innate or implicitly acquired and hence cannot be taught. This collides with the theory of skill acquisition in cognitive psychology, according to which practice can also be the pathway to transform a rule that has been initially presented explicitly to proceduralized and, ultimately, automatized knowledge. Interestingly, as DeKeyser noted, in recent revisitations of the notion Ellis (2003: 115) appears to have shifted toward a more comprehensive role of (output-focused) practice, where it is not only a mechanism to improve performance but also a way for learners to “construct and complexify their interlanguages” (see also Larsen-Freeman, 2003: 144).

Based on this review, DeKeyser (2007: 8) concluded:

> [I]n this book we define practice as specific activities in the second language engaged in systematically, deliberately, with the goal of developing knowledge of and skills in the second language.

Arguably, DeKeyser’s definition presents three strengths: (1) unlike Ellis’s (1993) definition, it is not restricted to one particular skill (comprehension or production); (2) it envisages a double role of practice as a way of “rescuing” and “developing” previously learned knowledge but also of “engendering” new knowledge; and (3) it is by default *deliberate*, which in DeKeyser’s terms presupposes a goal-orientation (cf. Bransford et al., 1999: 166, where “deliberate” refers to the presence of a tutor), in contrast to *incidental* practice, which “comes with activities engaged in frequently for work or...
personal routines” (DeKeyser, 2007: 309). DeKeyser’s definition however does not
directly address one of his opening observations when discussing the need of an explicit
definition of practice, that is, what constitutes a “specific activity,” a “task,” or whatever
instrument serves the intended goal of practice.

To this regard, Leow (2007: 21-2), albeit exclusively focused on receptive (as
opposed to productive) practice, offered a more detailed definition. Departing from an
attentional perspective, he stated:

[R]eceptive practice will be broadly defined in this chapter as follows: any exposure to
manipulated L2 input that provides not only various exemplars of 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. The general assumption is that via receptive practice, learners pay attention to the
targeted linguistic data in the input while processing the grammatical information substantially
enough to be capable of recognizing, interpreting, and/or producing such forms or structures after
exposure.

This definition therefore envisages receptive practice as a two-step process: (1) a pre-
requisite, i.e., exposure to L2 input, and (2) performing some kind of productive or
nonproductive activity with the L2 input that the learner has been exposed to (e.g.,
selecting one out of two options, completing a problem-solving task, translating). More
importantly, however, it brings forth into the picture two crucial components: attention
and targeted linguistic data. Thus, receptive practice in Leow’s terms cannot be defined
without considering the language learning goal, the specific L2 data (e.g., phonetics,
morphology, syntax, vocabulary, pragmatics, and more specifically, the particular
targeted element) that is intended to be practiced and developed, and which, based on
current SLA theories, needs to be attended to in order for any development to occur (see,
Interestingly, the absence or presence of targeted linguistic data, as envisaged by Leow, or more broadly, a communicative goal, as contemplated by DeKeyser, draws a very necessary line when defining the concept of practice, i.e., that which separates *incidental* from *deliberate* practice. That is the focus of the next section.

1.1.1.2 Incidental vs. deliberate practice

The definitions of practice by DeKeyser (2007: 8) and Leow (2007: 21-2) discussed above clearly refer to *deliberate* practice, since both of them contemplate a learning goal. Arguably, there is not such thing as deliberate foreign/second language practice per se, but rather, practice on, e.g., how to order a meal at a restaurant (communicative goal), or, alternatively, practice on the Spanish present indicative or names of foods (targeted linguistic form). Also, as noted by DeKeyser (2007: 5), the term “deliberate” in relation to practice has also been understood in a different way. For example, for Bransford et al. (1999: 166) deliberate practice necessarily involves the presence of a tutor. Yet, regardless the interpretation of deliberate practice (i.e., it requires the presence of a goal or a tutor), the role of the tutor cannot be left out of the equation, since it is the tutor who, through different pedagogical interventions (e.g., instructions to the task at hand, preliminary metalinguistic presentation, feedback) will attempt to direct the learners’ attention toward the communicative goal or targeted linguistic data at play. In this sense, asking learners to simply read a text is not an example of deliberate practice, but using pedagogical manipulations to direct learners’ attention to a communicative goal or
specific targeted linguistic data before or while they perform a productive or non-productive linguistic task or activity, is. The role of the tutor is crucial: the same pedagogical materials (e.g., a text) can be used as the basis of deliberate pronunciation practice (e.g., Spanish \( r \) vs. \( rr \) orthographic characters) and/or grammar practice (e.g., noun-adjective gender and number agreement), depending on the tutor’s intervention.

It should be noted, however, that despite the external attempts to manipulate learners’ attention deliberate practice may not always fulfill its intended purpose, as learners may allocate their attentional resources differently and redefine an activity or task as practice for other communicative goals or targeted linguistic data (see, e.g., Coughlan & Duff, 1994). For example, what may serve as pronunciation practice for one learner may serve as grammar practice for another, or even if both learners use an activity to practice grammar, they may focus on different grammatical structures. On this note, Swain (1995; 2001) pointed out that researchers should not evaluate the developmental effects of practice in terms of what they expect learners to do with it but rather in terms of what learners actually do, since that is largely determined by individual factors such as developmental readiness and allocation of attentional resources. Naturally, Swain’s point is even more relevant in the case of purely *incidental* (as opposed to deliberate) practice, where the only attempt is to use the L2, with no preset intended communicative goal or manipulation of targeted linguistic data. Therefore, engaging with a partner in casual conversation is still language practice, but of an incidental nature. Deliberate practice can present itself in four shapes, those of the four skills, and it does not require the presence of a tutor.
Both deliberate and incidental practice, however, are dynamic rather than static concepts. Thus, if, for example, in the course of casual conversation a tutor decides to focus on a particular problematic form and engages students in tasks or activities that require attending to that form and operating with it, then incidental practice is redefined as deliberate practice. Conversely, deliberate practice can give way to incidental practice, i.e., practice with no preset language development goals.

Now, in order to better understand the nature of deliberate and incidental practice and how they may be used in a pedagogical context, it seems adequate to review the different pedagogical approaches that may be adopted in the language classroom, i.e., meaning- vs. form-focused instruction. That is the goal of the next section.

### 1.1.1.3 Meaning vs. form-focused practice

A very concise and insightful review of L2 instruction types with special relevance to this discussion is provided by Loewen (2005). The following lines re-iterate his account, which may be best read in tandem with the visual representation in Figure 3.1 below.

Depending on the element on focus —meaning or form— L2 instruction can be classified into meaning-focused and form-focused instruction. Meaning-focused instruction involves tasks and activities in which learners are primarily concerned with the exchange of messages (Richards & Rodgers, 2001; Savignon, 2000). Conversely, form-focused instruction refers to “any pedagogical effort to draw learners’ attention to language either implicitly or explicitly” (Spada, 1997: 73). Within the realm of form-focused instruction, Long (1991b; 1996) further created a subdivision, i.e., so called “focus on formS” (FonFS) vs. “focus on form” (FonF), depending on whether the teacher
adopts a pre-emptive or reactive role. FonFS refers to a type of instruction that is pre-emptive on the part of the tutor, that is, instruction revolves around a syllabus containing specific linguistic items and structures to be taught (e.g., Monday: Spanish formal imperatives; Tuesday: present subjunctive, etc.). Conversely, the FonF approach is reactive, that is, it draws attention to specific linguistic elements “as they arise incidentally in lessons whose overriding focus is on meaning or communication” (Long, 1991b: 45-6).

As Loewen noted, the FonF/FonFS dichotomy was further expanded by Ellis (2001) to account for the fact that during FonF instruction attention to form can arise in a totally incidental manner or be the result of a preplanned pedagogical intervention. In *planned* FonF the tutor pre-selects certain linguistic items and draws attention to them as they occur in the course of a meaning-focused activity, either pre-emptively via input (e.g., input flood or input enhancement) or reactively in response to learners’ output (e.g., corrective feedback on errors). In contrast, in *incidental* FonF the linguistic items arise spontaneously in the course of meaning-focused activities.
At first glance, Ellis’s (2001) refinement of Long’s (1991b) FonF may strike some as unnecessary. If Ellis’s planned FonF involves the pre-selection, on the part of the tutor, of a number of targeted forms, how is this different then from Long’s FonFS? The absence or presence, respectively, of a laid-out syllabus, does not seem like a robust divergence. On the other hand, however, Ellis’s refinement can be productive when considering the weight of meaning or communication in the equation, and whether attention is raised preemptively on the part of the tutor or incidentally as the students complete an activity or task. Imagine two possible learning scenarios; disregard the syllabus criterion suggested by Long and assume that in both learning scenarios lessons are structured around a syllabus of preset targeted linguistic forms. The form of the day is, say, Spanish counterfactual conditional sentences (e.g., *Si yo fuera presidente de España, haría esto y lo otro* [“If I were the President of Spain, I would do this and that”]). In scenario A, communication has little weight. Students are engaged in a number of drills where they have to fill in the blanks with the appropriate form of the imperfect subjunctive or conditional tense. In contrast, in scenario B, students act as speechwriters for two
presidential candidates, outlining the policies that they would put into effect if they were elected President. The students then enact their campaigns assuming the role of presidential candidates and the instructor provides concurrent feedback on target. In this context, Ellis’s nuanced typology may come in handy to differentiate these two instructional approaches. Scenario A could be the most extreme representation of Long’s FonFS, whereas scenario B could be an example of Ellis’s planned FonF.

As shown above, many pedagogical decisions must be made to choose one approach or the other. The next section concisely identifies the main options that must be considered when designing practice-based instruction.

1.1.1.4 Designing deliberate practice

As discussed earlier, the presence or absence of a preset L2 learning goal categorizes practice as deliberate or incidental. Additionally, presetting a learning goal and devising the appropriate means to pursue it involves the presence of a tutor, human or computer-based, so deliberate and incidental practice are again separated by the tutoring factor.

Now, defining a language learning goal entails various specifications. For example, is the goal to develop communicative ability (the focus is on meaning conveyance), or is it L2 form development (the focus is on form)? Also, what is meant by development? The literature has typically distinguished between accuracy, fluency, and complexity, so objectives must be preset in this regard. Furthermore, if the focus is on L2 form development, then which is the relevant level (word-, sentence-, or discourse level), modality (written or aural), and mode (receptive, productive, interactive) under pursuit? These are all basic factors that language educators have to take into consideration when
designing deliberate practice. Table 3.1, excerpted from Ortega (2007: 182), provides a snapshot representation of these factors.

**Table 3.1** Some basic options in the design of L2 practice (from Ortega, 2007: 182)

<table>
<thead>
<tr>
<th>MODALITY</th>
<th>Spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Written</td>
</tr>
<tr>
<td>LANGUAGE DOMAIN</td>
<td>Word-level</td>
</tr>
<tr>
<td></td>
<td>Sentence-level</td>
</tr>
<tr>
<td></td>
<td>Discourse-level</td>
</tr>
<tr>
<td>MODE</td>
<td>Comprehension &lt; Interaction &gt; Production</td>
</tr>
<tr>
<td>DEGREE OF MEANINGFULNESS</td>
<td>Focus on meaning &lt; Focus on form &gt; Focus on forms</td>
</tr>
</tbody>
</table>

The majority of the factors above are either self-explanatory or have already been discussed. However, the notion of practice *mode* is often the source of various misunderstandings that need clarification. The next section delves further into this feature.

### 1.1.1.5 Receptive vs. productive practice

The formulation of historical landmarks in the SLA literature such as the Input Hypothesis, Processing Instruction, the Output Hypothesis, and the Interaction Hypothesis spawned a plethora of studies comparing the developmental effects of receptive vs. productive practice. These two constructs, however, call for important theoretical considerations, both of terminological bearing and of substance.

Although profusely used in the literature, labels such as receptive vs. productive practice (or their synonyms input- vs. output-based or -focused practice) have been criticized because they may lead to confusion. For example, output-focused practice
typically entails processing input first, and therefore in this particular case it could be
preferable to use a less misleading mathematical operator such as [+input, +output]
(VanPatten, 2004a: 43). There are, however, types of productive activities where input
processing is kept to a minimum. For example, in an activity like “write a speech in
which you outline the policies you would put into effect were you to become the next
president of Bolivia,” input processing is strictly limited to understanding the rubric or
instructions, and therefore this type of activity could be labeled as [-input, +output].
Arguably, there are many different types of receptive and productive activities, so for the
sake of systematicity there is a blatant need of precise definitions of these two practice
modes and a catalog of their possible instantiations. In principle, the division between
receptive and productive practice may be determined by the presence or absence of the
targeted linguistic form in the content of the activity and the demands imposed on the
learner. If the form is present in the learning activity, the practice is receptive; if,
conversely, the form is absent but the learner is required to use it, the practice is
productive. Now, the distinction between deliberate and incidental practice must be
brought into the equation. As pointed out by Leow (2007: 21-2), (deliberate) receptive
practice necessarily involves two steps: (1) a pre-requisite, i.e., exposure to the form, and
(2) performing some kind of productive or nonproductive activity involving the form the
learner has been exposed to. If this second step is missing, there is no “deliberate”
receptive practice; rather, all there is here is “incidental” receptive practice, which is
synonyms with exposure to L2 input. To this regard, as it follows from Leow’s (2007:
22) remarks, the experimental tasks used by many input enhancement studies are not
examples of deliberate receptive practice but, rather, incidental reading practice or
exposure to input. One particular example of this was the pedagogical task in Leow (2001), where learners were asked to read a text where the target structure, the Spanish imperative, had been underlined and bolded. Next, students were asked to complete a number of tasks, involving multiple choice recognition of the form and controlled written production; however, these tasks were used for testing rather than learning purposes. Consequently, the reading task in Leow (2001) is an example of incidental receptive practice or exposure to L2 input rather than deliberate receptive practice. If, however, Leow’s treatment had incorporated a post-reading task pushing understanding or production of the form as part of the treatment, the mere exposure to L2 input would have been transformed into deliberate receptive practice.

Now, within deliberate receptive practice at least three main subtypes may be distinguished, depending on the nature of the activity to be completed after exposure to the form, i.e., comprehension, interpretation, and recognition activities. Comprehension activities require the learner to understand the meaning of the target form. Examples of such activities include asking learners to dramatize a text, to translate the meaning of a sentence into the L1, or to answer open-ended questions, without providing a list of possible answers to select from. Crucially, comprehension questions that may be answered without understanding the meaning of the target form are not constitutive of deliberate receptive practice.

On the other hand, interpretation activities display the target form and then present the learner with a number of possible meanings, asking him/her to select the right one. Thus interpretation activities are distinguished from comprehension activities in that they do include the correct meaning of the form in the very rubric of the activity. This
meaning may be coded verbally or non-verbally. Examples of both types of interpretation activities in CALI are found in The Latin Project developed by Cristina Sanz, Harriet Bowden and Catherine Stafford (Sanz, Lin, Lado, Wood Bowden, & Stafford, 2009). In the activity in Figure 3.2 (left), learners are presented with one Latin sentence and two possible translations into English (the learners’ L1), one of which is correct. Conversely, in the activity in Figure 3.2 (right) the options are pictures.

**Figure 3.2** Examples of verbal and non-verbal interpretation practice (Sanz et al., 2009; Stafford, 2005)

As opposed to interpretation practice, where the learner is supposed to proceed from the L2 form to the meaning, *recognition* activities (a.k.a., “selected response” activities – Norris & Ortega, 2000: 440) require the learner to travel from meaning to form. In other words, in recognition activities the learner’s response consists of selecting, from a number of choices, the correct L2 form that best describes an L1 phrase, sentence, or picture. In CALI, the learner may be presented with a list of possible sentences differing in one or more constituents, or alternatively, with one sentence containing a drop-down
menu from which to select the right constituent. Figure 3.3 below provides an example of a recognition activity in CALI by Moreno (2007).

Figure 3.3 Example of recognition practice (Moreno, 2007)

There are, in addition, more sophisticated types of recognition activities in CALI. Observe, for example, the experimental task in Hsieh (2007), where learners are to navigate a series of forking paths with word or phrase options to sequentially translate a picture cue into its equivalent Spanish *gustar* structure. For example, to elicit the sentence *A Carmen le gustan los insectos* (“Carmen likes insects”), Hsieh’s tutor first displays a picture of a woman, a bunch of hearts, and a couple of insects. When clicking on a start button, two tree branches spread out, one with the subject constituent *Carmen* and another with the dative constituent *A Carmen* [“To Carmen”]. If the learner clicks on the wrong option, s/he runs into a dead end. If, however, s/he clicks on the right one, a checkmark appears and the tree branches out showing a number of options for the next constituent. The process repeats itself for every constituent until the correct sentence is
composed (see Figure 3.4). Clearly, this is a recognition task, but it differs from the one in Moreno (2007) in that it gradually focuses learners’ attention to form one constituent at a time.

**Figure 3.4** Example of sequential multiple recognition task (Hsieh, 2007)

Finally, other forms of recognition activities include so-called grammaticality judgments, where the learner infers the meaning of an L2 sentence, and based on that inference, s/he assesses whether or not that sentence is grammatically correct. Grammaticality judgment activities challenge somehow the previous definition of receptive practice in that the target form may not be present in all the items of the activity; however, they qualify as receptive practice because the learner does not need to provide the correct form. If this is the case, the grammaticality judgment activity turns into a revision activity, which as discussed below, qualifies as productive practice. Notably, grammaticality judgment activities are typically used for testing rather than practice purposes, but they can be used for both. Figure 3.5 below provides an example of grammaticality judgment test in CALI excerpted from The Latin Project (Sanz et al., 2009).
As opposed to receptive practice, productive practice per se does not involve exposure to L2 input that includes the targeted form. In productive practice, the target form is absent from the activity. The learner is thus exposed to either a visual stimulus, L1 input, or L2 input that does not contain the target form, and is then requested to produce L2 output involving that target form under more or less controlled and/or communicative situations, at either the word-, phrase-, sentence-, or paragraph-level.

Norris and Ortega (2000: 440) classify these activities into two blocks. “Constrained selected response” activities require the learner to produce the target form under highly regulated circumstances, ranging from the word to the phrase level, whereas “free constructed response” activities exert fewer constraints and give more agency to the learner (e.g., oral interviews, written compositions). Due to computational costs or constraints, production activities in CALI are most frequently of the first kind, usually involving drills, but as mentioned in section 1.1.2.2 below, drills may present themselves
in different shapes (i.e., mechanical, meaningful, or communicative). Figure 3.6 provides an example of word-level mechanical drill in CALI, excerpted from Nagata (1998a).

**Figure 3.6** Example of word-level mechanical drill (Nagata, 1998a)

Give a Japanese equivalent to the following verb.

(My supervisor) will buy it.

As opposed to word-level mechanical drills, in meaningful drills learners are provided with a context cue that they have to respond to. Figure 3.7 below provides an example of meaningful drill at the sentence level by Nagata (1998a).

**Figure 3.7** Example of sentence-level meaningful drill (Nagata, 1998a)

What would you say in Japanese if you were in the following situation[?] Use honorifics if the situation is appropriate.

Your boss is going to London on business. Ask him when he will return to America.

Examples of production activities beyond the sentence level are rare in CALI, due to computational costs and constraints. Nagata (1998a) attempted to achieve this by using a revision exercise, reproduced in Figure 3.8 below. In this exercise, the learner is asked to edit a text by reading it, deleting the words s/he wants to change, and typing in the correct words. However, there are two caveats here. First, while reading a text may be constitutive of text-level receptive practice, editing incorrect words is hardly constitutive of text-level productive practice. And second, it may be argued that this type of exercise involves not only production but also comprehension. In response to this, Nagata (1998a:
39) states: “the text is presented using non-honorific forms, so the target structures (honorifics) are not used for comprehension. That is, the target structures (honorifics) are not the ones to comprehend but the ones to be produced.” In any case, this type of revision exercise is hard to label, as it encompasses characteristics of both grammaticality judgments and word-level production.

**Figure 3.8** Example of revision practice (Nagata, 1998a)

The following is part of the note Mr. Mills wrote to his teacher. Revise the text using appropriate honorifics. Delete the words you want to change and type in the correct words.

今日パティのプランをします。
あした先生にそのプランを話します。
あしたがっこうにますけど、
先生はますか。

Additionally, Nagata (1998a) tried to elicit text level production by exposing learners to an aural narration and asking them to re-tell it in dialog form, thus eliciting the targeted forms (Japanese honorifics) without preliminarily exposing learners to them. In principle, this activity does qualify as deliberate productive practice. It should be noted, however, that the outcome of the activity depends on the comprehension of the L2 aural text, even if the text does not contain any exemplars of the target form. In order to avoid this, the L2 text could be replaced by an L1 text or visual cues (pictures or videos). All these tasks are genuine examples of text level production tasks; however, due to the high computational costs of analyzing L2 speakers’ free speech, the use of these tasks in CALI is usually limited to testing rather than practice purposes (see, e.g. the video retelling task in Sanz & Morgan-Short, 2004).
1.1.2 Practice into practice: Tasks, activities, exercises, and drills

Typically, definitions of deliberate practice envisage tasks, activities, exercises, and drills all as valid means through which practice may be materialized. However, since the particular means used largely affect the nature of the resulting practice, defining these means is in order. At present, tasks hold a central position in the SLA literature and pedagogy, so perhaps the easiest way to distinguish all these concepts is by starting out with what constitutes a task. To that aim, section 1.1.2.1 reviews some definitions of “task” and its defining components. Next, in contrast to the notion of task, section 1.2.2.2 draws some necessary precisions between often confused constructs such as activities, exercises, and drills.

1.1.2.1 Tasks

As argued by DeKeyser (2007: 2), the notion of task often generates doubts among second language instructors and scholars: “Is speaking French a task, or requesting a glass of water in French, or does using a conditional verb form to do this constitute a task?” In recent years, the field of SLA has witnessed a plethora of publications related to task-based learning and teaching, which spawned many different definitions of “tasks” (for monographic reviews see, e.g., Bygate, Skehan, & Swain, 2001; Crookes & Gass, 1993; Ellis, 2003; Language Teaching Research, 4(3), 2000; Lee, 2000; Skehan, 1998; Willis, 1996). According to Pica, Kanagy and Falodun (1993), most definitions of task share two features (emphasis in the original):

The first is that tasks are oriented toward goals. Participants are expected to arrive at an outcome and to carry out a task with a sense of what they need to accomplish through their talk or action. The second feature of task is work or activity. What this feature suggests is that participants take
an active role in carrying out a task, whether working alone or with other participants. In other words, a task is not an action carried out on task participants; rather, a task is an activity which participants, themselves, must carry out.

As Pica et al. (1993) noted, these criterial features fall short in that “almost any activity-generating, goal-oriented experience, even a classroom lesson or structured elicitation device, can get mislabeled as a task.” Therefore, based on the premise that “language is best learned and taught through interaction” (p. 10), they proposed a checklist to evaluate the potential contributions of tasks in terms of how they may promote interaction and, more specifically, trigger elicitation of its basic components, i.e., input comprehension, output production, and modifications thereof in response to feedback. Based on this checklist, Pica et al. proposed a typology of five main interaction task types: jigsaw, information gap, problem-solving, decision-making, and opinion exchange. Although Pica et al.’s (1993) task features checklist and typology are extremely valuable resources for both teachers (it helps them in making educated choices for lesson planning purposes) and researchers (it provides a paradigm to understand the published literature and contributes ideas for research design), it does not quite address our pursued goal of finding out what differentiates a task from other instantiations of practice such as activities, exercises, and drills.

To this regard, Ellis (2003: 16), based on an extensive review of the notion of task (Breen, 1989; Bygate et al., 2001; Crookes, 1986; J. F. Lee, 2000; Long, 1985; Nunan, 1989; Prabhu, 1987; Richards, Platt, & Weber, 1985; Skehan, 1996), presented a more specific definition:

A task is a workplan that requires learners to process language pragmatically in order to achieve an outcome that can be evaluated in terms of whether the correct or appropriate prepositional content has been conveyed. To this end, it requires them to give primary attention to meaning and to make use of their own linguistic resources, although the design of the task may predispose them
to choose particular forms. A task is intended to result in language use that bears a resemblance, direct or indirect, to the way language is used in the real world. Like other language activities, a task can engage productive or receptive, and oral or written skills, and also various cognitive processes.

Specifically, Ellis’s definition of “task” is constructed around seven task-dimensions or task-criterial features, namely:

(1) Teacher’s perspective: a task is a workplan, a series of steps that students are instructed to follow. This workplan is “successful” if students reach the expected outcome through meaning-focused communication.

(2) Scope: the scope of a task is to engage learners to use language pragmatically to convey meaning rather than to display language.

(3) Learner’s agency: the learner has freedom to choose the forms to convey meanings, and these forms cannot be specified in the rubric of the task.

(4) Authenticity: A task involves either “situational” or “interactional” authenticity, that is, it involves an activity that people usually carry out in the real world or, otherwise, it triggers language behavior comparable to real world tasks.

(5) Cognitive processing: A task engages not only linguistic but also extra-linguistic cognitive processes, ranging from the mere use of perceptual skills (e.g., to notice, detect, or identify the differences between two pictures) to deeper cognitive skills such as selecting, reasoning, classifying, sequencing, evaluating, and transforming information from one form of representation to another.

(6) Outcomes: Every task stipulates a non-linguistic outcome (e.g., to what extent two students are able to identify all the differences between two pictures).

(7) Language skills: A task can involve any of the four language skills.
Arguably, not all these dimensions are exclusive of tasks. For example, many exercises involve a workplan and can elicit any of the four language skills. Therefore these dimensions should be conceived of as a list of desiderata that can be used to determine whether a given activity is more “task-like” or “exercise-like” rather than a template to categorize activities into tasks or exercises. What are exercises, though? The next section examines this and other related constructs.

1.1.2.2 Activities, exercises, and drills

The terms activity, exercise, and drill appear profusely in the SLA literature on practice and task-based instruction; however, their definitions are rare to find. Departing from Ellis’s (2003) previous definition of task, it is perhaps easiest to define the term exercise first, since it is typically conceived of as the opposite extreme. Thus, an exercise would be defined as a language learning activity that has a primary focus on form over meaning, does not give total agency to the learner to choose the forms, lacks situational or interactional authenticity, does not require extra-linguistic processing, and does not result in the production of a non-linguistic outcome. Again, as pointed out by Ellis, not all language activities can be clearly defined as tasks or exercises, since they share features of both, and hence it is sometimes necessary to classify language activities as more task-like or exercise-like.

Now, moving on to the term activity, in this discussion it has been used as an umbrella term to refer to any instantiation of language learning practice. That is precisely the trend in the published literature, where activity is oftentimes used to encompass both tasks and exercises, and even appears as a wildcard word in their very definitions. The
literature abounds in definitions of tasks along the lines of “a task is an activity that…” (e.g., Bygate et al., 2001; Crookes, 1986; J. F. Lee, 2000; Pica et al., 1993; Prabhu, 1987; Richards et al., 1985; Skehan, 1996). Similarly, the word activity is sometimes used interchangeably with exercise (Ellis, 2003: 3). Cobb and Lovick (2007) provide a good analysis of the status quo of activities vs. tasks vs. exercises under the present paradigm of task-based language teaching:

Researchers generally make a distinction between tasks and so-called non-tasks, which are mainly two types of activities: exercises and (unstructured) conversation. In this regard, it makes sense to refer to everything the students are asked to do in the classroom as “activities,” and among these to distinguish between activities that meet the established characteristics for tasks and those that do not.

It should be noticed, however, that the term activity acquires a different meaning under the framework of Vygotsky’s Sociocultural theory of mind and its spin-off, Leont’ev’s Activity theory. Thus, as Coughlan and Duff elaborate in their 1994 article entitled “Same task, different activities,” an activity is the actual behavior produced as a result of doing a task. To this regard, the same task may be performed differently by different learners or even by the same learner in different moments, thereby translating into different activities (this interpretation of “activity” has elsewhere been referred to as “task as a process,” as opposed to “task as a workplan” — cf. Ellis, 2003).

Finally, one last term that appears very frequently in the literature as an instantiation of language practice is drill. As DeKeyser (2007: 10) noted, drills have been historically perceived very differently depending on the teaching philosophy in vogue. They were first massively advocated by the supporters of the form-focused audiolingual method, later “demonized” with the outburst of meaning-focused communicative language.
teaching, and recently resuscitated by the proponents of the Focus on forms approach. Today, the usefulness or appropriateness of drills is still a matter of heated discussion, as illustrated by the recent debate between Wong and VanPatten (2003; 2004) on the “nay” side, and Leaver et al. (2004) on the “yea” side. As noted by DeKeyser (2007: 10-11), this debate cannot actually be fully understood without underscoring that there are very different kinds of drills, and that far too frequently drills are understood in an overgeneralized, reductionistic sense. This point is illustrated by the very titles of Wong and VanPatten’s 2003 article (“The evidence is IN: Drills are OUT”) and Leaver et al.’s 2004 response (“Apples and oranges are both fruit, but they don’t taste the same”). To begin with, then, what are drills? If, as discussed earlier, everything that the students are asked to do in the classroom can be labeled as an “activity,” and activities are further classifiable as “tasks” or “exercises” depending on their features, where are drills in this taxonomy? Basically, drills conform, point by point, to the definition of “exercise” provided above: they prioritize form over meaning, restrict learner’s agency, lack authenticity, do not necessarily trigger extra-linguistic processing, and do not yield a non-linguistic outcome. Is there, then, a conceptual difference between drills and exercises? In my view, there is none, and drill is just a synonym of exercise, one however which is often reminiscent of the days of the audiolingual method and thus negatively loaded. As Leffa (2007) puts it:

The English language has one word for exercise and another for drill, with a more military and mechanical connotation. Using the word drill for classroom activities that could be described as meaningful and communicative is more disqualifying and seems to be the term preferred by some critics, instead of exercise.
Other authors, however, have pinpointed the difference between drills and exercises in quantitative terms, that is, amount of practice. For example, Helmke and So (1980) conceived drills as a concatenated sequence of exercises (emphasis in the original):

We introduced two different pronunciation practice conditions: The *experimental condition* [drills] which provides the learner with pronunciation drills so he can produce contrasting sounds *as often as time allows*, and the *control condition* [exercises], which gives the learner a chance to produce each minimal pair *only once*.

In sum, then, the terms exercise and drill point to the same reality, and the preferred use of one over the other depends on how they are subjectively perceived and whether they are considered in isolation or as a practice battery. Now, as mentioned earlier, there are different kinds of drills. A traditional classification can be found in Paulston (1970; 1972) and Paulston and Bruder (1976), according to whom there are three main types, i.e., mechanical, meaningful, and communicative drills, briefly discussed below. This description may be best read in tandem with the table in Appendix 3.1, including illustrative examples.

The target of numerous critics, *mechanical drills* exert total control of the response, since they accept only one correct answer that has to be expressed through one specific form. These drills enforce neither attention to, nor understanding of, the L2 and can be successfully completed by mere superficial analogy. The most extreme instantiation of this type of drill is repetition and “mim-mem.” According to Paulston (1970: 189), this type of drills originates from the classical Skinnerian conception that learning responds to habit formation, and is specially suited for practicing mechanical associations (e.g., adjective-noun agreement, verb endings).
In contrast to mechanical drills, in order to complete *meaningful drills* the learner has to understand the input both grammatically and semantically, making appropriate form-meaning connections, and responding accordingly. There is still control over the response, in that there is only one correct answer, but unlike the case of mechanical drills here the answer can be expressed through more than one form. This can be constrained by providing models. Drills of this type include comprehension questions based on written or aural materials or the classroom context.

Finally, *communicative drills* go one step further, requiring the learner to contribute real-world information that the other interactants (teacher or classmates) cannot infer from the learning materials or the classroom context. The response in these drills is consequently not controlled, since the learner has free choice of answer and may express it through a number of forms. As Paulston (1970: 191) puts it:

> Whatever control there is lies in the stimulus. ‘What did you have for breakfast?’ is likely to limit the topic to the edible world but not necessarily. ‘I overslept and skipped breakfast so I wouldn’t miss the bus’ is an answer I have heard more than once. It still remains a drill rather than free communication because we are still within the realm of the cue-response pattern.

### 1.2 Exposure

For many years now, and most likely as a result of Krashen’s influential Comprehensible Input Hypothesis (1984), the notion of *exposure* is typically associated to reading, or listening to, positive evidence (i.e., instances of correct L2 input). However, as pointed out by Mackey (1999: 559) in many classrooms and naturalistic contexts learners are often exposed to the interactions of others, receiving much richer input. After three decades of empirical work under the Interaction Approach, it has been shown that participation in interaction is more facilitative than exposure to L2 input alone (Norris &
Ortega, 2000 - more on this in Chapter 4) because in addition to pushing learners to comprehend input, it engages them in processes that have been proved to be developmentally helpful, such as the processing of input via selective attention, the production of output, the processing of feedback, the revision of hypotheses in reaction to it, and the subsequent modifications of input or output by the tutor or learner, respectively. While these processes have all been proved to aid L2 development, no claims have been made as to the role of the learner’s agency in them. In other words, it remains unclear whether the active participation in processes like output production and modification in response to feedback is a prerequisite for learning to occur, or if on the contrary, exposure to these interactional moves as performed by other peers may be equally helpful.

To date, very little effort has been devoted to analyzing the role of performing a task versus observing how others perform it. Arguably, from a psycholinguistic perspective the differences between these two processes are likely moderated by the type of practice at hand, including variables such as mode (receptive vs. productive), modality (written vs. oral), degree of meaningfulness (focus on meaning vs. form), and language domain (word-, sentence-, or discourse-level). For example, it may be argued that the gap between performing and observing is narrower for receptive than productive practice. Typically, in an input-focused task the role of a practicing learner consists of reading or listening to a restricted number of L2 choices and choosing one, subsequently receiving feedback on the adequacy of the selection. From the point of view of an observer the only difference is the selection part, but the usually very controlled nature of input-based tasks is likely to encourage subvocal participation anyway. To this effect, it may be easy for an
observer to retranslate the feedback provided to a performer as his/her own feedback via a simple elimination process. On the other hand, in an output-focused task there may be a much greater number of possible mismatches between a practicing participant and an observer that decides to complete the task subvocally. In relation to this, the modality of the task may also play a mediating role. For example, an observer that reads the output of a peer and the feedback he gets in response may use a number of strategies, such as backtracking, to identify the mismatches. Finally, the same applies to the degree of meaningfulness and the language domain of the task: the more controlled the task is, the easier it is for an observer to complete a task subvocally and to benefit from the feedback delivered to a performer.

**Section summary: Learner’s Agency**

In the present dissertation *Learner’s Agency* refers to the comparative effects of performing vs. observing second language grammar practice. Second language practice is understood in a broad sense and two restricted senses. Practice in a broad, general sense involves the process of using an L2 either receptively, productively, or both. More restrictively, two different instantiations of practice can be distinguished. *Incidental practice* involves using an L2 with no preset language learning goal and does not require the presence of a tutor. In contrast, *deliberate practice* is here understood as the process of using an L2, either receptively, productively, or both, via drills, exercises, activities or tasks with the ultimate goal of developing knowledge and skill of meaning and/or form conveyance. Deliberate practice presupposes the existence of a tutor, before practice (in
the form of a material developer), during practice (as a feedback provider), and/or after practice (as an evaluator).

In the language classroom, practice can take two main shapes, depending on the element on focus, meaning or form. Incidental practice is by nature meaning-focused. In turn, deliberate practice is necessarily form-focused, but there are different subtypes depending on the weight assigned to meaning. FonFS instruction (Long, 1991b; 1996) presupposes the pre-emptive design of a developmental sequence on the part of the tutor, and thus may reduce focus on meaning. In contrast, in FonF the attention to specific forms is drawn reactively, as they occur in communication. Additionally, Ellis (2001) distinguished between planned and incidental FonF, depending on whether or not the tutor designs his or her tasks with a predetermined set of linguistic targets in mind.

After deciding on the nature of practice (incidental or deliberate) and its focus (meaning or form), the tutor must determine additional factors, such as what to develop (language accuracy, fluency, or complexity), at what level (word-, sentence-, or discourse level), in which modality (written or aural), and through which mode (receptive, productive, or interactive). The mode factor is particularly important, given the current debate on whether or not declarative and procedural knowledge are transferable from one skill to another (DeKeyser, 2007). Receptive practice can present itself in the form of interpretation, comprehension, recognition, and grammaticality judgment activities. Interpretation and comprehension activities present an L2 form to the learner and ask him/her to define its meaning, with or without options to select from, respectively; recognition activities present a meaning and ask the learner to select the correct L2 form from a number of choices; and grammaticality judgment activities require the learner to
decide whether a given form is possible in the L2. In turn, productive practice can materialize in the form of “constrained selected response” and “free constructed response” activities, depending on whether or not they require the learner to produce the target form under highly regulated circumstances (Norris & Ortega, 2000: 440).

Notably, definitions of practice usually include or point to different ways in which practice may materialize, such as activities, tasks, exercises, and drills. These terms have received either vague or diverging definitions, thus calling for terminological normalization. Generally, the term activity is used as an umbrella term to refer to everything that occurs in the classroom (Cobb & Lovick, 2007). When practice is deliberate (i.e., there is a specific learning goal determined by the tutor), activities can be further refined into tasks vs. non-tasks or exercises. According to Ellis (2003), tasks are characterized by having a primary focus on meaning over form, giving total agency to the learner to choose the forms, entailing situational or interactional authenticity, requiring extra-linguistic processing, and resulting in the production of a non-linguistic outcome. Conversely, exercises presuppose the negation of these conditions. Nevertheless, not all language activities can be clearly categorized as tasks or exercises, since they may share features of both, and consequently it often makes more sense to classify language activities as task-like or exercise-like. Additionally, the literature on practice is populated by another term, drill, which as pointed out by DeKeyser (2007: 10) has been historically perceived very differently depending on the teaching philosophy en vogue. Arguably, the term drill is synonyms with exercise, as it shares the same features, so the preferred use of one over the other depends on how they are subjectively perceived (drills often having a negative connotation - Leffa, 2007) and whether they are considered in isolation or as a
practice battery (drills often involving more repetition - Helmke & So, 1980). Now, as DeKeyser (2007: 10-11) notes, drills are often understood in a too reductionistic way, and the debate they generate cannot actually be fully understood without underscoring that there are very different kinds of drills. Specifically, as shown by Paulston and colleagues (Paulston, 1970; 1972; Paulston & Bruder, 1976), drills can be divided into so-called mechanical, meaningful, and communicative drills, depending on the extent to which the correct answer is controlled and whether or not they enforce attention to, or understanding of, the L2 input.

As for exposure, the term has almost always been exclusively understood in a Krashenian sense, i.e., reading, or listening to, positive evidence. However, in many classroom contexts learners are not only exposed to positive evidence but also the practice performed by other peers (Mackey, 1999: 559). Research under the scope of the Interaction Approach has shown that participation in interaction is more facilitative than exposure to L2 input alone (Norris & Ortega, 2000 - more on this in Chapter 4) because in addition to pushing learners to comprehend input, it engages them in processes that have been proved to be developmentally helpful, such as the processing of input via selective attention, the production of output, the processing of feedback, the revision of hypotheses in reaction to it, and the subsequent modifications of input or output by the tutor or learner, respectively. However, it remains unclear whether the benefits of these developmentally helpful processes are moderated by the role of the learner, that is, whether active practice versus observation of practice by others translate into different learning outcomes. From a psycholinguistic perspective, several variables may contribute to narrow or widen the gap between these two cognitive processes and their facilitative
effects, including type of practice, mode (receptive vs. productive), modality (written vs. oral), degree of meaningfulness (focus on meaning vs. form), and language domain (word-, sentence-, or discourse-level). Therefore, future empirical investigations on learner’s agency (or the comparative effects of practice versus exposure to practice) must take all these factors into consideration.

2. Feedback

In the following sections the construct of feedback is discussed from four different angles. In section 2.1, feedback is placed within a classification of types of linguistic data for SLA, thus distinguishing it from other pedagogical interventions such as provision of modified input and preemptive metalinguistic explanation. Section 2.2 then presents a definition of feedback and an account of the specific components in feedback moves, i.e., negative evidence, positive evidence, and prompting for error repair. Section 2.3 reviews a number of feedback taxonomies and presents a comprehensive one that integrates all three components of feedback moves. Section 2.4 revisits Lyster and Ranta’s (1997) popular catalog of feedback labels, placing each particular type within the proposed taxonomy. Section 2.5 presents a comprehensive summary.

2.1 The place of feedback in SLA

In the process of learning a second language in either naturalistic, classroom, or laboratory settings learners can be exposed to different types of input or linguistic data. One of the earliest classifications of such potentially available data was published by Long and Robinson (1998), reproduced in Figure 3.9 below.
According to Long and Robinson’s taxonomy, linguistic data can first be divided into two main categories, i.e., positive and negative evidence. A number of authors (e.g., Chomsky, 1981; Gass, 1997, 2003; White, 1989) additionally noted the existence of a third type of evidence, i.e., indirect negative evidence (Plough, 1994); however this type will not be discussed here since, as pointed out by Gass (2003: 249), it is “the least relevant” or available type of evidence in SLA (see also White, 1989: 107). Going back to Long and Robinson’s binary distinction then, positive evidence is constituted by examples of what is possible in a language, that is, well-formed input, whereas negative evidence refers to information that learners receive about what is not possible in a language. Both positive and negative evidence can be divided into subtypes depending on the specific nature of their instantiations. Specifically, positive evidence can be divided into authentic and modified input, depending on whether or not the input that is provided to the learner has been pedagogically altered. One type of authentic input is often referred to in the empirical literature as baseline input, i.e., input that the tutor provides to the learners without necessarily taking into consideration their specific needs. This input may
be extracted from *realia* or may be generated by the tutor, either on the fly or based on e.g., observations of interactions among native speakers. Conversely, the tutor may incorporate learners’ needs into the equation and use so-called modified input, which can be subsequently classified as simplified or elaborated, depending on the strategy that the tutor or more competent interlocutor adopts to try to make the input more salient to the learner. For example, parents often simplify their input when speaking so-called *motherese* to their babies, as do some teachers when using so-called *foreigner-talk*.

In addition to positive evidence or instances of well-formed input, learners can be exposed to negative evidence or information about what does not work in a language. This type of evidence can be provided preemptively or reactively. Negative evidence is preemptive when it is delivered before the learner produces an error. For example, based on an analysis of the learners’ L1, the tutor may decide to give a presentation including cases of what is not acceptable in the L2. This is often referred to as aprioristic grammar explanation in the empirical literature. Conversely, negative evidence can be provided in response to learners’ errors. This type of evidence, which Long refers to as “reactive negative evidence” is precisely the focus of the present dissertation, i.e., “corrective feedback,” which is specifically discussed in section 2.1.2 below.

Overall, Long and Robinson’s taxonomy constitutes the earliest significant attempt to classify the different types of data that a language learner may be exposed to on the basis of useful dichotomies such as type of evidence (positive vs. negative), timing (preemptive vs. reactive), and explicitness of negative evidence (explicit vs. implicit). However, this taxonomy presents a number of problems on both branches of the tree, positive and
negative evidence, which have an important bearing on the understanding of the construct of feedback (see also Leeman, 2000).

Starting with positive evidence, timing is not used as a classifying criterion there. Yet, positive evidence, and modified input in particular, can be generated preemptively or reactively. Indeed, the empirical literature distinguishes between premodified input (i.e., input that has been generated preemptively by observing previous interactions with other learners) and interactionally modified input (i.e., input that is modified concurrently in reaction to learners’ specific requests for clarification). More importantly, interactionally modified input is not the only type of reactive positive evidence, as in response to a well-formed utterance the tutor may decide to produce so-called positive feedback, i.e., a message acknowledging the learner’s success, including metalinguistic comments and/or a repetition of the successful utterance. Therefore, Long and Robinson’s taxonomy of positive evidence could be tentatively re-graphed as follows:

**Figure 3.10** Long and Robinson’s (1998) taxonomy of positive evidence, revisited
As far as feedback is concerned, then, Long and Robinson’s taxonomy needs to be extended to account for the fact that feedback can not only be a representation of reactive negative evidence but also reactive positive evidence. Now, as mentioned earlier, Long and Robinson’s taxonomy of negative evidence also needs amendments in a number of areas. For example, although this taxonomy allows us to distinguish among different feedback types in terms of their explicitness (e.g., recasts vs. overt error correction), and different representations of the same feedback type (e.g., simple vs. complex recasts), it is not fine-grained enough to categorize all the feedback types that have been researched to date. Also, this taxonomy classifies all feedback types as instances of negative evidence, thus obscuring the fact that some feedback types (i.e., recasts and explicit corrections – cf. e.g., Lyster & Ranta, 1997) also convey positive evidence (Leeman, 2000). Section 2.2 below and its subsections elaborate on this.

Besides the mentioned problems with Long and Robinson’s characterization of positive and negative evidence there is an additional problem that affects both nodes. As pointed out by Leeman (2000), Long’s taxonomy typifies grammar rule presentation as a daughter node of preemptive negative evidence, which is problematic for two reasons. First, grammar rules may be presented both preemptively and reactively throughout the course of practice. For example, in response to a student’s ill-formed sentence like *La casa rojo (“The red-MASC-SING house-FEM-SING”) the tutor may say “Your sentence is incorrect because in Spanish nouns and adjectives must agree in gender”. Secondly, while provision of grammar rules or metalinguistic information is usually accompanied by examples of negative evidence (provided by the tutor or generated by the learner), it can also be accompanied by models, i.e., examples of positive evidence. Furthermore,
although this is not typically the case, grammar rules can be generic, that is, presented
without evidence of any kind, positive or negative. Consequently, according to Leeman,
since grammar explanation can be preemptive or reactive, and be accompanied by
positive, negative, or zero evidence, it should be considered as an independent node when
classifying input types in SLA.

In light of the previous discussion, Leeman (2000: 149) concluded:

Long’s classification scheme is a significant advance in the classification of input. It not only
incorporates a multitude of distinctions among various types of negative feedback, but also
includes various types of positive evidence. Nonetheless, the distinction between positive and
negative evidence is portrayed as a mutually exclusive contrast, and thus recasts are [inaccurately]
classified only as negative evidence. Moreover, the hierarchical nature of this classification
scheme may obscure similarities between input types on distant nodes.

To address these caveats, Leeman proposed a new taxonomy where input types are
defined in terms of three components that are considered independently rather than in
hierarchical relation: positive evidence, negative evidence, and metalinguistic rule
presentation (more on this in section 2.3). Arguably, Leeman’s classification scheme is
much more productive, allowing for a wider range of feedback types. On a different note,
however, recent literature (Nassaji, 2007; Ranta & Lyster, 2007) contends that feedback
moves should be classified depending on whether or not they push learners to repair their
errors, since this may push them to revise their hypotheses and generate new ones,
thereby bearing an impact on interlanguage development. In sum, then, a classification of
feedback types should be geared around the provision or absence of various components,
i.e., negative evidence, metalinguistic rules, positive evidence, and prompts or requests
for error repair. In the next sections these components are described separately,
concluding with a comprehensive taxonomy of feedback that allows for teasing out and comparing the different feedback types that have been empirically studied so far.

### 2.2 Defining corrective feedback and its components

In the field of cognitive psychology the term *feedback* appears as a construct in information processing models where it refers to “a mechanism by which the processor (or learner) is provided with information regarding the success or failure of a given process” (Leeman, 2000: 136). By definition, then, feedback is (1) reactive, i.e., contingent upon the execution of a process on the part of the learner; and (2) positive or negative, depending on whether it praises a successful execution or it signals the occurrence of an error. In the specific realm of SLA, process execution refers to language use, and although the empirical investigation of positive feedback is certainly warranted, it is negative feedback that has captured major attention over the last decades, most likely to address chomskian/nativist claims on the sufficiency of positive evidence for language acquisition. The present section deals with negative feedback exclusively.

From a terminological perspective, as several authors (e.g., El Tatawy, 2002; Leeman, 2000; Schachter, 1991) have pointed out, the literature abounds in terms that are often used interchangeably to refer to negative feedback, such as *negative evidence, error correction, corrective feedback*, and the like. Hence, for the sake of terminological consistency it may be useful to clarify the intended meanings of these terms. First, as discussed earlier, the term *negative evidence* generally refers to information of what is unacceptable in a language, but it does not necessarily capture the reactive nature of feedback. For example, an a priori grammar explanation containing examples of
 unacceptable utterances qualifies as negative evidence, while it clearly is not feedback. Conversely, the term *error correction* does capture the intended meanings of both reactive and negative evidence, but as several authors (e.g., Leeman, 2000: 136; Long, 2007: 77) point out, it is generally avoided in current SLA research because it seems to suggest that feedback may lead to the elimination of learners’ errors, which should remain an empirical question. Finally, *corrective feedback* stands as a perfect synonym of *negative feedback*, however, as Schachter (1991) noted, the first label is more frequently used in language teaching, while the latter is the preferred choice of cognitive psychologists. In this dissertation, therefore, in line with most SLA research, the term “corrective feedback” will be used.

Like for every construct in SLA, different definitions of corrective feedback have been provided; however, while some constructs (e.g., *task, working memory*) are conceived very differently and generate heated debates, most definitions of corrective feedback agree with, e.g., Lightbown and Spada (1999: 171), in that it is “[a]ny indication to the learners that their use of the target language is incorrect.” Now, some precisions must be made. First, these indications may include information about “the accuracy, the communicative success, or content” of the learner’s language use (e.g., Leeman, 2000: 136). Most SLA studies, however, focus on feedback that evaluates *accuracy*, either on vocabulary or grammar. This is the case of the present dissertation, which focuses on feedback on grammar accuracy exclusively. Second, the indications to learners may be provided by any source, be it the tutor (human or computerized), a more competent interlocutor (e.g., a native speaker or more advanced learner), or a peer learner. Third, feedback may be verbal and vary in modality (written, oral, or a
combination) or non-verbal (e.g., based on visual cues such as pictures, videos, or gestures). Finally, feedback on grammar accuracy is not necessarily restricted to learner’s productions, and hence it is possible to give feedback to a student who has made an incorrect choice in an input-focused task; for this reason, when discussing the learners’ processes that trigger feedback on the part of the tutor it seems more appropriate to talk about “instances of language use” rather than the often-found label “learners’ productions.”

While the definition of corrective feedback in SLA is greatly uncontroversial, there is no unanimously agreed-upon classification of feedback types yet. Arguably, until very recently the pivotal criterion to distinguish feedback types was the explicitness of their negative evidence component; however, to begin with, the term explicitness has been interpreted very differently. Some authors interpret it as how overtly the feedback signals the occurrence of an error, while others refer to the provision, or lack thereof, of metalinguistic information explaining the nature of the error. Section 2.2.1 below delves further into this issue. However, explicitness of negative evidence is just one criterion. As pointed out by Leeman (2000), classifying feedback types solely based on the explicitness of negative evidence often obscures the fact that feedback moves may either encapsulate, or be accompanied by, positive evidence, a construct that will be discussed in section 2.2.2. Finally, from a different angle recent research contends that feedback types must be classified on whether or not they push or prompt learners to repair their errors. This will be discussed further in section 2.2.3.
2.2.1 The explicitness of negative evidence

Hitherto, arguably the most popular criterion used by SLA researchers to distinguish among different feedback types has been the alleged explicitness of their corrective component or negative evidence. However, the term explicitness has been understood in two different senses that need to be teased out: (1) error signaling, i.e., how overtly the occurrence of an error is signaled, and (2) error explanation, i.e., how much detail is provided to explain the reason of the error.

2.2.1.1 Error signaling

The present section discusses the explicitness of error signaling from different angles. From the tutor’s perspective, explicitness is first presented as a continuum where three main milestones can be distinguished: recasts, negotiation, and explicit feedback. Next, explicitness is re-visited from the perspective of the learner. Both sides, the tutor’s and the learner’s, are brought together in the third subsection, where Carroll’s (2001) theoretical framework is used to measure feedback explicitness holistically. Next, the controversial issue of the alleged implicitness of recasts is discussed in light of recent empirical literature. Finally, the fifth subsection reviews the different methodological tools that have been used to measure learners’ perceptions of feedback.

*Three points in a continuum: recasts, negotiation, explicit feedback*

In a first sense, the dichotomy explicit vs. implicit has been interpreted as how overtly the tutor informs the learner that his/her instances of language use are not acceptable in the L2. Thus, explicit feedback overtly states that an error has been made, whereas implicit
feedback requires the learner to infer this (Carroll, 2001: 273; Carroll & Swain, 1993: 361; Leeman, 2000: 139). For that reason, explicit feedback is typically considered a more obtrusive pedagogical intervention, since it interrupts the normal flow of communication by momentarily switching the learner’s attentional focus from meaning to form. An early and often-cited example of explicit feedback in the empirical literature is provided by Carroll and Swain’s (1993) utterance rejection group, where in response to learners’ errors tutors produced messages along the lines of “That’s wrong”, “You can’t say that”, or “That is not good English” (Carroll, 2001: 366).

As opposed to explicit feedback, recasts are typically considered the paramount example of implicit feedback, since by providing a target-like reformulation of the learner’s ill-formed utterance they indirectly signal the occurrence of an error without explicitly stating so. The key issue here, then, is that learners may, but also may not, construe the feedback message as indicating a corrective intention on the part of the tutor (e.g., Chaudron, 1988; Fanselow, 1997; Lyster, 1998a, 1998b; Lyster & Ranta, 1997). Observe, for example, the following language related episode (LRE) or form-focused episode (FFE) from Mackey, Gass and McDonough (2000: 486), where the tutor corrects a deviant verb conjugation but the student does not notice it, as she herself acknowledges in a post-task stimulated recall session:

(1) Learner (L): It have mixed colors.
    Tutor (T): It has mixed colors.
    L: Mixed colors aha.
    L’s recall: Uh, I was thinking… nothing, she just repeat what I said.

In between recasts and instances of explicit feedback, many authors (e.g., Mackey & Goo, 2007) situate so-called negotiation moves such as clarification requests, elicitations,
or confirmation checks. As illustrated by (2) below (Pica, 1988: 6), these feedback moves do not explicitly state that an error has occurred; however, they are arguably more obtrusive than recasts for at least two reasons. First, they prompt the learner to respond to them, and in so doing they can hardly be perceived as a mere repetition of the learner’s utterance or a similar back-channeling strategy. As posed by e.g., Lyster and Ranta (1997: 49), producing a prompt after an implicit feedback move corrupts the pursued implicitness. Furthermore, although negotiation moves do not explicitly signal the occurrence of an error, they do signal a breakdown in communication. The learner, however, still needs to figure out the specific nature of the breakdown (choice of form, problems with meaning, mispronunciations, audition problems, etc.). Consequently, negotiation moves may be thought of as hybrid feedback moves that are neither completely explicit nor implicit.

(2) L: and they have the chwach there.
   T: the what?  NEGOTIATION MOVE
   L: the chwach – I know someone that –
   T: what does it mean?
   L: like uhm like American people they always go there every Sunday
   T: yes?
   L: you know – every morning that there pr- that- the American people get dressed up to go to um chwach
   T: oh to church – I see.

Consequently, then, as the hybrid nature of negotiation moves illustrates, “feedback explicitness” should not be understood as a two-level qualitative variable where feedback types are either implicit or explicit, but rather as a quantitative variable where feedback types fall on different points of explicitness in a continuum. In other words, some feedback types are more implicit-like or explicit-like than others. The problem then is
how to measure explicitness. This, as discussed in the next section, can be done from two
perspectives, the tutor’s and the learner’s.

Two sides of explicitness: the tutor and the learner

When measuring the explicitness of a feedback move, two different sides must be
considered: (1) the tutor’s side, i.e., the explicitness of the tutor’s corrective intent as
determined by the linguistic characteristics of the feedback message, and (2) the learner’s
side, i.e., the interpretation of the feedback by the learner as a corrective intent or
something else. As pointed out by e.g., Sheen (2006: 386), the former can be analyzed in
terms of a linguistic phenomenon, whereas the latter is a psycholinguistic construct
contingent upon learner’s individual differences. Hitherto the present discussion has only
envisaged the tutor’s side, as have most studies on feedback. However, looking at
explicitness from this perspective exclusively may be very deceiving. For one, explicitly
stating “that’s wrong” or “wrong answer” in the course of an LRE is not even a warranty
that the learner will interpret the tutor’s message as a metalinguistic comment on his/her
deviant performance. As Carroll (2001: 373) puts it:

*That’s wrong,* for example, is multiply ambiguous. The deictic pronoun could refer to an infinite
number of objects in the world. To count as feedback and correction of the sort pertaining to grammar
learning, the learner must eliminate all of those references which are irrelevant to the form-meaning
pairs of his own talk. *That* cannot refer to the truth value of the utterance (e.g. NOT *[JOHN BUY A
CAR]*) or to the reference of one of the determiner phrases of the sentences (e.g. MARY vs. JOHN
— it was Mary who bought a car and not John), or to one of the predications made (e.g., John bought
a car and did not spend his money on a boathouse).

Similarly, the reverse, i.e., not explicitly stating that an error has occurred, does not
entail that a learner will not perceive the tutor’s corrective intent. As a matter of fact,
some feedback moves which have been traditionally labeled as recasts, and hence thought
of as instantiations of implicit feedback, may hardly be perceived as anything but corrective feedback by any average learner. Observe, for example, the following succession of recasts in an LRE from Sheen (2006: 387):

(3) L: They hired me. But when I meet someone in…
    T: Met.
    L: When I met someone, there’s a foreigner in there.
    T: Was, was.
    L: Yeah, I felt so uncomfortable.
    T: Un-.
    L: Uncomfortable to speak with them.

In this LRE the tutor does not precede each recast with an explicit metalinguistic comment such as “that’s wrong”; however, it could be argued that doing such thing would only render the feedback more obtrusive but not necessarily more explicit. As Sheen herself argues:

The illocutionary force of the teacher’s short, segmented reformulations of the learner’s erroneous utterances make the recasts seem like explicit correction. It can be argued that there is not much difference between explicit correction and the kind of recasts illustrated in this example in terms of their perceived illocutionary force and their salience.

Along these lines, an increasing number of researchers are starting to note that recasts should not be viewed as necessarily implicit but, rather, as more or less implicit or explicit depending on their segmental and suprasegmental linguistic features and the discourse context in which they appear (Ellis & Sheen, 2006: 583; Loewen & Philp, 2006: 539; Sheen, 2006: 386). I will elaborate on this specific topic later; at this point, however, the ongoing question that needs to be answered is which pivotal criterion should researchers use to classify feedback moves as explicit or implicit, i.e., the tutor’s or the learner’s side.
If, as discussed above, the adoption of a purely linguistic perspective (the tutor’s side) may lead to inaccurate generalizations, the next logical step then is to envisage the issue from a psycholinguistic perspective (the learner’s side). Alas, this is not an easy task, since learners are subject to individual differences like developmental readiness and cognitive capacity which can play a role in how they process feedback. It stands to reason then that research in this arena must contemplate individual differences. Yet, large-scale studies analyzing the overall picture, that is, how large pools of learners generally notice the corrective intent of various feedback types should also bring about useful pedagogical implications. The problem here is a blatant shortage of empirical literature analyzing learners’ perceptions of feedback, both overall and in light of individual differences (for a review on the incipient research on this issue, see the subsection entitled *Constructs and methodological tools to measure learners’ perceptions of feedback*).

In principle, it seems reasonable to assume that explicit forms of feedback will typically be processed as such by most learners, despite Carroll’s (2001: 373) insightful remarks on the illocutionary force of “that’s wrong” quoted earlier. The crucial question, then, lies in finding out how learners process the most implicit types of feedback, particularly recasts. Hitherto, as Ellis and Sheen (2006: 583-4) point out, research on this arena has produced conflicting findings. For example, Mackey et al. (2000) used stimulated recall protocols to analyze learners’ perceptions of feedback and found that recasts on morphological errors went largely unnoticed as corrective illocutionary forces, thus suggesting that any potential pedagogical effects of recasts must be the byproduct of implicit learning. This finding, however, is counter-evidenced by Long et al.’s (1998) study, where all 8 of the 12 students who improved on their use of Spanish adverb
placement after a recast session were able to correctly formulate the rule at hand. Thus, if prior knowledge was accurately controlled for and learners did ignore the rule before exposure to recasts, then it can be posited that they did have to notice the corrective intent at some point, which led them to compare their productions with the tutor’s reformulations and generate a rule. Arguably, further research must seek to clarify the cause of conflict between Long et al. (1998) and Mackey et al. (2000), although one might assume that this conflict likely arises from differences in the type of learners and/or recasts that they used. This latter issue will be discussed in the next subsection Re-evaluating the implicitness of recasts.

In conclusion, then, it can be argued that the explicitness of feedback cannot be fully measured from either the tutor’s or the learner’s sides exclusively, but rather, both sides must be brought into the equation. As Han (2001: 20) puts it, “corrective feedback is a two-way, interdependent process, involving the giver and the receiver, with both being information providers.” What is needed, then, is a theoretical model that envisages feedback from the tutor’s and the learner’s sides. That is the goal of the next subsection.

**An integrative theoretical framework to measure feedback explicitness**

At this particular point, the debate on explicit/implicit feedback can only make headway by using a theoretical model that allows for a classification of feedback in terms of both the tutor’s intent and the learner’s perceptions. A good first step in this direction is an analysis of the actions that learners are hypothesized to carry out in order to process feedback for acquisition. As posed by a number of authors (e.g., Carroll, 2001; Gass & Varonis, 1994; Schmidt, 1983, 1990), these actions transcend noticing the tutor’s
corrective intent and extend to processing the segmental and suprasegmental information conveyed in the feedback message to figure out what in particular is being corrected.

Following Carroll’s (2001: 386) model, there are three main actions to perform: (1) recognizing the corrective intent of the tutor; (2) spotting the locus of the error, i.e., where in the utterance the deviant form is, and (3) recognizing the nature of the error, i.e., what level of execution (speech act, lexical choice, morphosyntax, linear order, articulation) is implicated. The latter two steps are often referred to in the literature as the “blame assignment problem” (Pinker, 1989) and are crucial for feedback to lead to grammar acquisition. When learners assign the blame, they are in a position to notice the gap in their interlanguage (IL), i.e., notice that there is a discrepancy between the L2 and their current use thereof, either receptive or productive (Doughty, 2001; Ellis, 1991b; Gass & Varonis, 1994; Philp, 2003; Schmidt & Frota, 1986). This constitutes, in terms of Schmidt’s (2001: 31) noticing hypothesis, the “first step in language building,” since it makes students more alert to positive evidence. Thus, in line with “classical” approaches to cognition (e.g., Fodor & Pylyshyn, 1988), learners can start to analyze the input by e.g., a process of IL-L2 comparison (Ellis, 1991a; Gass, 1991, 1997; Gass & Varonis, 1994; Long, 1996), forming their own hypotheses and subsequently testing them, which may ultimately result in interlanguage restructuring.

In a recent study, Ellis, Loewen and Erlam (2006) used Carroll’s model to explain the differential benefits observed for two types of feedback, i.e., partial recasts vs. explicit feedback in the form of error repetition plus provision of metalinguistic information. Along these lines, I will proceed to do the same for the three quintessential feedback types discussed so far: recasts, negotiation moves, and explicit feedback. As mentioned
earlier, all these feedback moves may take shape in very different instantiations with varying degrees of linguistic explicitness. At present, however, I will only consider them in their allegedly “baseline form”, in an attempt to inoculate potentially intervening variables (e.g., provision of positive evidence and prompting for repair, unless strictly required by definition, as in recasts and elicitations, respectively). Specifically, for the current purposes recasts are operationalized as full recasts, i.e., target-like reformulations of the whole deviant utterance without suprasegmental emphasis on the target form and no prompts for repair; in turn, as a subtype of negotiation moves I choose elicitations, here defined as a partial repetition of the student’s utterance followed by a pause before the problematic constituent, plus a prompt; finally, explicit feedback is here defined as a metalinguistic comment on the inaccuracy of the learner’s utterance, without provision of either the locus or the nature of the error. Example (4) below provides illustrative samples of all three feedback types:

(4) L: My mother have green eyes.
Recast: Uh-huh, your mother has green eyes. How about your father?
Elicitation: Your mother… what?
Explicit feedback: Sorry, that is not good English.

Following Carroll (2001), then, it can be hypothesized that these feedback moves will differently aid learners as described in Table 3.2 below.
Table 3.2 Facilitative potential of three quintessential feedback moves (recasts, negotiation, explicit feedback)

<table>
<thead>
<tr>
<th>Learner’s Actions</th>
<th>Tutor’s Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recast</strong></td>
<td><strong>Negotiation (elicitation)</strong></td>
</tr>
<tr>
<td>(1) Recognizing the corrective intent</td>
<td>The corrective force is not evident, since the learner may perceive the recast as a repetition of his/her utterance.</td>
</tr>
<tr>
<td>(2) Spotting the error locus</td>
<td>The provision of positive evidence immediately after the deviant utterance may help the learner to spot the error locus, subject to e.g., developmental level or working memory capacity.</td>
</tr>
<tr>
<td>(3) Recognizing the nature of the error (lexis, morphology, syntax, pragmatics, pronunciation, etc.)</td>
<td>The provision of positive evidence in contrast to the deviant utterance may help the learner to recognize the level of execution that is implicated.</td>
</tr>
</tbody>
</table>

Table 3.2 above illustrates a number of interesting points. First, according to the definition of explicitness used here, more explicit types of feedback do not necessarily provide “more help” to learners. Oftentimes, the implicit vs. explicit feedback (or teaching) issue has been conceived as creating the conditions for students “to fish” rather than “giving them the fish.” However, as illustrated here, due to the positive evidence that they embed in juxtaposition to negative evidence, recasts potentially provide more help in solving the blame assignment problem than negotiation and explicit feedback in their baseline forms. Secondly, though, it should be noted that the intuitive classification “recasts > negotiation > explicit feedback” as increasingly explicit types of feedback can
only be understood by bearing in mind that the learner’s actions described by Carroll
stand in causal relation; that is, one can notice the corrective intent but fail to assign
blame (locus and level), whereas one can hardly assign blame without noticing the
corrective intent. Therefore, noticing the corrective intent is the first requisite in
determining the explicitness of a feedback move, which is why recasts are here classified
as more implicit than the other two prototypical feedback types. Thirdly, the
classification presented here is arguably an abstraction that under-represents the variety
of feedback moves that naturally occur in either classroom settings or naturalistic
environments. For example, within the explicit feedback category it is not uncommon to
encounter messages that provide substantially more help than the baseline type illustrated
here (“Sorry, that is not good English”) in solving the blame assignment problem, e.g.:
“Sorry, that is not good English. We don’t say ‘My mother have’. We say ‘My mother
has’, because ‘My mother’ is third person singular”. However, feedback types like this do
more than signaling a corrective intent and helping with blame assignment (locus and
level): they also provide positive evidence and explain the nature of the error via a
metalinguistic rule. In other words, feedback messages like this are multi-componential
moves that entangle a number of variables, each of which may bear a different impact on
L2 development and must be individually researched. What is needed, then, is a
taxonomy that thoroughly discriminates between feedback types in terms of the presence
or absence of each of these components.

A number of recent reviews and meta-analyses contend that mainly all feedback types
can be subsumed into the three categories described above, i.e., recasts, negotiation, and
explicit feedback (e.g., Mackey & Goo, 2007). There are, however, more complex tones
in this picture. First, as mentioned earlier each category boasts a variety of feedback subtypes with different characteristics and possibly different effects on L2 development (e.g., “negotiation” subsumes clarification requests, repetitions, elicitations, and confirmation checks), which warrants empirical investigation of each separate subtype and calls for more refined feedback taxonomies. And second, even the same feedback type is understood differently by different authors, which may lead to confusing overgeneralizations. I will work toward a more comprehensive taxonomy all the way down to section 2.1.3. Let me now focus on the second front, which has an important bearing for the implicit vs. explicit debate.

**Re-evaluating the implicitness of recasts**

An overview of the empirical literature shows that authors often use the same label to refer to very different operational definitions of a feedback type. This is particularly patent in the case of recasts, which given the number of interpretations are increasingly recognized as an “elastic” rather than a “monolithic” construct (e.g., Ellis et al., 2006; Loewen & Philp, 2006; Mackey & Goo, 2007; Sheen, 2006). Arguably, the existing hurly-burly of operational definitions is a logical byproduct of the popular status of recasts as pedagogical instruments in SLA over the last twenty years. By and large, research has privileged recasts over other feedback types for two main reasons: first, descriptive studies show that they are the most frequently used feedback type in and out the classroom environment (Braidi, 2002; Chaudron, 1977; Doughty, 1994; Ellis, Basturkmen, & Loewen, 2001; Fanselow, 1997; Iwashita, 2003; Loewen, 2004; Lyster & Ranta, 1997; Morris, 2002; Oliver, 2000; Oliver & Mackey, 2003; Panova & Lyster,
2002; Roberts, 1995; Seedhouse, 1997) although their relative frequency viz à viz other feedback types varies depending on the setting; and second, recasts are commonly conceived of as the most implicit and less obtrusive of all feedback types. Specifically, they are said to be time-saving, to help maintain a focus on meaning to a greater extent than other techniques —thereby adding authenticity to the classroom— and to be less threatening to student confidence, which presents them as optimal instructional interventions (e.g., Han & Kim, 2008; Loewen & Philp, 2006; Nicholas et al., 2001). These arguments, however, are often taken as an irrefutable premise, and authors rarely question whether they remain true in their particular operationalizations.

In an attempt to address this issue, a number of recent studies (e.g., Loewen & Philp, 2006; Nassaji, 2007; Sheen, 2006), building upon previous research (Philp, 2003; Sheen, 2004), have challenged the traditional view of recasts as purely implicit feedback moves and contend that recasts do vary in their degree of explicitness and salience. Based on classroom observations, Loewen and Philp (2006) and Sheen (2006) present a collection of linguistic factors that supposedly contribute to modulate the alleged implicitness of recasts. To my knowledge, these factors, which must be researched empirically, can be distilled, merged, and classified into three main levels, namely, the segmental level, i.e., the content of the message where the recast is embedded; the suprasegmental level, i.e., the prosodic features of the recast; and the discourse context, i.e., the conversational moves that precede and follow the recast. Figure 3.11 below presents a taxonomy of recasts mainly based on Loewen and Philp (2006) and Sheen (2006), with additions from other sources (J. H. Kim, 2004; Nassaji, 2007), where features are ordered according to
the proposed three levels. To minimize terminological confusion, synonymous labels used in other sources have been included.
Fig 3.11 A taxonomy of recasts based on segmental, supra-segmental, and discourse-level features that may modulate their implicitness (based on Loewen & Philp, 2006; Sheen, 2006, with additions from Kim, 2004 and Nassaji, 2007)

- **Segmental level**
  - Number of changes
    - One-change or simple
    - Multiple-changes or complex
  - Contextual information
    - Full, whole, or isolated
    - Partial or segmented
    - Expanded, extended, or incorporated
  - Length of recast
    - Word-level
    - Phrase-level
    - Clause-level
  - Type of change
    - Addition
    - Deletion
    - Substitution
    - Reordering
    - Combination

- **Supra-segmental level**
  - Prosodic emphasis (aural mode) / Textual enhancement (written mode)
    - Stressed or enhanced
    - Unstressed or non-enhanced
  - Intonation
    - Declarative
    - Interrogative

- **Discourse level**
  - Consistency of delivery
    - Intensive
    - Non-intensive
  - Consistency of focus
    - Form-specific
    - General purpose
  - Feedback moves surrounding the recast
    - Single-move
    - Multi-move
      - Corrective
      - Repeated
      - Combination
  - Topic continuation after recast
    - Immediate topic continuation
    - Non immediate topic continuation
At the segmental level, one of the features that may moderate the saliency of a recast is the number of changes that exist between the student’s utterance and the tutor’s reformulation. Thus, depending on whether they fix one or more errors, recasts can be divided into simple and complex, with the former likely being more noticeable as a correction than the latter (Farrar, 1992; Lyster & Ranta, 1997; Philp, 2003). Also, recasts can be divided into partial, full, or expanded, depending on the amount of contextual information surrounding the reformulated error. Partial recasts just reformulate the erroneous form in isolation, while full recasts keep the surrounding targetlike context intact, and expanded recasts include additional information for topic continuation purposes. The rational here is that the narrower the focus of the recast, the more noticeable it becomes (Doughty & Varela, 1998; Mackey & Philp, 1998; Philp, 2003). By the same token, shorter recasts are argued to be more easily noticed than longer ones. Crucially though, length of recast should not be confused with segmentation. Segmented or partial recasts pinpoint the problem, whereas short recasts are contingent upon the length of the learner’s utterance and do not necessarily pinpoint the problem (Loewen & Philp, 2006: 541). To date, different units of length have been proposed, ranging from the number of syntactic constituents (word, phrase, clause - Sheen, 2006) to the number of morphemes (fewer or more than five - Loewen & Philp, 2006; Philp, 2003). Finally, a last segmental factor that may modulate the explicitness of the recast is the type of change carried out upon the triggering utterance, i.e., addition, deletion, substitution, reordering, or a combination. Research on this factor is particularly scant, but authors
such as Sheen (2006: 380) contend that recasts using substitution might be considered the most salient to learners.

At the suprasegmental level, while some authors have delivered recasts in their baseline form, others have used prosodic cues such as pitch, pausing, and emphasis (in oral interaction) or textual enhancement techniques (in written interaction environments such as CMC) to highlight the deviant constituents, thereby allegedly increasing the explicitness of the correction (e.g., Sachs & Suh, 2007). Additionally, intonation, i.e., declarative vs. interrogative, has also been argued to modulate the saliency of recasts. The thesis here is that interrogative recasts are less likely to be perceived as corrective forces because they may be read as a request to the learner to confirm the intended meaning rather than to focus on form; notably, however, declarative recasts can also be perceived as a mere tutor’s repetition of the learner’s utterance for back-channeling purposes (Loewen & Philp, 2006: 541).

Finally, at the discourse level the alleged implicitness of recasts may also be modulated by the consistency and amount of recasts, both holistically and in relation to other feedback moves. Thus, so-called intensive recasts (produced consistently after every error) and form-specific recasts (produced selectively for a particular form) are deemed to be more salient than non intensive, general purpose recasts (Nassaji, 2007). Also, the corrective force of recasts can be enhanced by repeating them or using them in combination with other more explicit feedback moves such as prompts (Lyster, 1998b), repetition of the deviant form with rising intonation (Doughty & Varela, 1998), clarification requests (Muranoi, 2000), or elicited recall via non-verbal stimuli (Philp, 2003). Finally, the existence of a topic continuation move on the part of the tutor, or even
the time lapse between the recast and such a continuation move may determine whether or not learners perceive the corrective intent of the recast or can successfully assign the blame (cf. also James, 1998; J. H. Kim, 2004; Lyster, 1998b).

All in all, this re-evaluation of the notion of recast is extremely important because it discredits earlier claims on the potential benefits of implicit vs. explicit feedback based on comparisons between recasts and other allegedly more explicit feedback types. As it has been shown, labeling recasts as implicit feedback types solely based on the fact that they do not verbally state “your answer is wrong” is an inexhaustive approach that may lead to inaccurate conclusions. Therefore, large-scale empirical research is needed to elucidate which linguistic features are usually perceived by learners as red flags signaling a corrective intent, both in light of learners’ individual differences and regardless. On this token, the incipient research has resorted to various methodological tools that are briefly discussed in the next section.

**Constructs and methodological tools to measure learners’ perceptions of feedback**

Recently, a number of researchers have availed themselves of various constructs and methodological tools to investigate learners’ perceptions of feedback, including measures of uptake, measures of noticing via stimulated and immediate recalls, and tests of metalinguistic awareness, all with their own strengths and weaknesses.

To date, uptake is probably the most popular construct in research addressing learners’ perceptions of feedback, as shown by a growing body of research (Ellis et al., 2001; Loewen & Philp, 2006; Lyster, 1998b; Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2006). Usually defined as the learner’s immediate response after a feedback
move (Lyster & Ranta, 1997), uptake may well constitute an indirect indicator of feedback noticing, but it is not fully reliable for a number of reasons. First, some feedback moves like recasts do not typically create an opportunity window for uptake to occur because they encourage topic continuation, so the absence of observable uptake does not mean that the corrective intent went unnoticed; in other words, uptake is an optional discourse move (Ellis et al., 2001; Loewen, 2004; Mackey & Philp, 1998; Oliver, 1995; Sheen, 2004). Additionally, although successful uptake may indicate that the learner has noticed a recast (Ellis et al., 2001; Lightbown, 1998; Loewen, 2004), it may also be the case that the learner is repeating the tutor’s reformulation in a parrot-like fashion, without really noticing the corrective intent or the mismatch between his/her interlanguage and the L2 (e.g., Ammar & Spada, 2006: 565).

In addition to measuring uptake, Mackey et al. (2000) used a procedure known as stimulated recall (Gass & Mackey, 2000) to directly ask learners about their perceptions of feedback. Immediately after task-based interaction, learners were played a video-recording of their interaction sessions and were asked by the researchers to articulate what they were thinking at the time they took part in the feedback episodes. While this procedure may solve the optionality problem associated with measures of uptake, it also presents a number of problems. For example, stimulated recalls may not be reliable because of memory decay, i.e., learners may not accurately remember their thoughts at the time the feedback episode took place. Also, as argued by Gass and Mackey (2007: 62) learners “may not be aware of their thought processes and/or may not wish or be able to reveal them” or, as posed by J. H. Kim (2004), they may even make up their recalls at the moment they are asked.
As an alternative to stimulated recalls, recent studies by Philp (2003) and Egi (2004) used another type of verbal protocols known as immediate recalls. These differ from the former in that they are not carried out after task completion, but immediately after each feedback episode throughout the course of interaction. For example, in Philp’s study the researcher informed the learners that she would occasionally knock on the table twice as a recall prompt, and that they would then have to repeat the last thing that they had heard prior to that sound. Similarly, Egi used the same auditory stimulus but rather than asking her participants to repeat the recasts, she asked them to verbalize their thoughts on language episodes during a 10-15 second-long conversational turn. Arguably, immediate recalls greatly solve the memory decay problem posed by stimulated recalls. Yet, as mentioned earlier learners are not always aware of their thought processes, may have trouble verbalizing them, may make them up, or may decide not to reveal them. Also, the immediate nature of these verbal protocols may interfere with task performance (i.e., reactivity) and in this particular case trigger increased noticing of the recasts (e.g., J. H. Kim, 2004).

Typically, recommendations for both stimulated and immediate recall research include avoiding priming students to detect the goal of the study. Authors are thus suggested to ask indirect questions along the lines of “What were you thinking back then?” or give instructions such as “Can you repeat what you just heard?” Some authors, however, have used more direct techniques to determine perceptions of feedback, by directly asking students to watch interaction sessions and spot instances of tutor’s corrections. That is the case of Roberts’s (1995) study, in which three volunteers were asked to watch a tape of a teacher-centered class and jot down the meter reading every
time they thought that the teacher had corrected someone. This technique, however, may clearly predispose students to detect implicit feedback moves that they might have not detected in a regular classroom setting, where teachers seldom alert students to analyze the force of their illocutionary acts.

Finally, one last way to measure learners’ perceptions of feedback is to administer tests of metalinguistic awareness after a treatment session and any applicable posttests. For example, after a practice session targeting Spanish adverb placement Long et al. (1998) asked students who had been exposed to recasts whether they had learned anything new. Interestingly, all 8 of the 12 participants that showed a learning gain were able to produce the grammar rule correctly, unlike participants in another experimental group who received models instead of recasts. This finding can have various readings; for example: (1) implicit feedback via recasts can lead to explicit learning (i.e., noticing the corrective intent, blame assignment, and rule formation); (2) the recast-based treatment in this study was actually not that implicit: the task-essential nature of the practice discouraged open-ended utterances and elicited production of the target structure in isolation, and errors were followed by intensive, form-focused recasts that may have been noticed as corrections; and (3) implicit feedback may indeed have promoted implicit learning, with students generating the rule only after being instructed to do so. Clearly, further research must be conducted to elucidate this, and extreme caution must be exercised when making statements such as “implicit instruction can generate explicit learning” on the basis of experiments like Long et al.’s (1998).

To sum up, then, all the procedures discussed above have limitations. Yet, they are greatly contributing to show that feedback cannot solely be categorized as explicit or
implicit depending on whether or not the tutor overtly indicates the occurrence of an error via statements like “Your answer is wrong.” Rather, the explicitness of a feedback message is a twofold coin that must be envisaged from the tutor’s and the learner’s side. From the tutor’s point of view, the explicitness of a feedback move is a linguistic construct determined by the segmental, suprasegmental, and discourse features of the message. From the learner’s perspective, however, feedback explicitness is a psycholinguistic construct determined by whether or not students perceive the corrective intent of a feedback move and spot the gap in their interlanguage. In this context, the constructs and methodological tools discussed above are, despite their limitations, the only way to contemplate how the two sides of the coin are interconnected.

2.2.1.2 Error explanation

So far, the ongoing discussion on implicit/explicit feedback has exclusively focused on how overtly tutors signal the occurrence of an error by using statements such as “Your answer is wrong” or various segmental, suprasegmental, and discourse-level features which modulate how learners detect a corrective intent and a gap in their interlanguage. However, many authors have understood the implicit/explicit feedback dichotomy in a different way, based on whether or not a feedback message provides a rationale for the error at play, i.e.: “Your answer is wrong because…” Interestingly, these latter authors have used different labels that must be clarified to avoid terminological confusions. For example, in the field of CALI Rosa and Leow (2004a; 2004b) used the labels “implicit” vs. “explicit” to signify the absence or presence of metalinguistic explanation. However, from the “error signaling” perspective discussed above, both feedback types qualify as
“explicit” because they overtly state the occurrence of an error. For this reason, when comparing similar types of computerized feedback Sanz and Morgan-Short (2004) avoided the use of the label “implicit,” using instead the labels “less” vs. “more” explicit feedback. Yet, the mathematical formulation that they used, i.e., [+/- explicit feedback], can be misread as “presence” vs. “absence” of explicit feedback.

At this point, there is a clear need for conceptual and terminological normalization. Specifically, since “explicitness” has been interpreted both as “error signaling” (“Your answer is wrong”) and “error explanation” (“Your answer is wrong because…”), it is necessary to determine whether this is an issue of essence (i.e., error signaling and explanation are two independent constructs) or degree (i.e., both error signaling and explanation are instantiations of explicit feedback, with the latter rating higher than the former on the gradual scale of explicitness). One of the earliest attempts to distinguish feedback explicitness in terms of error signaling and explanation, both terminologically and operationally, was provided by Carroll and Swain’s (1993) empirical investigation on the developmental effects of various feedback types. In response to errors during tutor-student paired interaction, participants in the so-called “explicit utterance rejection” condition were simply informed that their utterances were off-target, whereas those in the “explicit hypothesis rejection” group were additionally presented with a detailed explanation of the governing rule at hand. According to Leeman (2000: 141) the fact that both feedback types were labeled as “explicit” seems to suggest that the difference between the two is one of degree rather than essence (bolding added):

[U]sing the single feature of explicitness to classify negative feedback seems to imply that these two treatments represent points on a single explicitness-implicitness continuum. While there is no question that these treatments are different, and thus may have different effects on L2
development, this difference is not one of degree; if explicitness is operationalized as referring only to whether or not the NNS is overtly informed of the (un)grammaticality of her utterance, the two feedback types are identical in this regard, although ‘explicit hypothesis rejection’ contains an additional metalinguistic component. While Carroll and Swain’s study did not include such a treatment, it is feasible that grammar explanation could be offered in response to a non-target utterance, without ever explicitly commenting on the unacceptability of the NNS utterance. Their definition of explicit feedback above would require this response to be classified as implicit, which is counterintuitive, given that rule presentation is commonly considered explicit teaching. For this reason, it is useful to classify the explicitness of feedback independently from whether or not metalinguistic information is provided.

In Leeman’s view, then, the distinction between Carroll and Swain’s “explicit utterance rejection” (error signaling) and “explicit hypothesis rejection (error explanation) is one of essence, because both types are explicit in nature and the difference between the two resides in the presence of a variable component, metalinguistic explanation. While I concur with Leeman in that error signaling and explanation are two separate components of feedback that must be teased out in order to build exhaustive feedback taxonomies, it is my stand that in line with a vast body of previous research, feedback that includes metalinguistic explanation constitutes the very extreme of the explicitness continuum. Hence, the difference between Carroll and Swain’s groups is simultaneously one of essence and degree, as I elaborate on further below.

Arguably, as discussed earlier in section 2.1 (The place of feedback in SLA) there are two robust arguments in support of Leeman’s thesis that metalinguistic explanation constitutes an independent construct in the classification of input for SLA: first, although grammar explanation is typically conceived of as a preemptive pedagogical intervention, it clearly may be presented reactively (Leeman, 2000: 141); and second, albeit metatalk is usually accompanied by examples of what does not work in a language, or negative
evidence, it can also be presented along models of what does work, i.e., positive
evidence, or even produced in isolation, that is, zero evidence.

Now, precisely because metalinguistic explanation is an independent type of input in
SLA, not every instance of metalinguistic explanation that is presented reactively in
response to learners’ errors must necessarily constitute feedback. By definition, both in
cognitive psychology and SLA feedback is “a mechanism by which the processor (or
learner) is provided with information regarding the success or failure of a given process
[or instance of language use]” (Leeman, 2000: 136). That is, feedback intrinsically
contains a tutor’s evaluation on the quality of the language used by the student. As
discussed earlier, this evaluation may be explicit (e.g., “Your answer is wrong”) or
implicit, provided via complex feedback moves such as recasts, but if such evaluation is
absent, then no feedback is provided. Therefore, the contradiction of an “implicit rule
presentation” group posed by Leeman is actually resolved. In other words, what Leeman
is hypothesizing is a case of reactive rule presentation but not feedback, and that is an
independent construct that certainly warrants empirical investigation.

Undoubtedly, then, the difference between Carroll and Swain’s (1993) “explicit
utterance rejection” and “explicit hypothesis rejection” is one of essence; however, as I
will now argue, it is also one of degree. Based on Carroll’s (2001) model, the explicitness
of a feedback move can be measured by how it helps learners in carrying out the actions
that they purportedly need to process feedback for acquisition, namely: (1) uncovering
the corrective intent; (2) spotting the locus of the error; and (3) identifying the nature of
the error, i.e., the level of execution (speech act, lexical choice, morphosyntax, linear
order, articulation). In this sense, explicitly stating that an error has been made only
facilitates the first process; conversely, signaling the error and explaining the rationale behind it facilitates all three processes. Therefore, feedback with metalinguistic explanation clearly represents the extreme in the explicitness continuum.

Now, as mentioned by a number of researchers (e.g., Gass et al., 2005) the explicitness of a feedback move may be moderated by the features of the task at hand. This is very pertinent in the case of Carroll and Swain’s (1993) study, where all subjects were informed of the experimental goal (practice with English dative alternation) at the outset of the study (p. 364):

We are doing a study concerned with English as a second language. I will give you a sentence and I would like you to think of a different way of saying the same thing. For example, in English you can say Peter wrote a letter to Theresa. Once again, Peter wrote a letter to Theresa. But you can also say Peter wrote Theresa a letter. I repeat: Peter wrote Theresa a letter. These two sentences, Peter wrote a letter to Theresa and Peter wrote Theresa a letter, have the same meaning; they “alternate.”

After receiving the instructions, learners were presented with a battery of prepositional sentences like Peter wrote a letter to Theresa and asked to produce an alternative form. Since dative alternation in English is constrained by the phonological and semantic properties of the verb at play, there were many opportunities for learners to make mistakes due to overgeneralization (e.g., notice My uncle donates a lot of money to charity but *My uncle donates charity a lot of money). When learners erred, they received tailored feedback depending on their condition. Arguably, under these conditions “explicit utterance rejection” and “explicit hypothesis rejection” are analogously explicit, since the instructions and the features of the task specify the locus and nature of any occurring error. In other words, both feedback types help learners to assign the blame.

Yet, intuitively “explicit hypothesis rejection” is a more explicit feedback type. From a
theoretical perspective this is because it further assists learners in the analytical processes that follow blame assignment towards L2 development, specifically, hypothesis formation. Therefore, the previously proposed framework to measure feedback explicitness must be extended to include a fourth facilitative dimension of feedback: hypothesis formation through provision of the rule at hand. To sum up, then, the explicitness of feedback moves is increasingly determined by how they (per se and in combination with extrinsic factors) assist learners in carrying out the following subprocesses towards L2 development: (1) recognizing the corrective intention > (2) spotting the error locus > (3) recognizing the nature of the error > (4) forming hypotheses.

It should be noted, however, that stating a metalinguistic rule does not entail its immediate understanding on the part of the learner. As argued by Carroll and Swain (1990: 362), metatalk may pose serious problems of interpretation, depending on factors such as the intrinsic nature of the rule, the metalinguistic awareness of the learner, and the type of language register used. For example, the rule in Carroll and Swain’s (1993) study is a rather complex one because it operates at two levels, semantic and phonological, and since dative alternation can usually only occur when both sub-rules apply, the learner cannot make do by understanding only one of them. To illustrate this, let us reproduce each sub-rule below (p. 374):

[SEMANTIC EXPLANATION:] For example, if a subject gave an incorrect response to “He fixed the lawnmower for his neighbor,” he or she was told “You cannot say He fixed his neighbor the lawnmower because with alternating verbs there is always an exchange of possession. Here the neighbor has the lawnmower to start with, and also has it after it is fixed. The person who fixed the lawnmower never gives the lawnmower to its owner. It always belonged to the owner.”
[PHONOLOGICAL EXPLANATION:] For example, if a subject gave an incorrect response to “The students pronounce a new word for their teacher,” he or she was told “You cannot say *The students pronounce their teacher a new word*. *Pronounce* does not alternate; this is because it is a long verb that is louder at the end and such verbs don’t alternate. *Pronounce* is louder at the end than at the beginning.”

Interestingly, in Carroll and Swain’s study the two rules were never provided together but in an interspersed way, quite possibly to avoid processing overload on the part of the students. It would be interesting, however, to learn whether or not the students comprehended the two rules with equal rates of success. If that was not the case, it would also be interesting to explore whether the successful interpretation of a metalinguistic rule may be contingent upon not only the clarity of the rule but also the student’s individual learning style or metalinguistic awareness of each level (i.e., some students may find it easier to analyze one type of information, semantic or phonological, depending on their level-specific linguistic sensitivity or prior education). One recent step in this direction has been taken by Cerezo and Martínez (2007), who isolated two experimental conditions to observe the effects of two types of metalinguistic feedback (grammatical vs. semantic-based) during practice based instruction with Spanish *tough*-constructions with *ser* and *estar*. Their study, although quasi-experimental in nature due to a low number of participants, does show greater gains for the semantic-based rule over the grammar-based one. Also, preliminary analysis of concurrent verbal protocols suggested individual differences in linguistic sensitivity. Consequently, as argued à propos of implicit feedback earlier, even very explicit feedback is a two-sided coin that involves both the tutor and the learner.
2.2.2 The provision of positive evidence

As argued by Leeman’s (2003) seminal work, corrective feedback is not only a source of negative evidence but may also be a source of positive evidence. From a theoretical perspective, the juxtaposition of these two components in a single speech move, such as in recasts, is argued by many (e.g., Long, 2007) to be more facilitative than exposure to positive evidence alone because it allows the learner to notice a gap in his/her interlanguage and fill it in immediately, thus contributing to the strengthening of targetlike form-meaning mappings.

Hitherto, however, very little effort has been devoted to analyze the individual and combined effects of positive and negative evidence, posing some interesting questions. For example, it has been posited that feedback moves such as recasts can be used to teach new forms because their embedded positive evidence helps learners to resolve the knowledge gap, whereas more explicit types that do not include positive evidence (e.g., elicitation) are thought to require at least latent knowledge of the targeted linguistic form (Long, 2007; Loewen & Philp, 2006; Loewen, 2002). However, this remains an empirical question. Also, it has been argued that recasts are ideal pedagogical interventions because they are minimally disruptive, allowing learners to keep a focus on meaning. However, it would be interesting to investigate whether the benefits of positive evidence are actually contingent upon the degree of explicitness of the negative evidence that accompanies it. Also, positive evidence can be accompanied by tutor’s prompts for learners to repair their errors, which may moderate the way in which feedback is perceived (Nassaji, 2007; Ranta & Lyster, 2007). That is the focus of the next section.
2.2.3 Prompting for error repair

In her reformulation of the Output Hypothesis, Swain (1993) argued that output-focused practice can materialize in two different ways. Bare-bones output refers to practice where the learner is pushed to write or speak, without any further interventions on the part of the tutor or a more competent interlocutor. However, quite often the action of speaking or writing alone may not be sufficient to facilitate L2 development because learners may find their way to convey meaning successfully at the expense of grammatical or pragmatically accuracy. For that reason, in contrast to the notion of bare-bones output Swain introduced that of pushed or modified output, i.e., output that the learner reformulates in a target-like way in response to the tutor’s feedback. As Swain put it (1993: 160): learners “need to be pushed to make use of their resources; they need to have their linguistic abilities stretched to their fullest; they need to reflect on their output and consider ways of modifying it to enhance comprehensibility, appropriateness, and accuracy.”

Interestingly, the importance of pushing learners to repair their own language errors was already highlighted as early as 1977 by Chaudron’s descriptive work based on classroom observations, but until the last decade it has been largely overlooked both in the theoretical and empirical literature. From a psycholinguistic perspective, pushed output can contribute to L2 development because it promotes additional grammatical encoding, either through the generation of a new message or the reprocessing of an old one (Izumi, 2003), thus leading to the formulation of new hypotheses or the re-evaluation of old ones. Additionally, as posed by Mackey (1999), when learners produce a particular structure they are likely to use it repeatedly in subsequent utterances. This “clustering”
phenomenon is important because if the reformulation is successful, it can strengthen and ultimately automatize knowledge representations through repeated use.

In sum, the last three sections have demonstrated that feedback moves must not be classified exclusively on the basis of the explicitness of the negative evidence that they provide. Rather, feedback moves can be multi-componential sources of information, also providing positive evidence and prompting for error repair on the part of the learner. Each of these components, negative evidence, positive evidence, and prompting for repair, may differently affect L2 development on their own right, so it is important to tease them out and study their independent and combined contributions. Additionally, in order to extrapolate findings from previous and ongoing empirical research, it is necessary to create a comprehensive and usable taxonomy that envisages the role of these three components. This is precisely the goal of the next two sections.

2.3 A comprehensive taxonomy of corrective feedback

A necessary pre-requisite for building a classification of feedback types is compiling a catalog of the different feedback samples that do occur in the classroom or in naturalistic environments. Probably the earliest significant effort in this direction was conducted by Lyster and Ranta (1997), who based on classroom observations distinguished between six different types of feedback: explicit correction, recasts, clarification requests, metalinguistic feedback, elicitation, and repetition (more on these in section 2.4 below).

Building upon descriptive studies, the first attempt to classify feedback types based on their components was carried out by Long and Robinson (1998). As part of a more ambitious plan to classify input in SLA, these authors implemented useful dichotomies
such as type of evidence (positive vs. negative), timing (preemptive vs. reactive), and explicitness of negative evidence (explicit vs. implicit) (see section 2.1 for details).

As posed by Leeman (2000), however, the linearity of Long and Robinson’s (1998) taxonomy under-represented the multi-dimensional nature of many feedback moves. For example, feedback does not necessarily have to provide negative evidence, as it can be positive and reinforcing. Conversely, some negative feedback moves such as recasts can also embed positive evidence. To address these and other caveats, Leeman proposed a new taxonomy where input types are defined in terms of three components that are considered independently rather than in hierarchical relation: positive evidence, negative evidence, and metalinguistic rule presentation. Additionally, each of these components is defined according to five values that determine the availability of the input (present vs. absent), its explicitness (implicit to explicit), the timing (reactive vs. proactive), the saliency of the target form (high to low), and the clarity of the meaning conveyed (high to low). The resulting taxonomy is a non-linear matrix like the one in Table 3.3 below.
Table 3.3 Leeman’s (2000: 151) input classification scheme

<table>
<thead>
<tr>
<th></th>
<th>Availability</th>
<th>Explicitness</th>
<th>Timing</th>
<th>Salience of form</th>
<th>Clarity of meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive evidence</strong></td>
<td>present vs. absent</td>
<td>implicit to explicit</td>
<td>reactive vs. proactive</td>
<td>high to Low</td>
<td>high to Low</td>
</tr>
<tr>
<td><strong>Negative evidence</strong></td>
<td>present vs. absent</td>
<td>implicit to Explicit</td>
<td>reactive vs. proactive</td>
<td>high to Low</td>
<td>high to Low</td>
</tr>
<tr>
<td><strong>Metalinguistic rule presentation</strong></td>
<td>present vs. absent</td>
<td>Explicit</td>
<td>reactive vs. proactive</td>
<td>high to Low</td>
<td>high to Low</td>
</tr>
</tbody>
</table>

Arguably, Leeman’s classification of input constitutes a significant improvement in that it is extremely productive and it identifies key components of input that should be isolated as independent variables. However, with regards to feedback specifically it presents a number of shortcomings. First, the multidimensional nature of Leeman’s matrix diminishes the visual power of linear classifications, where different operationalizations of feedback can be appended off every node, thus contributing to a bird’s eye view of what has been covered in the literature and facilitating comparisons across studies. Second, Leeman’s definition of explicitness as overt vs. non-overt error correction is arguably inexhaustive. As discussed by Carroll, the explicitness of a feedback move can be measured by how it helps learners in carrying out the actions that they purportedly need to process feedback for acquisition, namely: (1) uncovering the corrective intent; (2) spotting the locus of the error; and (3) identifying the nature of the error, i.e., the level of execution (speech act, lexical choice, morphosyntax, linear order, articulation). Third, while Leeman’s isolation of metalinguistic rule presentation makes sense for a general classification of input in SLA (metalinguistic rules can be presented
either preemptively or reactively, with or without positive and/or negative evidence) it is 
arguably unnecessary for a classification of feedback. According to Leeman (2000: 241),
from a hypothetical perspective a feedback move could be simultaneously implicit (i.e.,
the learner is not explicitly informed about the occurrence of an error) and provide
metalinguistic information (i.e., the feedback message provides an explanation of how the
L2 form at play works). However, by definition feedback must provide an evaluation of
the learner’s utterance, either implicit or explicit. Therefore, in this dissertation feedback
with metalinguistic rule presentation is considered to represent, as most authors contend,
the most explicit extreme along a continuum of feedback types, because it not only assists
learners with processes 1-3 above but goes one step further, helping learners to (4) form
hypotheses. Fourth and finally, Leeman’s taxonomy does not capture the fact that
feedback moves are often accompanied by specific tutor’s prompts for learners to repair
their errors, which may modulate the way in which feedback is perceived (Nassaji, 2007;
Ranta & Lyster, 2007). As Ranta and Lyster (2007: 152) put it:

[i]n our present view, all corrective feedback moves belong to one of two kinds: either
reformulations or prompts. Reformulation includes recasts and explicit correction because both
these moves supply learners with target reformulations of their non-target output (see Lyster &
Ranta, 1997 for definitions of feedback categories). Prompts include a variety of signals, other
than alternative reformulations, that push learners to self-repair. These include elicitation,
metalinguistic clues, clarification requests, and teacher repetition.

This Copernican turn of sorts is interesting because it unites under the same group
feedback moves such as recasts and explicit correction, two types which have been
traditionally confronted by most authors in terms of implicit vs. explicit. On the other
hand, however, ignoring the explicitness of negative evidence overly simplifies the
multidimensional nature of feedback moves. Therefore, merging the underlying ideas in
Leeman (2000), Carroll (2001), and Ranta and Lyster (2007), a comprehensive taxonomy of feedback should envisage the combination of three components: (1) provision of negative evidence, with various levels of explicitness: implicit and explicit, subdivided into overt error signaling (“Signal error”), error locus and nature spotting (“Spot error”), and rule presentation (“Explain error”); (2) provision of positive evidence (“PE”), and prompting for error repair (“Prompt”). Based on these three components, the following taxonomy is proposed:

Table 3.4 A comprehensive taxonomy of corrective feedback

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

This taxonomy is not intended as definitive classification scheme, but rather as a usable template that assigns independent cells to the vast majority of feedback moves in the literature. By filling in this template, one can obtain a bird’s eye view of the research to date and spot under-researched areas for future work, or compare the overall effectiveness of various feedback types against each other. Notably, this template contains various over-simplifications. First, it is for corrective feedback only. Second, it escapes the complexity of recasts discussed in section 2.2.1. Third, error “locus” and “nature” have been merged under one cell, “Spot error,” but whenever appropriate they should be considered as independent cells. And fourth, this template only contemplates
the existence of implicit feedback moves where positive evidence is provided, such as recasts. This does not mean, however, that feedback moves of the [implicit, -PE] type are impossible. Imagine, for example, the use of irony in the following Tutor (T) – Learner (L) interaction.

L: *Yo gusto el cine. [“I like cinema”]

T: Sí, y yo *me llamo es Luis. [“Yes, and I am called is Luis”]

In this episode, L produces an ungrammatical sentence, and in reaction to that, T produces another ungrammatical sentence that has been discussed earlier in the classroom. This type of response is [implicit] feedback, since by using irony T is indirectly signaling that L has made a mistake, and L may or may not realize this. At the same time, unlike recasts, this type of feedback is [-PE], since it does not reformulate L’s utterance in any way. The occurrence of this type of feedback move, however, though not impossible, is negligible, which is why it is not contemplated in this taxonomy. For illustrative purposes, a list of examples for every node in this taxonomy is provided in Appendix 3.2.

Moving away from classroom-based interaction to the computer laboratory, it could be posited that due to the idiosyncrasy of computers, their potential and limitations, a specific classification of computerized feedback types must be created. Premised on this belief, various authors (e.g., Heift, 2004; Zourou, 2008) have provided their own classifications, most often based on Lyster and Ranta’s (1997) seminal work. The next
pages discuss whether or not the creation of an independent taxonomy of feedback types in CALI is justified from a theoretical or practical perspective.

Hitherto, existing research on computerized feedback has approached the taxonomy issue from two different perspectives: the “how approach” (i.e., the way in which the software analyzes the learner’s communicative moves and presents feedback in response to them), and the “what approach” (i.e., the content of the information conveyed by the feedback). An example of the “how approach” is provided by Garrett (1987), according to whom there are four main types of feedback in relation to error analysis in CALI grammar lessons:

<table>
<thead>
<tr>
<th>Table 3.5 Feedback types in CALI (Garrett, 1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
</tr>
<tr>
<td>Type 2</td>
</tr>
<tr>
<td>Type 3</td>
</tr>
<tr>
<td>Type 4</td>
</tr>
</tbody>
</table>

As it can be observed, Garrett’s (1987) taxonomy comes in very handy to distinguish between so-called traditional vs. intelligent CALI systems. Thus, traditional CALI systems would be those that can handle feedback types up to level three, whereas intelligent CALI systems are able to provide the fourth level of feedback, analyzing the student’s response on the basis of linguistic structure. While this distinction is crucial
from a computational perspective, from a language acquisition standpoint Garrett’s
taxonomy is not detailed enough to distinguish between the different CALI feedback
types that have been reported on. In other words, from an acquisitionist perspective what
the field needs is a taxonomy along the lines of the “what approach” that allows for a
classification of feedback types according to the information they convey, to study their
differential effects on L2 development.

Perhaps the first step in this direction was taken by Heift (2004), who extrapolated to
the CALI environment Lyster and Ranta’s (1997) classic taxonomy of feedback types in
oral interaction. Heift argued that “[d]ue to the medium, feedback in a CAL[I]
environment cannot be identical to feedback in the oral classroom,” and subsumed Lyster
and Ranta’s six-tiered classification into four basic CALI counterparts:

<table>
<thead>
<tr>
<th>Feedback Type</th>
<th>Oral Classroom</th>
<th>CALI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit Correction</td>
<td>You mean…</td>
<td>Correct answer</td>
</tr>
<tr>
<td>Recast</td>
<td>Teacher Reformulation</td>
<td>Correct answer</td>
</tr>
<tr>
<td>Clarification</td>
<td>What do you mean?</td>
<td>Try again!</td>
</tr>
<tr>
<td>Meta-linguistic Feedback</td>
<td>Explanation of error type</td>
<td>Explanation of error type</td>
</tr>
<tr>
<td>Elicitation</td>
<td>Ellipsis</td>
<td>Highlighting</td>
</tr>
<tr>
<td>Repetition</td>
<td>Intonation</td>
<td>Highlighting</td>
</tr>
</tbody>
</table>

Interestingly, Heift first conflated “explicit correction” and “recasts” under one single
category, i.e., “correct answer.” Her point here is that while introducing a correction with
“You mean…” presents the correct form more overtly than recasting, both feedback
moves deprive students from opportunities to correct their mistakes: “[o]nce the correct
answer has been supplied by the system, learner uptake and thus a negotiation of form
between the learner and the CAL[I] program is not an option” (p. 418). While this could
partially hold for a CALI environment where the written sentence is the main unit of
communication (written recasts may preclude uptake more than oral feedback), it does
not capture the potentiality of CALI to simulate oral communication or written
communication beyond the sentence level. Furthermore, CALI systems may invite
students to repair their ill-formed utterances, by e.g., asking or allowing students to
reproduce the correct answer or reformulate their own response after a timed
computerized feedback message has disappeared from the screen, similarly to what
occurs in oral speech. Hence, this conflation between “explicit correction” and “recasts”
is unjustified. Second, Heift conflated “elicitation” and “repetition” under a single CALI
feedback move, i.e., “highlighting,” arguing that the ways in which these are respectively
manifested in the oral classroom, pauses and intonation, are “not possible” in a CALI
environment, which is why she proposed highlighting as the next best thing to account
for both feedback moves. Once again, this conflation is too reductive, since even in the
simplest CALI systems creative use of points of ellipsis and italics can do the job, not to
mention the possibilities of different forms of text enhancement or audiovisual CALI. In
conclusion, then, Heift’s (2004) classification of CALI feedback types clearly
underestimates the capabilities of CALI technology to simulate classroom interaction,
unnecessarily reducing Lyster and Ranta’s (1997) typology.

In contrast to Heift (2004), and switching the focus from CALI to CMC, Zourou’s
(2008) classification of computerized feedback not only does not reduce Lyster and
Ranta’s (1997) typology but expands it considerably, introducing new labels such as
“multiple” and “collective” feedback, “insertion of HTML tags,” and “use of
commenting and advanced markup tools,” among others. However, while this taxonomy
allows for a more fine-grained classification of feedback moves, it fails to group them up based on the components that they share.

To date, no known CALI study has attempted to provide a taxonomy of computerized feedback based on its components along the lines of e.g., Long and Robinson (1998) and Leeman (2000). The closest step in this direction is to be found in the longer-established literature on educational technology, where Dempsey, Driscoll and Swindell (1993) distinguished between three basic types of computerized written feedback based on the information that they convey (see Table 3.7 below).

**Table 3.7** Computerized feedback types (Dempsey et al., 1993)

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) knowledge of response</td>
<td>informs learners whether their response was correct or incorrect.</td>
</tr>
<tr>
<td>(b) knowledge of correct response</td>
<td>additionally informs the learner about what the correct response should be.</td>
</tr>
<tr>
<td>(c) elaborated feedback</td>
<td>provides an explanation for why the learner’s response is correct or incorrect, or allows the learner to review material relevant to the attributes of a correct response.</td>
</tr>
</tbody>
</table>

Arguably, this classification scheme is theoretically based on some of the components of feedback. Thus, “knowledge of response” corresponds to [+explicit negative evidence (Signal error), -PE]; in turn, “knowledge of correct response” maps to [+explicit negative evidence (Signal/Spot error), +PE], and finally, “elaborated feedback” can be defined as [+explicit negative evidence (Explain error), -PE]. However, as it can be noticed, this classification presents two main problems: first, it does not envisage the potential multiple combinations of positive and negative evidence (with different degrees of
explicitness), and second, it does not contemplate the prompting component advocated by Ranta and Lyster (2007).

In conclusion, then, the creation of a specific classification for computerized feedback is unjustified from a theoretical and practical perspective. Ideally, feedback of all sorts, regardless the source (the teacher or a computer) should be classified according to the specific values of its three components, positive evidence, negative evidence, and prompting for repair, in order to gauge their independent and combined contributions to SLA. From a terminological perspective, this calls for a revisitation of the labels that have been manipulated in the literature. How are recasts different from explicit corrections or clarification requests in terms of their components? That is the focus of the next section.

2.4 Revisiting Lyster and Ranta’s (1997) catalog of feedback types

In order to implement the new feedback taxonomy proposed in section 2.3 it becomes necessary to review the definitions of the traditional feedback labels used in the oral interaction literature. Following is an account of Lyster and Ranta’s (1997) six feedback moves: (a) explicit correction, (b) recasts, (c) clarification requests, (d) metalinguistic feedback, (e) elicitation, and (f) repetition.

(a) “Explicit correction refers to the explicit provision of the correct form. As the teacher provides the correct form, he or she clearly indicates that what the student had said was incorrect (e.g., “Oh, you mean,” “You should say”)” (p. 46). In principle, the provision of PE, the correct form, in juxtaposition to negative evidence greatly helps to resolve the blame assignment problem but does not provide a rationale for hypothesis
formation. According to this, Lyster and Ranta’s “explicit correction” moves could be initially be tagged as [explicit (Signal/Spot error), + PE, - prompt]. Further in their discussion, Lyster and Ranta (p. 49) noted that the label “explicit correction” also accounts for combined feedback moves where elicitation (e.g., “Please, try again”) accompanies either a recast or an explicit correction. Therefore, by “explicit correction” is meant a feedback move where positive evidence is always provided and negative evidence is provided either implicitly in conjunction with prompting, or explicitly, with or without prompting for error repair. Thus, the full-fledged tagging of “explicit correction” should be [implicit, +PE, +prompt] U [explicit (Signal/Spot error), +PE, +/- prompt], which fills in the following categories in the chart:

Table 3.8 Lyster and Ranta’s (1997) definition of “explicit correction” revisited

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

(b) “Recasts involve the teacher’s reformulation of all or part of a student’s utterance, minus the error… [They] are generally implicit in that they are not introduced by phrases such as “You mean,” “Use this word,” and “You should say.” However, some recasts are more salient than others in that they may focus on one word only, whereas others incorporate the grammatical or lexical modification into a sustained piece of discourse” (Lyster & Ranta, 1997: 46-7). According to this, “recasts” should be tagged as [implicit].
An arising question here is whether a recast followed by a request for error repair can still be labeled as an implicit type of feedback. Regarding this, Lyster and Ranta (1997: 49) argue that when recasts are followed by elicitation moves (e.g., “Please, try again”), the implicit nature of recasts is corrupted, and they turn into “explicit correction.” Consequently, recasts should be tagged as [implicit, +PE, -prompt], thus checking the following cell in the checkchart:

**Table 3.9** Lyster and Ranta’s (1997) definition of “recasts” revisited

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) “Clarification requests (…) indicate to students either that their utterance has been misunderstood by the teacher or that the utterance is ill-formed in some way and that a repetition or a reformulation is required. This is a feedback type that can refer to problems in either comprehensibility or accuracy, or both. (…) A clarification request includes phrases such as ‘Pardon me’ (…). It may also include a repetition of the error as in ‘What do you mean by X?’” (Lyster & Ranta, 1997: 47). According to this definition, “clarification requests” are particularly hard to label. They are not implicit because they do signal the occurrence of a breakdown in communication and, unlike recasts, they cannot be perceived as a mere repetition of the student’s utterance or another type of
back-channeling strategy. However, when compared to many types of explicit feedback that clearly indicate the nature of the error, clarification requests are vaguer (is the problem in lexis, grammar, pronunciation?). Consequently, from a conservative perspective “clarification requests” could be tagged as [-PE, +prompt] and some level of explicitness ranging from [Signal error] to [Spot error] at best. On this token, the inclusion of metatalk would corrupt the ambiguous nature of clarification requests, i.e., that a repetition or a reformulation is needed due to problems in comprehensibility or accuracy, in favor of the latter. Therefore, the final tagging for “clarification requests” should be [explicit (Signal/Spot error), -PE, +prompt].

Table 3.10 Lyster and Ranta’s (1997) definition of “clarification requests” revisited

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

(d) “Metalinguistic feedback contains either comments, information, or questions related to the well-formedness of the student’s utterance, without explicitly providing the correct form. Metalinguistic comments generally indicate that there is an error somewhere (“Can you find your error?”… “No, not X”… or even just “No.”). Metalinguistic information generally provides either some grammatical metalanguage that refers to the nature of the error (e.g., “It’s masculine”) or a word definition in the case of lexical
errors. Metalinguistic questions also point to the nature of the error but attempt to elicit the information from the student (e.g., “Is it feminine?”).” From this definition, it can be drawn that metalinguistic comments, information, and questions do not necessarily intend to push learners to fix their errors, but rather push them to make some sort of metalinguistic introspection to reflect upon the nature of their errors. In other words, as argued by Lyster and Ranta (1997: 48), metalinguistic questions can be simply answered with a yes or no (e.g., “Do we say that in French?”), whereas questions that explicitly ask learners to fix a linguistic error (e.g., “How do we say that in French?”) are considered elicitation moves, discussed below. Wrapping up, then, “metalinguistic feedback” as understood by Lyster and Ranta can be tagged as [explicit, -PE, -prompt], which fills in the following cells in the taxonomy:

Table 3.11 Lyster and Ranta’s (1997) definition of “metalinguistic feedback” revisited

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

(e) “Elicitation refers to at least three techniques that teachers use to directly elicit the correct form from the student. First, teachers elicit completion of their own utterance by strategically pausing to allow students to “fill in the blank” as it were (…). Such “elicit completion” moves may be preceded by some metalinguistic comment such as
“No, not that. It’s a…” or by a repetition of the error (…). Second, teachers use questions to elicit correct forms (… e.g., How do we say X in French?”). Third, teachers occasionally ask students to reformulate their utterance” (Lyster & Ranta, 1997: 48).

Based on this, it can be concluded that “elicitation” maps to [explicit, - PE, +prompt], which checks the following cells in the proposed taxonomy:

Table 3.12 Lyster and Ranta’s (1997) definition of “elicitation” revisited

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

(f) “Repetition refers to the teacher’s repetition, in isolation, of the student’s erroneous utterance. In most cases, teachers adjust their intonation so as to highlight the error” (Lyster & Ranta, 1997: 48) Repetition, as these authors noted, often occurs with all other feedback types with the exception of recasts. However, given its frequency overall, it is the feedback move it accompanies that overrides repetition in terms of illocutionary force: clarification requests (“What do you mean by X?”), metalinguistic feedback (“No, not X. We don’t say X in French”), elicitation (“How do we say X in French?”), and explicit correction (“We don’t say X in French; we say Y”). Following this rationale, repetition can be tagged as [explicit (Signal/Spot error), -PE, +prompt].
Table 3.13 Lyster and Ranta’s (1997) definition of “repetition” revisited

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

In conclusion, according to the discussion above, Lyster and Ranta’s (1997) six feedback types can be tagged as follows:

(a) Explicit correction: [implicit, +PE, +prompt] U [explicit (Signal/Spot error), +PE, +/- prompt];
(b) Recasts: [implicit, +PE, -prompt];
(c) Clarification requests: [explicit (Signal/Spot error), -PE, +prompt];
(d) Metalinguistic feedback: [explicit, -PE, -prompt];
(e) Elicitation: [explicit, -PE, +prompt];
(f) Repetition: [explicit (Signal/Spot error), -PE, +prompt].

Table 3.14 below visually illustrates the re-analysis of Lyster and Ranta’s (1997) feedback types based on their individual components:
### Table 3.14 Reinterpretation of Lyster and Ranta’s (1997) feedback taxonomy according to the basic components of feedback moves

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal error</td>
<td>Spot error</td>
</tr>
<tr>
<td>PE Prompt</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>LABEL</td>
<td>Recast</td>
<td>Explicit correction</td>
</tr>
</tbody>
</table>
As it can be observed from Table 3.14 above, the Signal and Spot error columns in this taxonomy completely mirror each other. Therefore, Table 3.14 can be simplified by collapsing these two columns into one, as shown in Table 3.15 below. In contrast, the Explain error column is slightly different because it does not contain clarification requests, repetition, or explicit correction. This is because adding metalinguistic explanations to clarification requests or repetition corrupts the ambiguous nature of these feedback moves, completely solving the blame assignment problem and stressing the fact that the breakdown in communication is not due to problems of comprehension but accuracy. Similarly, adding metalinguistic explanations to explicit corrections overrides the illocutionary force of these moves. With this precision in mind, Lyster and Ranta’s (1997) feedback catalog clearly boasts remarkable strengths. First, even if this taxonomy was aimed at drawing a map of naturally occurring feedback types, it impressively populates most of the potential combinations of the three dimensions distinguished here: negative evidence, positive evidence, and prompting for error repair. Also, Lyster and Ranta’s labels can be used to distinguish between different feedback types that share the same components, such as clarification requests, elicitation, and repetition (explicit, -PE, +prompt). On a negative note, however, this taxonomy falls short in distinguishing different granularity levels of explicit feedback, which based on Carroll (2001) have been narrowed down here to three, i.e., Signal error, Spot error, and Explain error.
Table 3.15 Simplified reinterpretation of Lyster and Ranta’s (1997) feedback taxonomy

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>IMPLICIT</th>
<th>EXPLICIT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signal/Spot error</td>
<td>Explain error</td>
</tr>
<tr>
<td>PE</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Prompt</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>LABEL</td>
<td>Recast</td>
<td>Explicit correction</td>
</tr>
</tbody>
</table>

Section summary: Feedback

In the SLA literature “corrective feedback” typically refers to a discourse move that evaluates the quality of a learner’s utterance. Corrective feedback is therefore (1) reactive, i.e., contingent upon the execution of a process on the part of the learner; and (2) positive or negative, depending on whether it praises a successful execution or it signals the occurrence of an error. While there is little discrepancy around the definition of corrective feedback, there is some debate surrounding its typology. Until very recently, the pivotal criterion to distinguish feedback types has been the explicitness of the negative evidence they encapsulate. Yet, feedback explicitness has been understood from two different angles: (1) error signaling, i.e., how overtly the occurrence of an error is signaled, and (2) error explanation, i.e., how much detail is provided to explain the reason of the error. While these perspectives are clearly different, both are premised on the linguistic features of feedback, that is, feedback is envisaged from the tutor’s side; however, as recent studies point out, feedback is a two-sided phenomenon that can also be tackled from a psycholinguistic perspective, i.e., how learners decode.
This double nature of feedback calls for a re-evaluation of the notion of explicitness. From the error signaling perspective, feedback moves cannot be classified as explicit or implicit merely based on the presence or absence of messages such as “Your answer is wrong.” For example, recent research shows that recasts, which have typically been placed at the implicit end of the continuum, can actually be perceived very explicitly by learners depending on their segmental, supra-segmental, and discourse features. By the same token, simply saying “That’s wrong” does not entail that learners will automatically interpret that statement as a metalinguistic comment, and if they do, they still have to determine which level of language is affected and why there is an error. This is precisely where error explanation comes in handy, and yet stating a rule does not imply its immediate understanding on the part of the learners.

In light of these different interpretations of feedback explicitness, there is a blatant need for theoretical and terminological normalization. In order to extrapolate refined conclusions from the now vast and often contradictory body of research on feedback, it does not suffice to classify feedback moves as “implicit” or “explicit,” or as “more” or “less” the former or the latter; it is necessary to measure explicitness more precisely, taking into consideration both the linguistic and psycholinguistic dimensions. One way to do this is by observing how feedback moves may differently facilitate the actions that learners are supposed to carry out in order to process feedback for acquisition, all the way from the exposure to the feedback message up to the intended goal of L2 development or, more ambitiously, interlanguage restructuring. These actions are mainly four, namely: (1) recognizing the tutor’s corrective intent; (2) spotting the error locus; (3) recognizing the nature of the error (i.e., which level of language is implicated —speech act, lexical
choice, morphosyntax, linear order, articulation); and (4) forming hypotheses.

Consequently, the higher the level that feedback facilitates, the more explicit it is. Under this definition, then, feedback with metalinguistic explanation represents the most explicit type in the continuum. Nonetheless, access to the aforementioned levels must be attested via methodological tools like measures of uptake, noticing via stimulated and immediate recalls, and tests of metalinguistic awareness. All these procedures have limitations, but they constitute so far our only peepholes into the learners’ minds.

Now, various factors, such as the nature of the practice and the consistency and focus of the feedback, may moderate feedback explicitness. For example, in all existing CALI studies even the less explicit types of feedback used (i.e., “Your answer is wrong”) typically facilitate access up to level three (recognizing the corrective intent, error locus, and error nature), since they involve intensive focused feedback in the written mode. Therefore the explicitness of feedback cannot only be measured by intrinsic factors (such as explicitness) but also by external factors. Furthermore, the explicitness of negative evidence is only one intrinsic factor of feedback. Feedback moves are multi-dimensional in nature, since they may also provide positive evidence and/or prompt learners to fix their errors. Each of these components may have an impact on L2 development, and therefore future investigations of feedback must isolate them to unveil their separate and combined effects. With that goal in mind, I have proposed a classification scheme that includes the possible combinations of negative evidence (with different degrees of explicitness), positive evidence, and prompting for error repair, and I have used it to revisit Lyster and Ranta’s (1997) influential classification of feedback types. In the next chapter, I will use it to reinterpret the published literature on CALI feedback.
3. Grammatical form

Empirical research on L2 grammar acquisition is typically geared around a specific structure. Most published studies include a section especially devoted to justify the choice for targeted structure in the research design, and many limitations sections ask readers not to overgeneralize findings, calling for replication studies with other forms. Despite this fact, the role of type of grammatical form remains largely unexplored in the SLA literature. As pointed out by DeKeyser (2005: 11), “[r]elatively few studies have actually attempted a systematic empirical investigation […] by comparing acquisition for a broad range of language structures.” Perhaps the best places to search for such type of evidence, DeKeyser adds, are studies in the areas of fossilization (e.g., Han, 2000; Lardiere, 1998, 2000; Long, 1997, 2003), ultimate attainment by adult learners (e.g., Birdsong, 1992; DeKeyser, 2000; DeKeyser, Ravid, & Alfi-Shabtay, 2005; Flege, Yeni-Komshian, & Liu, 1999; McDonald, 2000), and order of acquisition (Eckman, Bell, & Nelson, 1988; Eckman, Moravcsik, & Wirth, 1989; Gass, 1979; Pienemann & Johnston, 1987; Pienemann, Johnston, & Brindley, 1988).

In my view, existing literature on the effects of type of linguistic form on second language acquisition can be arranged into three main areas, depending on whether the focus is on the L1, the L2, or the interlingua (IL). Research with a focus on the L1 has spawned taxonomies based on order of acquisition; research with a focus on the L2 (independently from, or in relation to, the L1) has attempted to categorize structures in terms of their relative difficulty; finally, research looking at the IL has introduced the notions of learnability, fossilization, and ultimate attainment. Now, some of these notions
are more difficult to define (and investigate) than others. For example, based on research on acquisition order, it can be concluded that a structure A is acquired earlier than a structure B in the L1; similarly, research on IL development can show that, for a specific language combination, an L2 structure A is more learnable than an L2 structure B at a given point; on the contrary, though, it is far more difficult to posit, on the basis of a solid theoretical foundation, that an L2 structure A is less “difficult” than an L2 structure B merely by comparing their inherent features or their differences with respect to their L1 equivalents.

In an attempt to solve the problem of determining what makes grammar more or less difficult, researchers have come up with different solutions. Some have avoided operational definitions altogether, appealing to instructors’ intuitions (e.g., Robinson, 1996). Others have adopted a cognitive rather than linguistic perspective, addressing the complexity of the rules that govern the use of a structure. For example, Reber (e.g., Reber, Allen, & Reber, 1999) argued that the greater the complexity of a rule, the less likely that its explicit presentation will prove effective, or to put it another way, conscious learning of rules can only prove effective if the rules are simple. Yet, he did not provide a detailed list of features to classify rules as complex or simple. From the opposite standpoint, Hulstjin and de Graaff (1994) hypothesized that the advantage of explicit rule presentation arises precisely when rules are complex, because they save learners considerable time in discovering the intricacies of the linguistic phenomenon that they describe, while simple formal phenomena may be salient enough in the input for L2 learners to discover them spontaneously. Differently from Reber, Hulstjin and de Graaff did provide a definition of rule complexity, understood as “the number (and/or the type)
of criteria to be applied in order to arrive at the correct form” (p. 103). To illustrate this, they presented the following examples of two purely formal (phonological) and two semantic (aspectual) rules (p. 102-3):

Example 1: In language M, verbs have different endings for two aspects, perfect and non-perfect; in language N, however, verbs have different inflections for three aspects, perfective, durative, and punctual. Example 2: In language X there are two plural suffixes for nouns: the suffix –s is added to singular nouns ending on a vowel, and the suffix –os is added to nouns ending on a consonant. Language Y has the same two suffixes: -s for nouns ending on a vowel and for nouns ending on a consonant containing a front vowel in the penultimate syllable, -os for nouns ending on a consonant and containing a back vowel in the penultimate syllable. If we were to make flow charts for the derivation of inflections in languages M and N and for the derivations of the plural suffix in languages X and Y, the charts for languages M and X would contain fewer steps than those for N and Y respectively.

In addition to rule complexity, Hulstijn and de Graaff introduced the notions of rule scope and reliability, both of which may have implications for explicit L2 instruction. By rule *scope* is meant the number of cases covered by the rule, whereas *reliability* refers to the extent to which the rule is valid. Specifically, the scope of a rule is large if it covers more than 50 cases, while reliability is high if it applies in 90% or more of all cases. To illustrate this, Hulstijn and de Graaff presented several examples concerning noun gender in German (p. 103, taken from Mills, 1986: 33):

(1) Of the ca. 15,000 singular nouns ending on –e about 13,500 (90%) are feminine.
(2) Of the 15 monosyllabic nouns beginning with Kn-, 14 are masculine (93%).
(3) Of the 107 monosyllabic nouns ending on –CnasalC, 75 are masculine (70%).
(4) Of the 45 nouns ending on –ier, 27 are neuter (60%).

Following Hulstijn and de Graaff’s definitions, then, rule (1) has a large scope because it covers more than 50 cases and high reliability because it applies in 90% of the cases; on the opposite extreme, rule (4) has small scope and low reliability; in turn, rule (2) has small scope but high reliability, and rule (3) is the reverse. The notions of rule scope and
reliability are relevant for research on explicit instruction because for practical purposes, as posed by Hulstijn and de Graaff, if rules have a small scope and low reliability, explicit grammar teaching may not be worth it anyway. Finally, in addition to the notions of rule complexity, scope, and reliability, Hulstijn and de Graaff discussed the role of a purely linguistic factor, semantic redundancy, in explicit instruction. Specifically, they argued that for the development of L2 comprehension explicit instruction can be more beneficial for forms with a semantic contribution (e.g., the –s suffix for plural marking in English) than for forms that are semantically redundant (e.g., the –s suffix for third person singular), since in the latter case learners will understand the meaning from the surrounding context anyway (e.g., the adjacent third person pronoun). Conversely, when it comes to L2 production, explicit instruction may prove more beneficial for forms that are semantically redundant, because learners may be prone to avoid producing these forms as opposed to those that do bear a semantic value.

Overall, Hulstijn and de Graaff’s (1994) proposal constituted one of the first serious attempts to entangle the interaction between type of instruction (explicit vs. implicit) and type of object of instruction (linguistic form). However, as mentioned earlier, it tackled the problem mostly from a cognitive perspective (i.e., the nature of the metalinguistic rules) rather than a linguistic one (i.e., the inherently linguistic features of a given form). Also, their formulation of rule complexity as “the number (and/or the type) of criteria to be applied in order to arrive at the correct form” needs to be expanded, since it cannot determine which of two rules is more complex if they contain the same number of processes to arrive at the correct form. In other words, Hulstijn and de Graaff do not provide guidelines to determine which types of criteria or cognitive processes are more
complex or demanding. To fill in this gap, authors such as Ellis (2007) have resorted to Pienemann’s (1998) Processability Theory to compare and rank, by purported order of acquisition, the processing operations that underlie different grammatical structures (e.g., category vs. phrasal procedure). Along these lines, and based on a meta-analysis of empirical studies, Goldschneider and DeKeyser (2001) proposed that the order of acquisition of grammatical morphemes in the L2 is determined to a large extent by at least five linguistic properties: perceptual salience, semantic complexity, morpho-phonological regularity, syntactic category, and frequency.

Perceptual salience refers to how easy or difficult it is to perceive a given structure in written or aural input. Leaving aside suprasegmental factors such as emphatic stress, Goldschneider and DeKeyser posited that perceptual salience is determined by at least three subfactors. The first subfactor is phonetic substance, i.e., the number of phones of the morpheme. For example, the plural -s morpheme in English has an average of 1.33 phones, resulting from dividing 4 phones by 3 allomorphs, [s, əz]. The second subfactor is syllabicity, i.e., whether or not the morpheme contains a vowel. The third factor is sonority, i.e., the perceptual auditory quality of a sound. For example, according to Laver’s (1994: 504) hierarchy (vowels [low > mid > high] > glides > liquids > nasals > obstruents [fricatives > affricates > stops]), the present progressive -ing morpheme has a sonority score of 11, which results from adding up 7 points of the high front vowel and 4 points of the nasal. According to Goldschneider and DeKeyser, the greater the phonetic substance, syllabicity, and sonority, the greater the perceptual salience of a morpheme, and the more salient a morpheme is, the earlier it will be acquired.
Semantic complexity refers to the number of meanings conveyed by a particular form. For example, in English the plural -s morpheme has a semantic complexity of 1, since it only expresses number, whereas the third person singular -s scores 3, because it expresses person, number, and present tense. However, determining the semantic complexity of a given form might be subject to debate (p. 25). For example, according to Brown (1973) English articles (a vs. the) should score 1 point because they express “definiteness,” while for other authors they should be assigned a score of 2 points because they can also convey a “generic meaning,” as in “The beaver has a flat tail.” This debate aside, the thesis here is that semantic complexity is related to acquisition indirectly, that is, the greater the semantic complexity of a form, the later it will be acquired.

Morpho-phonological regularity refers to whether or not a form presents allomorphy and homophony. For example, the English plural -s morpheme presents three allomorphs [s, z, əz] and is homophonous with both the saxon genitive and the third person singular morphemes. Conversely, English present progressive -ing has no allomorphs and is not homophonous with other morphemes. The prediction here is that the more regular a morpheme is (the fewer allomorphs, and the fewer the cases of homophony), the earlier it will be acquired.

Syntactic category refers to the characteristics of the morpheme as defined by Functional Category theory. According to Goldschnieder and DeKeyser, syntacticians have long abandoned the thesis that the difficulty to acquire a form is directly correlated to the number of syntactic transformations needed. Rather, research by Zobl and Liceras (1994) suggests that the crucial concept here is quality over quantity, with lexical items
likely to be acquired before functional ones, and within each group, free morphemes appearing to be acquired before bound ones.

Finally, frequency refers to the number of times a given structure occurs in the speech, with the assumption that the higher the frequency, the sooner it will be acquired. Following this logic, Goldschneider and DeKeyser posited that English articles should be acquired earlier than the Saxon genitive, since based on a sample of Brown’s (1973) frequency data they stand on a ratio of 8:1.

Arguably, Goldschneider and DeKeyser’s (2001) account is a remarkable attempt to provide a list of strictly linguistic parameters that may determine the difficulty of L2 forms. In addition, based on a meta-analysis of empirical studies they concluded that phonological salience and syntactic category had the highest correlation with accuracy, suggesting that they are better predictors of acquisition, although all the factors combined did predict accuracy too. In this context, future research should pursue two different goals. First, it should attempt to determine the relative weight of the aforementioned linguistic factors in acquisition, and whether and how they work in concert. And second, it should account for syntax more extensively, since Goldschneider and DeKeyser’s proposal was restricted to a set of English morphemes.

Precisely comparing morphology and syntax, and also vocabulary, Gass, Svetics, and Lemelin (2003) posited that there are three main linguistic factors that determine the efficacy of form-focused pedagogical interventions: abstractness, complexity, and the L1-L2 differences. Based on these factors, Gass et al. hypothesized that form-focused attention may have the most facilitative effect on vocabulary and the least on syntax, with morphology in the middle. Specifically, they argued:
“attention might be a less significant factor in learning when there is greater complexity (and abstractness), because in addition to there being a problem for learners in isolating forms (isolation undoubtedly being easier for morphology than for syntax), they also have to learn the correctness of those forms.”

Yet, Gass et al. did not provide a clear-cut set of criteria to determine how syntax, morphology, and vocabulary are purportedly more or less complex, abstract, or L1-L2 deviant than one another, or how different forms within each domain can be arranged hierarchically along these parameters.

Much more exhaustively, in what possibly constitutes the most ambitious review to date, DeKeyser’s (2005) article “What makes learning second-language grammar difficult? A review of issues,” proposed three different classification criteria for “difficulty” of grammatical forms, i.e., problems related to meaning, problems of form, and problems of form-meaning mapping. Problems of meaning arise, for example, when the meaning of the L2 form can constitute a source of difficulty due to its novelty, abstractness, or both. Thus, articles, classifiers, grammatical gender, and verbal aspect are difficult to acquire for speakers of L1s that either lack them or use a very different system. In turn, problems of form are related to the number of choices that the speaker has to make in order to pick all the right morphemes and allomorphs to express or decode a meaning, as well as putting them together in the right place. To this regard, existing literature indicates that e.g., word order is typically nonproblematic past the initial stages of acquisition, whereas rich morphology requires high amount of time and special attention to form while processing input. However, as posited by DeKeyser and Sokalski (1996: 620) the same structure may present different problems of form depending on whether the intended goal is reception or production. In other words, some structures are
easier to recognize than to produce and vice versa: “[a] morphological complex structure may be easier to notice but harder to produce correctly than a simpler structure; a simpler structure may be inconspicuous and therefore harder to notice, but easier to produce by virtue of its simplicity.” Finally, form-meaning problems arise when the link between the form and its meaning is not transparent. According to DeKeyser, such lack of transparency can be due to at least three factors: redundancy (the form in question is not semantically necessary because its meaning is expressed elsewhere in the utterance—e.g., verb endings when the subject is explicit); optionality (the form may be present or absent, often contributing to subtle pragmatic differences that are very hard to grasp—e.g., pronoun dropping in Spanish); and opacity (the correspondence between the form and its meaning is ambiguous—e.g., the -s morpheme in English, which can be the third person singular of the verb, the plural of the noun, or the genitive of the noun).

In relation to all problems of meaning, form, and form-meaning connections, many authors have approached the notion of difficulty of a language structure by looking at its salience and its frequency of occurrence in the input, factors which, to a certain extent, may be manipulated through certain pedagogical interventions (see e.g., Sharwood-Smith, 1991: 121 on “naturally occurring” vs. “deliberately engineered” salience). Overall, most researchers have viewed salience in terms of how easy it is to perceive a given structure, either aurally or visually (i.e., perceptual salience). In turn, frequency of forms, meanings, and form-meaning connections may also play a role in determining the difficulty of a structure. Thus, as pointed out by DeKeyser (personal communication, June 4, 2006), “if a form occurs multiple times with different meanings, or a meaning occurs multiple times with different forms, this high frequency on one side of the
equation actually makes learning more difficult, whereas high frequency of form-
meaning mapping makes it easier.”

Hitherto, very few SLA studies have compared the relative difficulty of two or more
morphological or grammatical forms on the basis of a comprehensive list of features such
as the ones proposed by Hulstijn and de Graaff (1994), Goldschneider and DeKeyser
(2001), and DeKeyser (2005). Authors have often focused on a very inexhaustive set of
features, often leading to conflicting classifications. For example, as discussed by
DeKeyser (2005):

[T]he same structure, third person –s in English as a Second Language (ESL), was classified by
Krashen (1982) as easy to learn because it is simple, and by Ellis (1990) as hard to learn because it
is complex. One might think that the reason for this discrepancy is that Krashen was dealing with
learning in the narrow sense here (as opposed to implicit acquisition) and Ellis with a broader
meaning of learning. But when one looks at their reasons for classifying this structure as easy or
difficult, it is clear that they used different criteria for deciding on the complexity of –s: Krashen
pointed to the simple dichotomous choice between supplying this simple morpheme or not,
whereas Ellis, referring to Pienemann (1984), pointed to the long-distance relationship between
the grammatical number of the subject and the presence or absence of –s on the verb. Nor is the
disagreement due to a mere focus on formal complexity by Krashen versus a broader look at the
form-function relationship by Ellis: The latter actually goes beyond Krashen by considering even
the form-function relationship for –s to be simple (‘‘transparent’’); it is only because of the
processing operations required that Ellis considered the structure to be complex (1990, p. 167).

Recently, however, studies have addressed the issue of type of linguistic form in
greater depth, considering a larger number of factors. For example, in the most recent
published study on the role of linguistic form on type of corrective feedback, Ellis (2007)
included eight different factors to compare the learning difficulty of English regular past
tense -ed morpheme and comparative constructions: grammatical domain, input
frequency, learnability, explicit knowledge, scope, reliability, formal semantic
redundancy, and experts’ opinion. Specifically, Ellis posited that past tense –ed is easier
to acquire than comparative constructions because it constitutes a purely morphological
feature, as opposed to involving syntactic operations (grammatical domain); it is two
times more frequent than comparatives (input frequency); it invokes a processing
operation that needs to be mastered earlier in the developmental sequence, i.e., “category
procedure” as opposed to “phrasal procedure” (learnability); its underlying rule is easier
to understand (explicit knowledge) and it applies in a larger number of cases (scope); and
second language tutors deem it to be an easier construction (experts’ opinion). On the
other hand, the rules for the use of both structures apply in less than 90% of the cases
(reliability) and both structures contribute meanings that can be easily inferred from the
neighboring constituents most of the time (formal semantic redundancy), so these latter
two factors do not help in making predictions. Clearly, then, in order to compare two or
more structures in terms of their relative difficulty a multiple number of factors ought to
be considered.

The present dissertation analyzes the effects of different computer-based pedagogical
interventions (involving various degrees of learner’s Agency and type of Feedback) in the
development of two L2 grammatical structures with posited different levels of difficulty.
These two structures under comparison (Spanish Present subjunctive and Spanish
Preposition pied-piping in adjectival relative clauses) are separately discussed in Chapter
5 using DeKeyser’s (2005) operational criteria.

**Section summary: Grammatical form**

The role of type of grammatical form in second language learning and instruction
remains largely unexplored (DeKeyser, 2005: 11). Hitherto, the incipient literature has
tackled the construct from three different perspectives, namely the L1, the L2, and the
interlingua. Research with a focus on the L1 has spawned taxonomies based on order of acquisition; research with a focus on the L2 has attempted to categorize structures in terms of their relative difficulty; finally, research analyzing the IL has introduced the notions of learnability, fossilization, and ultimate attainment.

While data-driven research premised on the L1 and the IL can yield relatively objective taxonomies of grammatical structures, it is harder to posit, on the basis of strong theoretical grounds, whether an L2 structure is more or less “difficult” than another just by looking at their inherent linguistic features or their differences from their L1 equivalents. To tackle this problem, authors have proposed different solutions. Some have avoided operational definitions altogether, appealing to instructors’ intuitions (e.g., Robinson, 1996). Others, in turn, have adopted a mainly cognitive approach, addressing the nature of the rule that governs the use of a structure rather than its linguistic features per se (e.g., Hulstijn & de Graaff, 1994; Reber et al., 1999). For example, Hulstijn and de Graaff proposed that rules can be compared in terms of their complexity (number of processes involved to arrive at the L2 form), scope (number of cases covered by the rule), and reliability (percentage of cases covered by the rule).

In contrast, a third group of studies addressed the issue from a purely linguistic perspective. For example, Goldschneider and DeKeyser (2001) proposed that the difficulty of an L2 morphological form may be determined by at least five linguistic properties: perceptual salience (how easy it is to notice the structure in the input), semantic complexity (how many meanings are contributed by the form), morphophonological regularity (whether or not the form presents allomorphic variation and homophony), syntactic category (whether the form involves lexical or functional
operations and whether it is free or bound), and frequency (how often the form occurs in the input).

Goldschneider and DeKeyser’s (2001) proposal was later extended by DeKeyser (2005) to account for grammatical structures in general, as opposed to morphemes exclusively. Specifically, DeKeyser proposed three classification criteria for linguistic difficulty, i.e., problems related to meaning, form, and form-meaning mapping. Problems of meaning arise when the meaning of the L2 form can constitute a source of difficulty due to its novelty, abstractness, or both; problems of form are related to the number of choices that the speaker has to make in order to pick all the right morphemes and allomorphs to express or decode a meaning, as well as putting them together in the right place; finally, problems of form-meaning mapping arise when the link between the form and its meaning is not transparent, due to redundancy (the form is not semantically necessary), optionality (the form may be present or absent), and/or opacity (the correspondence between the form and its meaning is ambiguous). Finally, in addition to these problems many authors have approached the notion of difficulty of a language structure by looking at its salience (how easy it is to perceive the structure) and frequency (how often it occurs, in terms of form, meanings, and form-meaning mapping).

Existing empirical studies have sometimes showcased contradicting classifications of grammatical structures in terms of their difficulty, differently categorizing the same form as “easy” or “difficult.” Arguably, this is due to the fact that authors have sometimes made judgments based on a very limited set of criteria. Consequently, in order to avoid this future research must analyze grammatical structures based on an exhaustive set of cognitive and linguistic features.
CHAPTER 4.
REVIEW OF THE LITERATURE

Over the last three decades the field of SLA has witnessed a heated debate on the effects of practice and feedback on L2 grammar development, per se or in contrast to the provision of input. Positions have shifted from denying their roles, to admitting their facilitative effects, to analyzing the particular subtypes that yield the best results. However, little attention has been devoted to studying whether the facilitative effects of practice and feedback are conditioned by the agency of the learner, who can take an active part in the interactive process or act as a mere observer. In addition, the effects of practice and feedback may be moderated by a number of factors, including learners’ individual differences and the grammatical form under instruction. The present chapter highlights the main accomplishments and shortcomings of the empirical literature on the three variables investigated in this dissertation: Learner’s agency, i.e., the compared effects of Practice vs. Exposure to practice (Section 1); type of corrective Feedback (Section 2), and type of grammatical Form (Section 3).

1 Learner’s agency

The following review of the literature is structured into two sections. Section 1.1 discusses the empirical literature on Practice, while section 1.2 does the same for Exposure to practice. A comprehensive summary of both sections is provided at the end.
1.1 Practice

The next sections provide a review of empirical research on practice since the ground-breaking formulation of Krashen’s Comprehensible Input Hypothesis to the present date. Section 1.1.1 revises the backbones of Krashen’s framework, where the role of practice was neglected. Sections 1.1.2 and 1.1.3 revisit the first reactions against Krashen, based on the historic evaluations of the Canadian French immersion programs and White’s theoretical critiques. In turn, the next three sections summarize the body of work of three different strands of research providing empirical support for the role of practice. Section 1.1.4 presents the results of studies under Processing Instruction. Section 1.1.5 focuses on research under the Output Hypothesis. Finally, section 1.1.6 reviews the work premised under the Interaction Hypothesis.

1.1.1 Krashen’s Input Hypothesis

Typically, practice is assumed as a necessary condition for adult L2 learning, both by lay people and researchers (e.g., DeKeyser, 2007). However, even today it is not rare to come across recommendations like the following: “If you want to learn Spanish, watch a lot of movies, read a lot of books, and listen to music in Spanish,” actions which, as discussed in Chapter 3, do not qualify as deliberate practice but as exposure to input (or its synonym incidental practice). As a matter of fact, for many years this was not exclusively the belief of lay people but the informed stand of many academic researchers under the paradigm of Krashen’s Comprehensible Input Hypothesis (e.g., 1978; 1982; 1984; 1985; 1989, and elsewhere), according to which learners acquire an L2 incidentally and subconsciously by being exposed to, and comprehending, input that is one step beyond
their competence level ($i+1$). According to Krashen, comprehending $i+1$ is not one particular way of acquiring an L2; it is the “only true cause of second language acquisition” (Krashen, 1984: 61), and failed outcomes are explained by affective barriers alone. Therefore, under this hypothesis the role of language instructors is restricted to creating an amiable environment and procuring learners with input that is comprehensible, which can be best achieved via preemptive simplification strategies such as the ones present in foreigner- or teacher-talk. Interestingly, in Krashen’s view, practice, back then almost exclusively understood in its productive mode, is just a mere sign that acquisition has taken place, and any potential benefits that may result from it are due to its utilization by the learner as a source of self-input. Furthermore, practice of mechanical nature is deemed to obstruct the learner’s focus on meaning, thereby not only not facilitating but impeding acquisition (e.g., Krashen, 1982; Prabhu, 1987). Similarly, feedback, arguably one typical component of pedagogic practice, is deemed to play no facilitative role either because learners do not notice it or make any use of it (e.g., Krashen, 1984). In sum, according to Krashen, form-focused pedagogical interventions, whether instantiated as metalinguistic presentation, practice, or feedback, can only help conscious learning and monitoring, but neither of these is relevant to acquisition.

1.1.2 The Canadian French immersion programs

For several decades now, Krashen’s ideas have been rebated. While no one denies that exposure to comprehensible input is necessary (e.g., Carroll, 2001; Gass, 1997; Larsen-Freeman & Long, 1991), there is substantial evidence attesting that it alone is not sufficient for SLA. One often-cited example is the case of the evaluations of the Canadian
French immersion programs that started out in the late 1960s (see e.g., Lambert & Tucker, 1972; Swain, 1978, 1981, 1991; Swain & Lapkin, 1982, 1986). According to these evaluations, L1 French pupils exposed to great amounts of L2 English input over long periods of time rated comparably to native speakers on measures of reading and listening comprehension, while they failed to achieve native-like competence in production tests, with a blatant lack of basic vocabulary and sustained grammatical errors in domains such as verb tenses, prepositions, and gender agreement (Harley & Swain, 1978, 1984; Lapkin, Hart, & Swain, 1991). The same was observed for adult learners living abroad, as shown by a vast body of studies on fossilization (see the recent books by Han, 2004; Han & Odlin, 2006 for comprehensive reviews on the last three decades of research). Clearly, then, exposure to comprehensible input, while necessary, may not suffice for SLA, but why?

**1.1.3 White’s Incomprehensible Input Hypothesis**

Many arguments have been presented to account for the insufficiency of exposure to comprehensible input for SLA. For example, White (1989; 1991) claimed that it is actually *incomprehensible* rather than comprehensible input that may trigger acquisition. To illustrate this, White hypothesized how passive structures in English may be incorporated into a learner’s interlanguage on the basis of input alone. Arguably, a learner who is yet to acquire the passive voice is bound to interpret a sentence such as *John was kissed by Mary* as an active sentence, where John is the kisser. However, this putative learner will have a conflict upon hearing a sentence like *The book was read by John*, since books cannot read. Similarly, if the learner hears the sentence *John was hit*,
s/he is bound to misinterpret the sentence as ill-formed or incomplete, because one must necessarily hit something. In addition, the visual input of the environment in which these sentences are uttered may force the learner to rethink his or her parsing. To sum up then, it is “nonsense” rather than comprehended input that may push a learner to re-evaluate and restructure interlanguage grammar, and this nonsense may be brought about through different routes, such as linguistic cues (*hit* subcategorizes for a theme), real world knowledge (books do not read), and the environment (the learner sees Mary giving a kiss to John). In this context, the benefits of simplified input advocated by Krashen are seriously questioned, since instructors who talk to their learners only in a simplified way may actually deprive them from the linguistic complexity that pushes learners to expand their interlanguages (White, 1989: 102).

In addition to these arguments, White claimed that certain structures may not possibly be acquired through mere exposure to correct L2 input, irrespective of how (in)comprehensible that is. This is especially relevant when learners have to retreat from non-target forms that they borrowed from their L1s. For example, French allows for adverbs to be placed between a verb and a direct object, while English does not (e.g., *John drank slowly his coffee*). In English, adverbs may appear in an array of positions, including at the beginning or end of a sentence, before a simple tense, or after a modal. Quite possibly, an L1 French learner that is exposed to all these variants will perceive that in English, as in French, adverb placement is relatively free, and no amount of L2 comprehensible input can possibly indicate the mismatch between English and French. As White puts it, there are cases in which “there are ‘gaps’ in the L2 input” (p. 107). Arguably, from a theoretical standpoint it can be posited that learners may be able to use
L2 input as “indirect negative evidence,” i.e., they may detect non-occurring forms and re-label these as ungrammatical. However, as White argues (p. 107) existing empirical evidence seems to suggest that this is actually not the case, at least for the majority of learners. For this reason, then, according to White, pedagogical interventions such as metalinguistic explanation and practice with corrective feedback may actually serve a greater purpose than that of improving the learner’s monitoring ability: they may actually lead to the acquisition of certain structures that no amount of exposure to comprehensible input can provide.

The data from the Canadian French immersion programs and the theorizing by White illustrate that exposure to comprehensible input alone may not suffice for SLA; yet, they do not provide empirical evidence for the role of practice. Such evidence is found however in the growing body of experimental studies framed under at least three strands of research: Processing Instruction, the Output Hypothesis, and the Interaction Hypothesis.

### 1.1.4 Processing Instruction

Processing Instruction (PI) is a type of FoF instruction or technique couched on VanPatten’s model of input processing (e.g., VanPatten, 1993, 1996, 2002; 2004b and elsewhere). According to this model, SLA is the result of a succession of learners’ internal processes that starts out with the conversion of input to intake, continues with the subsequent partial or total accommodation of intake into the developing system, and ends with the ulterior access to these data as output or production. In line with Krashen, VanPatten’s PI concedes maximum importance to the first process, input processing, or
the conversion of input to intake. However, both frameworks envisage different goals and means. Krashen’s goal is to facilitate intake to learners, and the best way to achieve this is by providing them with comprehensible input via simplification strategies. More ambitiously, though, PI’s goal is to help learners derive “richer” intake and make better form-meaning connections, and to achieve this it proposes two pedagogical interventions largely derided by Krashen: (1) preliminary exposure to metalinguistic information; and (2) subsequent practice. Each of these components is narrowly operationalized.

Metalinguistic information consists of (a) “traditional” explicit information about how the targeted linguistic form works, and (b) “psycho-linguistically motivated” information on the strategies that L2 learners may activate to process input on their own (“you would tend to interpret this as X”), and the strategies that they should be using instead in order to process input correctly (“don’t interpret this as X but Y”). In turn, the subsequent practice consists of two components: (a) input-based, interpretation activities using “structured input,” i.e., input that has been manipulated to increase the frequency and saliency of the key form that enables richer form-meaning connections; and (b) provision of feedback after learners’ responses.

Overall, a growing body of empirical research shows that PI can lead to L2 development of a wide range of grammatical structures on a variety of measures (e.g., Allen, 2000; Benati, 2001, 2005; Cadierno, 1995; Cheng, 2002; Collentine, 1998; Erlam, 2003; Farley, 2001a, 2001b; VanPatten & Cadierno, 1993a, 1993b; VanPatten & Sanz, 1995). However, given the bi-componential nature of PI, it may be argued that it is hard to determine whether the observed learning gains are due to the metalinguistic presentation, the practice, or a combination. This issue has been investigated by a
representative number of studies (Benati, 2004; Farley, 2004a; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996; Wong, 2004), all of which showed that practice with structured input may suffice to experience learning gains. There are, however, discrepancies as to the impact of additional metalinguistic presentation, with most studies showing no significant contribution and Farley (2004a) proving otherwise, likely as a result of the greater complexity of his target structure, the Spanish subjunctive (Farley, 2004a: 241-2). Regardless, research on PI clearly establishes a link between receptive practice and L2 development. Based on these results, VanPatten and many others have claimed that PI directly affects acquisition, understood as “the development of some underlying competence on which eventual (if not developmental) skills in language use depend” (VanPatten, 2004a: 39).

Arguably, then, PI constitutes a direct attack to Krashen’s claims about the negligible role of practice. Both theories, however, are more in synch as to the role of productive practice. In his early work, VanPatten (1993: 436) harshly criticized “traditional” drill-based productive practice because “it asks learners to produce when the developing system has not yet had the relevant intake data” or, as he phrased it, because it puts “the cart before the horse.” Recently, however, VanPatten (2004a; 2004b) has acknowledged a greater role of output in SLA. First, the negotiation of meaning that often arises in the course of productive practice can lessen task demands, thus freeing up attentional resources and allowing learners to process aspects of the input “they might miss otherwise” (2004b: 12). Second, the juxtaposition of the learner’s ill-formed output with another’s reactive input or feedback “may trigger noticing that is useful for making form-meaning connections” (2004b: 13). In a nutshell, then, VanPatten believes that interaction
can turn learners into “better processors of input” (2004a: 39), helping them to notice the “holes” and “gaps” in their interlanguage (cf. Doughty & Williams, 1998; Schmidt & Frota, 1986); however, in his view this only applies to vocabulary items and simple grammatical structures, with no studies thus far demonstrating “any impact from interaction on subtle and/or abstract properties of the grammar” (2004a: 40). Third and finally, output practice can also speed up the natural acquisition process by developing fluency through repeated accessing and productive procedures. Indeed, according to Lee and VanPatten (2003), input alone is not sufficient for developing the ability to use language in a communicative context, and PI is intended as but one component of a broader communicative approach that also includes structured output activities. In sum, then, VanPatten admits that productive practice may play a “facilitative” role in SLA but he opposes the view that it is “necessary” or that it may constitute a “direct path” to acquisition. Only input is necessary and directly affects the acquisition of an “underlying grammar” or “implicit system.”

1.1.5 The Output Hypothesis

A greater role of productive practice in SLA is envisaged by the supporters of the Output Hypothesis (e.g., Swain, 1984, 1985, 1988, 1993, 1995; 2005, and elsewhere). According to this hypothesis, output-focused practice can not only increase fluency and automatization of already acquired-knowledge but also bring about mental processes that both directly and indirectly improve accuracy and affect acquisition. This is achieved via three functions that may (but do not necessarily have to) operate whenever learners
produce the target language: (a) the noticing function, (b) the hypothesis-testing function, and (c) the metalinguistic function.

The noticing function refers to the fact that, in the process of producing speech, learners are confronted with what they want to say and what they can actually say, and as a result they may notice what they do not know at all (a “hole” in their interlanguage – Doughty & Williams, 1998) or what they know only partially (a “gap” in their interlanguage – Schmidt & Frota, 1986). A number of studies have illustrated that this function does actually occur, with productive practice triggering noticing of problems in vocabulary (e.g., Bialystok, 1990; Faerch & Kasper, 1983; Kellerman, 1991; Tarone, 1977) and grammar (e.g., Swain & Lapkin, 1995). However, as Swain and Lapkin (1995) noted, language proficiency seems to play an intervening role, with more advanced learners carrying out more grammatical analysis.

A second function of output is hypothesis testing. When learners produce an utterance, they reveal their hypotheses as to how the target language works. In turn, these utterances may trigger feedback that can cause learners to “reprocess” their hypotheses and produce modified output. This function of output is often referred to as the “comprehensible output” hypothesis, in clear contrast to Krashen’s theorizing, and refers to the fact that in addition to providing learners with comprehensible input, learners ought to be pushed to deliver messages that are not only conveyed, but that are conveyed “precisely, coherently, and appropriately” (Swain, 1985: 248). Existing literature has proven that this pushing does work because learners do modify their output in response to feedback moves such as confirmation checks or clarification requests (e.g., Iwashita, 1993; Pica, Holliday, Lewis, & Morgenthaler, 1989), although modification rates may be
low (e.g., one-third of the cases in Pica et al., 1989) and output may not always be successfully modified in the short term (Gass & Varonis, 1994).

Finally, output can also serve a *metalinguistic* function. According to Swain, certain tasks may elicit both production for communication purposes and conscious reflection about language form. That is the case of collaborative tasks where learners are asked to work in tandem to produce a written or oral text, including planning sessions for class presentations (Donato, 1994), passage reconstructions (LaPierre, 1994), and “dictogloss” tasks, where learners are first read a text and then work in pairs to produce a faithful reproduction (Kowal & Swain, 1997). Hitherto, as posited by Swain, a number of studies using verbal protocols have illustrated that in the process of completing these tasks learners do pay attention to grammatical form, engaging in negotiation episodes that may lead to successfully modified output (Donato, 1994; Kowal & Swain, 1997; LaPierre, 1994; Swain, 1998; Swain & Lapkin, 1998, 2002).

Arguably then, there is enough evidence proving that productive practice can promote the three functions hypothesized by Swain. However, whether productive practice in general or these functions specifically may contribute to second language acquisition remains an empirical question. Interestingly, the first studies in this direction were not published until the 1990s. As Shehadeh (2002) notes, in the decade that followed Swain’s formulation of the Output Hypothesis most studies were descriptive in nature, examining the different learner and contextual factors that may determine or promote the occurrence of output. Investigations on learner factors covered issues such as gender differences (Gass & Varonis, 1986; Pica, Holliday, Lewis, Berducci, & Newman, 1991; Pica et al., 1989; Shehadeh, 1994); ethnic background (Scarcella, 1983, 1992); proficiency level
(Varonis & Gass, 1985); status and expertise (Woken & Swales, 1989; Zuengler, 1989); and language background (Pica, Lincoln-Porter, Paninos, & Linnell, 1996; Shehadeh, 1999). In turn, research on contextual factors dealt with variables such as group size (Shehadeh, 2004); task type (Duff, 1986; Iwashita, 1993, 1999; Long, 1980; Pica, 1987; Pica & Doughty, 1985; Pica et al., 1996; Plough & Gass, 1993; Samuda & Rounds, 1993; Shehadeh, 1999); feedback type (Pica et al., 1989); and feedback source (Shehadeh, 1999, 2001).

The studies above illustrated how output may be promoted on the part of the learner, but they assumed rather than tested the facilitative role of output. Since the mid-1990s, though, an increasing number of studies have shown a link between productive practice and L2 learning (e.g., Izumi, 2002; Izumi & Bigelow, 2000; Izumi et al., 1999; Loewen, 2005; Mackey, 1999; McDonough, 2005; Nobuyoshi & Ellis, 1993; Silver, 2000). Based on this evidence, Swain (e.g., 2005) and other researchers contend that output is not only the outcome or “product” of the language acquisition device but also part of the “process” of learning that can affect acquisition. A number of authors, however, contend that while productive practice may be facilitative, it is not “necessary” for acquisition (e.g., Farley, 2001a: 295; Long, 1996: 448; VanPatten, 2004a: 43). For example, VanPatten (2004a) argued that productive practice can only affect acquisition indirectly, by sensitizing learners to their specific needs and leading them to process the subsequent input better. To illustrate this, VanPatten cites Izumi (2002), where groups reconstructing texts learned more about English relativization than those merely engaged in input comprehension activities. According to Izumi (p. 566, cited in VanPatten, 2004a), the greater gains experienced by the input+output groups would be due to the fact that “on
exposure to relevant input immediately after their production experience, the heightened
sense of problematicity would lead [learners] to pay closer attention to what was
identified to be a problematic area in their IL. In short, pushed output can induce the
learners to process the input effectively for their greater IL development.” A similar take
is shared by Farley (2001a; 2001b; 2004b). Addressing important critiques against the
processing instruction literature, Farley (2001a; 2001b) conducted two studies comparing
this technique against truly meaning-oriented output instruction. Contrary to previous
research, he found that productive practice led not only to comparable gains for
production but also for interpretation. To explicate this latter finding, Farley (2001a: 295)
argued that learners’ output served “as incidental input for other learners.”

It seems widely agreed then that productive practice can push learners to notice the
gaps and holes in their interlanguages, allowing them to selectively focus their attention
on subsequent input. The question that remains, however, is whether output can
independently contribute to SLA. According to Toth (2006: 327-8), several frameworks
provide support to this view. For example, sociocultural theory and skill acquisition
theory abandon the notion of an underlying linguistic system fed by input processing and
envision productive practice as a cognitive activity where doing and learning go hand in
hand. From a different angle, Ellis (2003: 114-115) contended that by monitoring their
production learners can bootstrap their way to syntactical rules that add to their linguistic
system independently of input processing procedures. Regardless the approach, the
purported independent contribution of output to SLA needs to be validated empirically.
Specifically, research must show that output can result in acquisition beyond the
provision and processing of additional input.
In this sense, an incipient number of empirical studies (e.g., Swain & Lapkin, 1995; Toth, 2006) contend that in the course of productive tasks learners do not necessarily have to resort to positive evidence in the ambient input to fill their gaps and holes. For example, Swain and Lapkin (1995), using verbal protocols, illustrated the different attempts of a learner at producing a noun he did not know yet based on a verb he just used (“détruire”) and his previous knowledge of French suffixation (“Et la déstruction. No, that’s not a word […] détruction, détruisision, détruisision, la détruisision des arbres”). As noted by Swain (2005: 474), the solution reached by the learner was incorrect, but this is precisely why this example is “revealing”: “[i]t allows us to conclude that new knowledge has been created through a search of the learner’s own existing knowledge, there being no other source. The learner’s search was triggered by his own output, which he noticed was incorrect.” This episode, however, clearly doesn’t result in new target-like knowledge, so one cannot establish a link between output and learning here, and, in line with VanPatten’s (2004a: 40) reasoning, it is restricted to vocabulary items.

In contrast, Toth (2006) provided an example of a learner-generated satisfactory solution for Spanish anticausative se, “a problematic clitic construction for English-speaking learners” (p. 328). Following the trend initiated by Farley (2001a; 2001b), Todd compared PI with a communicative output (CO) condition, obtaining comparable gains for receptive measures and a greater gain of the CO group for production. Beyond his quantitative analysis, of special interest here is his qualitative analysis of a particular interaction episode in the CO group (Toth, 2006: 358-9). In this episode, the students were asked to describe a picture showing the inside of a house after a hurricane, taking part in so-called “IRE” episodes where the teacher (a) initiated Interaction (¿Qué ocurrió
en el dibujo? [“What happened in the picture?”], (b) elicited a student’s Response (La ventana se abrió [“The window opened”]), and finally (c) delivered an Evaluation (Bien… En el dibujo B la ventana se abrió [“Good…. In drawing B the window opened”]). For some minutes, the teacher and the students engaged in a number of IREs with the same pattern, with all of the correct answers containing a middle-voice se (la ventana se abrió, la luz se cayó [“the window opened, the light fell”]). At one point, however, a student overgeneralized the structure, formulating the incorrect sentence La lluvia… se… mojó (“The rain got itself wet”). In response, the teacher delivered a complex feedback move including error repetition and metalinguistic information, without providing positive evidence (“The rain got wet? That is very metaphysical. That’s very, like, mysticism. How could the rain—? The rain didn’t get itself wet. The rain, itself.”) Next, the student stripped se from her sentence, but the sentence was still ungrammatical because it needed a direct object, so the teacher delivered an elicitation move (La lluvia mojó ¿qué? [“The rain got what wet?”]) and the student eventually reached a satisfactory solution (La lluvia mojó la alfombra [“The rain got the carpet wet”]). According to Toth, in this episode the solution reached by the student cannot possibly result from the processing of positive evidence in the ambient input, either prior to the episode (the structure had not occurred) or after the error (the instructor used elicitation and did not provide positive evidence). Instead, it was the “scaffolding” technique used by the instructor that helped the student use feedback with and without metalinguistic information to generate a target-like solution. Based on this qualitative analysis and the quantitative results of his study, Toth concluded that productive practice can affect acquisition beyond the processing of further input: it can elicit metalinguistic
information on the part of the teacher or activate existing learner-internal metalinguistic knowledge; it can draw increased attention to remembered input in addition to ongoing input, and it can restructure an extant system that contains incomplete or overgeneralized knowledge of a newly presented structure (p. 373).

It could be argued that, despite Toth’s claims on the difficulty of the target structure, the nature of the task and the teacher scaffolding reduced the cognitive operations imposed on the student to simple binary choices: adding or subtracting a *se*, adding or subtracting a direct object. Additionally, this information had been delivered either implicitly or explicitly in the preliminary metalinguistic explanation session, where models were also included. Therefore, despite PI’s research minimizing the role of prior explicit information (Toth, 361), it should be noted that students had already been explained the contributions of the middle-voice *se* and had been given examples of sentences with and without them. On this note, McDonough’s (2005) research is especially interesting, because it cancelled out the prior metalinguistic explanation variable and pushed learners to reformulate their errors in English question formation, a structure that requires complex syntactic operations far beyond binary choices. Interestingly, transcripts of her verbal protocols illustrated that, in response to negative evidence without positive evidence, some learners improved the accuracy of their original utterances in ways that exceeded the structural complexity of the interlocutor’s input. Nevertheless, McDonough (p. 96) stated that “the next challenge is to determine why” pushed output is beneficial, and in advancing the possible causes she did not make any explicit claims about independent contributions (p. 83): “[i]n sum, producing modified output may contribute to L2 development by strengthening knowledge representations
that learners already have stored (Nobuyoshi & Ellis, 1993) and by encouraging automatic retrieval of linguistic forms (de Bot, 1996).”

Overall, the studies by Swain and Lapkin (1995), Toth (2006) and McDonough (2005) are enlightening because they illustrate that productive practice can go beyond pushing learners to attune their attention to subsequent input. Learners can form, test, and revise hypotheses based on prospective and retrospective information, and this information may encompass not only positive evidence in the input but also metatalk and feedback. Clearly, further research must be conducted to determine how receptive and productive practice may contribute to SLA. To date, many studies have compared these two practice modes for a variety of structures, with mixed results. Arguably, though, arguments in favor of one or the other seem to be inextricably linked to one main property: their ability to raise learners’ awareness of the target structure, not only in quantitative terms but qualitatively, with deeper levels of analysis yielding better results (Izumi, 2002: 569). For this reason, frameworks such as Schmidt’s Noticing Hypothesis (Schmidt, 1990, 1993, 1995, 2001) and Craik and Lockhart’s (1972) Depth of Processing Hypothesis are often invoked. Furthermore, given that both receptive and productive practice have been shown to promote SLA, research must elucidate which particular subtypes are more beneficial and how they correlate to what is taught (the target form), where it is taught (the context), and to whom (the learner). No other framework has devoted more interest to studying these variables than the so-called Interaction Hypothesis.
1.1.6 The Interaction Hypothesis

Far from colliding with the Input and Output Hypotheses, the Interaction Hypothesis (e.g., Long, 1981, 1983b, 1996), recently rechristened as the Interaction Approach (Gass & Mackey, 2006), has essentially embraced the main tenants of these frameworks, subsuming them under one single paradigm. However, as noted by a number of authors (e.g., Mackey, 2007), over the three decades of its existence, the Interaction Hypothesis has undergone several transformations, constantly redefining itself and its areas of inquiry. Early work in the 1980s was geared around two pivotal ideas: Krashen’s Comprehensible Input Hypothesis and Hatch’s (1978: 63) proposal that “language learning evolves out of learning how to carry on conversations, out of learning how to communicate.” Bridging these two ideas together, Long’s (1981) initial formulation of the Interaction Hypothesis claimed that “participation in conversation with native speakers, made possible through modification of interaction, is the necessary and sufficient condition for SLA” (p. 275, emphasis added). Specifically, the following syllogism was proposed: (1) conversational interaction with native speakers can trigger adjustments in linguistic form, discourse structure, and/or message content that render L2 input more comprehensible to the learner; (2) as posed by Krashen, comprehended input leads to acquisition; and (3) based on the former premises, interaction can promote SLA. Focusing on the first premise, early empirical research in the 1980s and early 1990s sought to demonstrate a link between interaction and language comprehension, with positive results (e.g., Pica, Young, & Doughty, 1987). Additionally, following Long’s (1983a) remarks that instruction makes a difference in L2 acquisition when compared to naturalistic exposure, researchers investigated different task and learner features,
uncovering the learning conditions that best promoted the interactional adjustments advocated by the Interaction Hypothesis. For example, it was found that two-way tasks, where each partner holds relevant but incomplete information, prompt more negotiation work than one-way tasks (Doughty & Pica, 1986; Long, 1989; Pica & Doughty, 1985); the same applies to tasks where the input is split rather than shared (Newton, 1991); where the information exchange is obligatory rather than optional (Doughty & Pica, 1986); where the outcome is closed, with only one possible solution (Crookes & Rulon, 1985; Long, 1989); where learners collaborate toward convergent rather than divergent goals (Duff, 1986; Pica et al., 1993); where learners collaborate with each other as opposed to with a teacher or native speaker (Rulon & Mc Creary, 1986; Varonis & Gass, 1985); and where less proficient learners assume the role of information providers in dyadic interaction with more proficient learners (Yule & Macdonald, 1990).

This early body of studies illustrated that interaction could indeed facilitate comprehension, as well as the different ways in which it could be promoted. However, whether interaction could lead to acquisition remained an empirical question. To that aim, another batch of studies in the late 1980s and all throughout the 1990s investigated whether interaction could lead to L2 development, per se and in contrast to non-interactionist instructional approaches, showing learning gains for both vocabulary and grammar (e.g., Ellis & He, 1999; Ellis et al., 1994; Gass & Varonis, 1994; Loschky, 1994; Mackey, 1999; Pica, 1992; Pica et al., 1987). Even more interestingly, though, research illustrated that contrary to Krashen’s claims, comprehension gains did not always correlate to developmental gains, and that it is processed more than merely comprehended input that seems to be crucial for acquisition. For that reason, in his
reformulation of the Interaction Hypothesis, Long (1996) emphasized the role of interaction in promoting noticing, which as argued by a number of frameworks (e.g., Noticing Hypothesis, Consciousness-raising - Schmidt, 1990, 1993, 1995, 2001; Sharwood-Smith, 1981, 1993) helps learners process input for intake. Additionally, drawing on the positive results of empirical literature examining the role of feedback and productive practice (cf. Swain’s Output Hypothesis), Long was more explicit about the backbones of interaction: not only positive evidence in the input, but also negative evidence about what does not work in a language, and the learner’s active involvement in producing and/or reformulating speech, can promote L2 development. Thus, Long’s (1996: 451-2) currently ubiquitous quote reads: “negotiation for meaning, and especially negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways.”

Long’s reformulation of the Interaction Hypothesis encapsulated the spirit of a new direction in interaction research. First, interaction was no longer seen as the necessary and sufficient condition for SLA (Long, 1981) or as the way through which “language is best learned and taught” (Pica et al., 1993) but as one route that is highly facilitative. Second, having established links between interaction and L2 development, the focus shifted to investigating which particular types of interaction are more beneficial (Doughty, 1991; Long, 1991a). Previous research showed that interaction by itself is not always the answer, that it can promote comprehension but not necessarily acquisition. *Incidental practice* is far from ideal. As posed by Long (1996: 448): “Free conversation is notoriously poor as a context for driving IL development for a number of reasons,
because the lack of any fixed topics or outcomes permits rapid, superficial treatment of topics and the dropping of any that cause linguistic trouble.” Instead, what is sought is interaction that best promotes the tenants of the hypothesis: comprehended and noticed input, output, and modifications thereof in response to feedback. Thus, researchers started to compare the developmental effects of different forms of deliberate practice which preserved meaningful interaction while drawing attention to form (e.g., Focus on Form and Functional-Analytic instruction - E. Day & Shapson, 1991; Harley, 1989; Lightbown & Spada, 1990; Lyster, 1994; J. Williams & Evans, 1998). The field of Task-Based Language Teaching (TBLT) boomed, spawning numerous investigations within the interaction approach but also other frameworks. Researchers compared the effects of receptive vs. productive practice, different variants of each practice mode against each other, different types of feedback, and different types of tasks, bringing in new issues such as the timing and content of each of the phases of practice, e.g. task-planning and task-repetition (Bygate et al., 2001; Skehan, 2003; Willis, 1996), as well as the cognitive demands imposed on the learner, e.g., task-complexity and task-sequencing (Robinson, 2001, 2003, 2005).

Finally, in addition to considering extrinsic factors, i.e., those which can be manipulated by the instructor, interaction research in the 1990s and 2000s has devoted increased attention to intrinsic factors, gauging the weight of the targeted linguistic form (e.g., Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; N. Ellis, 1993; Robinson, 1996) and individual differences such as age, working memory, aptitude, language experience, motivation, and the like (Robinson, 2002). Researchers have thus come to realize that there is no “magic recipe” to L2 development through interaction, but rather that there is
a very assorted menu, and that the ideal choice depends on a combination of factors that are inextricably linked. As Mackey (2007: 24) puts it:

To date, there is no claim that one or more of the interactional processes ‘work better’ than any of the others, although it does seem likely that processes such as negotiation for meaning, feedback, and modified output may be differentially effective for various aspects of language, learner characteristics, contexts, and task demands. Indeed, the vast majority of research in this area suggests that these developmentally helpful interactional processes may work in concert or in unique ways.

1.2 Exposure to practice

In a pioneering study, Norris and Ortega (2000) synthesized and meta-analyzed the effects of the main variables explored in instructed SLA research published between 1980 and 1998, addressing research questions such as (1) the role of L2 instruction overall and relative to simple exposure or meaning-driven communication; (2) the role of explicitness in form-focused instruction; (3) the role of the assessment measures used; (4) the incidence of length of instruction; and (5) the duration of the observed developmental effects.

Of special interest here is the first of their research questions. Based upon a sample of 49 studies, Norris and Ortega found that form-focused instruction significantly yielded higher developmental gains in L2 grammar than meaning-oriented instruction or exposure to input containing the targeted form. More interesting even is the coding protocol that they used to classify the experimental groups in the literature. Specifically, Norris and Ortega (2000: 446) identified three main types of conditions. “Experimental groups” are cells of students that are exposed to the targeted form via form- or meaning-focused instruction. “Instructional comparison groups” are deprived from practice but still receive non-focused exposure to L2 input containing the structures being taught in
the experimental condition. Finally, “true control groups” are those which receive neither instruction nor exposure to the target form except in pretests and posttests. As this division illustrates, “exposure” is here understood in a Krashenian sense, where “instructional comparison groups” have access to the targeted forms by reading written texts or listening to aural texts, with no deliberate practice on their part. Norris and Ortega’s operational definition of the term “exposure” thus clearly illustrates a lurk in the SLA empirical research since 1980: the vast majority of studies in PI, the Output Hypothesis, and the Interaction Hypothesis have compared the effects of different types of practice against each other or versus exposure to L2 input alone; however, very few studies have compared the effects of performing a task vs. observing other learners performing it. Filling this research gap, however, is sound for theoretical and pedagogical reasons, both in the classroom setting and in CALI.

From a theoretical perspective, the proponents of the Interaction Hypothesis contend that, in contrast to teacher-centered instruction settings — where communication is unidirectional and involves learners’ exposure to the input delivered by the teacher — participation in interactive practice can promote SLA because it triggers developmentally helpful processes such as input processing, output production and modifications of input and output in response to feedback. However, no specific claims are made as to the role of the agency of the learner in the interaction process. In other words, it remains unclear whether the effects of interaction and its quintessential components, input, output, and feedback, are mediated by the role of the learner, who can act as an “interactant” or an “observer.”
Additionally, from a pedagogical perspective, empirical research on the effects of exposure to practice is required to better understand the very nature of classroom instruction and CALI. In classroom contexts, especially since the advent of the Interaction and Output hypotheses, teachers have sought to motivate students to assume an active role in interaction episodes. However, as Mackey (1999: 559) puts it, “[i]n many second language classrooms as well as naturalistic contexts […] learners often observe the output of others without producing their own output. Is it helpful for learners to observe output [and the input and feedback that output may trigger] without actually taking part in it?” Almost every class has a number of students that seem to feel uncomfortable when asked to assume an active role; yet instructors and/or language program coordinators almost always encourage participation, oftentimes even penalizing low rates. If both active practice and exposure to practice are proven equally effective, then teachers and language program directors may want to rethink their policies.

Similarly, in the realm of CALI it has been posited that in contrast to other technologies based on textbooks or audiovisual media, the benefits of language learning software arise from its ability to engage learners in interactive practice (Nagata, 1996). However, CALI is intrinsically expensive. Developing computerized tutors requires computationally complex tools that are able to analyze the student’s response and react accordingly. In contrast, filming interactions is arguably less expensive, but so far even the highest-budget non-interactive audiovisual series (e.g., Sol y Viento) only provide positive evidence and occasionally metalinguistic information, and very rarely, if ever at all, do they display episodes of feedback in response to L2 errors or output modification. Therefore, research must address whether the benefits of CALI are due to its capability to
actively engage learners in interactive practice or due to a poor use of less costly, non-
interactive media (see, e.g., Stockwell, 2007; C. Williams & Brown, 1991).

Both in oral interaction and CALI research on the effects of exposure to practice on
L2 grammar development is notably scant. The next sections provide a review of related
literature. Section 1.2.1 focuses on studies analyzing the role of exposure to practice in
oral interaction, while section 1.2.2 focuses specifically on CALI.

1.2.1 Exposure to practice in oral interaction

In SLA, the effects of exposure to practice on L2 grammar development have been
observed from two different perspectives, i.e., naturalistically, in classroom
environments, and experimentally, in laboratory settings. On the one hand, a batch of
studies mostly published in the late 1970s and the 1980s recorded classroom interactions
and observed how the amount of voluntary oral participation may impact learning. On the
other hand, an incipient number of studies since the late 1990s experimentally isolated
two conditions in laboratory settings to compare the effects of practice versus exposure to
practice. The next two sections deal with these blocks of studies separately.

**Classroom contexts**

Since the late 1970s, a number of observational and empirical studies have investigated
how voluntarily assuming a tacit or an active role in the classroom correlates to learning
rates. Some studies have peripherally investigated this for vocabulary learning (Ellis et
al., 1994), finding no significant differences, while other studies have done so for L2
grammar development. These latter studies have targeted a wide variety of populations,
ranging from infants (Ellis, 1984) to adult students (R. Day, 1984; Ely, 1986; Seliger, 1977). The following pages exclusively summarize the studies dealing with adult SLA and some type of measure of grammar development.

In a pioneering study, Seliger (1977) investigated the correlation of what he termed “practice” with grammar development and aural comprehension. His definition of practice was “any verbal interaction between the learner and others in his environment.” In other words, he measured voluntary oral participation in and outside the classroom without a focus on a particular structure. Based on class observations, he distinguished between more versus less practicing students (“high vs. low input generators”). The former “cause input to be directed at them, by calling out, answering out of turn, and working out answers to questions or drill cues directed to others,” while the latter “sit quietly but rarely and in some extreme cases never participate unless specifically asked to do so.” To operationalize this, two observers counted the number of output production turns (self-initiated and in response to teacher solicits) of 12 ESL students during four hours of classes, and selected the top three and the bottom three. All students were pre-tested for aural comprehension and written structure. Results of achievement scores at the end of a 15 week semester showed that while initial placement test scores could not predict final achievement, active participation correlated positively with both aural comprehension and written grammar (measured by Lado-Fries test); however, active participation did not correlate with an additional measure of written proficiency (discrete item cloze test). Seliger’s findings must be interpreted with caution due to a number of limitations. First, the sample size was of only 3 students per cell. Second, his coding procedure for student participation might involve some circularity, since by including
responses to teacher personal solicits he ignored Rubin’s (1975, cited on Day, 1984: 75) point that a teacher might tend to call more often on the better student, so one may wonder what comes first, being a better student or being a student who participates often. Third, Seliger’s cloze test was seemingly a vague measure of written proficiency, since asking students to provide the exact word for every fifth deleted token in a text randomly addresses accuracy in grammar, vocabulary, and/or pragmatics.

In a partial replication of Seliger (1977), Day (1984) addressed some important limitations. From a terminological point of view, he relabeled Seliger’s “practice” notion as “the phenomenon which we call classroom participation” (p. 79). From a methodological perspective, he expanded the sample size from 6 to 26; additionally, he improved Seliger’s (1977) allegedly circular coding system by disregarding responses to personal solicits, over which the student has little control, and counted only self-initiated turns and responses to general solicits, in which the teacher asks a question to the entire class. Similarly to Seliger, Day found no correlation between in-class participation and written performance on a cloze test, which may not be particularly striking given the randomness of this assessment measure. However, while Seliger found a positive correlation between participation and both aural comprehension and written grammar production, Day found no such correlation for his dependent variable, oral communicative competence in a structured interview. Three factors may contribute to explicate these contrasting findings. First, the subjects were quantitatively and qualitatively different (Day had a larger sample size, which was also more heterogeneous in terms of level and L1 background). Second, the coding procedures were different (Seliger counted total speech acts, while Day only considered voluntary ones). And third,
the assessment tests measured very different skills (by measuring oral competence only, Day could not assess whether any progress in written proficiency occurred).

Building on previous research, Ely (1986) investigated the effects of voluntary participation on L2 development in a sample of 75 college students of first year Spanish. Echoing Day (1984), he operationalized class participation by counting only voluntary speech acts (self-initiated turns and responses to teacher general solicits). Proficiency was measured by three dependent variables: fluency and accuracy on a story retelling task, and accuracy on a final written examination. Interestingly, Ely’s findings provide counterevidence to both Seliger (1977) and Day (1984). Contrary to Seliger, Ely did not find any effects of participation on a written grammar test, which may be due to either a lack of skill transfer (the treatment was aural) or, as argued by Ely (p. 21), the fact that non-participating students may have equally benefited from other learning strategies such as silent study or exposure. Additionally, Ely’s results also contradicted Day’s, since he did get a significant positive correlation, as opposed to none, between participation and oral accuracy for a subset of his participants.

To sum up, then, SLA research investigating the effects of high versus low in-class participation on L2 grammar development is scant and contradictory. Seliger (1977) found positive effects of active participation on a written grammar test, whereas Ely (1986) did not. In turn, Day (1984) did not obtain any correlation between active participation and oral accuracy, whereas Ely (1986) did. Additionally, the lack of focus on a particular grammatical structure poses some concerns on the internal validity of these studies. Finally, from a theoretical perspective, this research offers only indirect evidence about the effects of practice vs. exposure to practice. There are many variables
that can determine the rate of active participation of a student: motivation, personality, language experience, etc. As a result, this early research may show us whether there is a correlation between voluntary in-class participation (and hence more output practice) and L2 development, but it cannot show us whether students, regardless of their personal drive to participate, may differently benefit from being pushed to practice or observe others practicing.

**Laboratory contexts**

Hitherto very few oral interaction studies have experimentally isolated two conditions to observe the effects of practice versus exposure to practice, and only a subset has done so for L2 grammar development. Probably the first study to isolate these conditions was Pica (1992) where students who completed an information-gap task with the teacher achieved statistically similar comprehension rates than students who merely witnessed those interactions. On the other hand, only two studies to date have used L2 grammar development as a dependent measure, Mackey (1999) and Muranoi (2000).

Targeting English question formation, Mackey (1999) was the first study to observe the effects of learner’s agency in L2 grammar development. Specifically, 14 students at different developmental stages (7 readies and 7 unreadies) were asked to complete a series of tasks (picture drawing, story completion, story sequencing) with a native speaker that provided interactionally modified input. In turn, 7 other students observed some of the interactions. These observers were deprived from L2 output practice, but in order to ensure that they paid attention to the interactors, they were asked to complete the experimental tasks by supplying the necessary information in their L1 or drawing the
pictures that were described. Finally, a control group received premodified input from the researcher, but had no opportunity to negotiate for meaning. Learning outcomes were based on the same tasks used for the treatment. Results showed that although observers successfully completed the tasks in 98% of the cases, they did not experience a significant learning gain, similarly to the control group, whereas interactors did. Despite the groundbreaking nature of this study, results must be interpreted with caution due to the relatively small sample size of the exposure group and the fact that the pool of participants was hybrid, including both developmentally “ready” and “unready” students.

Building upon Mackey (1999), Muranoi (2000) used a much larger sample size to address the effects of type of agency (or “type of participation”) in the development of the English article system. In his experiment (N = 91), 30 Japanese students completed problem-solving tasks in dyadic oral interaction with the teacher, while other 61 students observed those interactions. Unlike Mackey’s study, these observers were not required to perform any meaning-focused task, although to ensure attention they were instructed to act as consultants when their performing peers needed assistance. In addition to type of learner’s participation, Muranoi included another variable, type of instruction. Specifically, Muranoi compared two experimental groups, IEF and IEM, against each other and a contrast group, NIE. In the experimental groups (IEF and IEM), the participants who performed the tasks received form-focused instruction based on the Interaction Enhancement (IE) technique, thus receiving feedback in the form of recasts and prompts after errors in their use of English articles. Additionally, after completing the task, the IEF group received a “formal debriefing,” consisting of a grammatical explanation on the targeted form, while the IEM group received a “meaning-focused
debriefing” in which the instructor made comments on the students’ performance in terms of accuracy in communicating messages, rather than accuracy on the target forms. In turn, the contrast group, NIE (Non-enhanced Interaction), completed the same interactive tasks without receiving recasts or prompts, and then received a meaning-focused debriefing. For all three groups, the treatment evolved along three sessions of 30 minutes. Interestingly, Muranoi found that type of student participation did not play a role in L2 development, since students in both groups (performers and observers) experienced comparable learning gains across all assessment tasks, including two types of oral production tests (story description, picture description), a written production test (picture description), and a grammaticality judgment test. Moreover, the obtained non-significant interaction for Instruction x Participation-Type indicated that learner’s agency in interaction played no role for any of the treatments, the more explicit one (IEF), the less explicit one (IEM), and the contrast group (NIE). Therefore, Muranoi’s results provide counterevidence to Mackey (1999), who found an advantage for practice over exposure to practice. Muranoi’s study, however, is not without limitations. The author himself acknowledged several of them, such as the composite nature of the treatments (the effects of practice, feedback, and post-task presentations on meaning or form were not isolated); possible researcher expectancy effects (he himself administered all the treatments); and lack of participant randomization (intact classes were used). Additionally, other caveats may have moderated the results. First, it is unclear whether or not the two agency levels were properly isolated. Muranoi states that he took part in dyadic interaction with three or four students in each of the three treatment sessions, but he does not report what these students and their observers did after finishing their tasks. If the performers remained in
class and watched other interactions, then the line between the two agency levels may have been blurred. And second, the obtained non-significant interaction Instruction x Participation-Type may have resulted from the fact that the ANOVA was run on all experimental groups holistically, including the control condition, where the performers experienced minimal gains. Consequently, any potential differences in the two experimental conditions, IEF and IEM, may have been obscured.

1.2.2 Exposure to practice in CALI

While there are several CALI studies analyzing the role of type of practice in terms of mode (i.e., receptive vs. productive - Morgan-Short & Wood Bowden, 2006; Nagata, 1998a, 1998b) and task features (task-complexity - Medina, 2008; task-essentialness - Moreno, 2007), only two CALI studies to date have empirically addressed the effects of practice versus exposure to practice, one in L2 vocabulary learning (deHaan, Reed, & Kuwada, 2010) and another in L2 grammar development (Hsieh, 2007).

In a very recently published study, deHaan, Reed, & Kuwada (2010) divided 80 Japanese undergraduate students of L2 English into two conditions. In one condition students played an English-language music video-game for 20 minutes, whereas in the other students watched the game simultaneously through a monitor. Results showed that although both players and watchers recalled vocabulary from the game, the players recalled significantly fewer words than the watchers. According to deHaan et al., the poorer performance of the players could be the result of the extraneous cognitive load induced by having to play the game.
Switching from vocabulary to grammar, Hsieh’s (2007) doctoral dissertation investigated the effects of input-based practice on the development of Spanish *gustar*. A total of 52 participants of first semester Spanish were divided into 6 experimental groups, as per the interaction of two variables, i.e., practice (practice vs. exposure to practice) and type of feedback (more explicit, less explicit, no feedback). The experimental task consisted of a game where “practicing” participants were prompted to navigate a series of forking paths with word options in order to sequentially translate a picture cue into its equivalent *gustar* structure. Participants in the “exposure” group were instructed to watch, in real time, the completion of this task by a peer. The results of achievement tests were mixed. While both conditions achieved statistically significant learning gains in written recognition and production of the form, the practice group outperformed the observers in oral production. However, Hsieh’s findings must be interpreted with caution given important methodological decisions and caveats. First and foremost, Hsieh compared practice vs. exposure to practice for all three feedback conditions together (more explicit, less explicit, no feedback), which does not allow to determine whether the results applied to some or all conditions and may have masked additional differences. Second, the study had low power and small sample sizes (as low as 6 participants in some cells). Third, timing was managed inconsistently (students completed the immediate posttest with a 1-5 day difference due to technical problems). Fourth, prior L3 experience may have played a role (40 out of the 52 participants had over three years of prior exposure to either Latin, French, or Italian, languages with parallel *gustar* constructions). Fifth and last, all the participants in the exposure condition were instructed to think aloud, which may have washed out the differences between the practice and exposure groups.
Arguably, one could posit that participants deprived from practice may subvocally complete the tasks they observe anyway; however, the imposed requirement of thinking aloud may further encourage them to do so, which is problematic for pedagogical and methodological reasons. From a pedagogical standpoint, a thinking aloud prerequisite is not ecological, as instructors hardly ask their students to think aloud while observing the interactions of their peers. From a methodological perspective, concurrent verbal protocols may induce reactivity, although the incipient debate on this area of inquiry is open (see the recently published book by Bowles, 2010). Clearly, then, the singularity of Hsieh’s study and its limitations warrant further research in the CALI arena.

**Section summary: Learner’s agency**

Since the formulation of Krashen’s Comprehensible Input Hypothesis to the present date, researchers have envisaged the role of practice in SLA very differently. According to Krashen, practice, as well as other pedagogical interventions such as explicit grammar information and feedback may at best lead to conscious learning and monitoring, but neither of these is relevant to acquisition (Krashen, 1984). Furthermore, practice of mechanical nature is deemed to obstruct the learner’s focus on meaning, thereby not only *not facilitating* but *impeding* acquisition (e.g., Krashen, 1982; Prabhu, 1987).

Consequently, then, in Krashen’s view the role of the teacher should be restricted to providing comprehensible input to students, which can be achieved via simplification strategies.

In reaction to Krashen, some authors claimed that while necessary, comprehensible input alone may not suffice for SLA, as evidenced by the evaluations of the language
immersion programs that started out in Canada in the 1960s. These evaluations showed that students failed to achieve native-like competence in production measures (e.g., Harley & Swain, 1978, 1984; Lapkin et al., 1991) despite being exposed to great amounts of L2 English input over long periods of time. From a theoretical standpoint, White (1989: 102) argued that it is incomprehensible rather than comprehensible input that pushes learners to expand their interlanguages, and that some L2 structures that involve retreating from L1 transfers cannot possibly be learned via the ambient input alone. Based on these two arguments, White questioned the benefits of simplified input posited by Krashen and contended that pedagogical interventions such as metalinguistic explanation and practice with corrective feedback may not only improve the learner’s monitoring ability but also facilitate acquisition.

From an empirical perspective, validation for the benefits of practice in SLA can be found in at least three strands of research. The studies in Processing Instruction illustrate that input-based practice with structured input can lead to acquisition because it helps learners in making better form-meaning connections and deriving richer intake. On the other hand, the studies framed under the Output Hypothesis show that productive practice can promote L2 development, possibly because it helps learners to notice the gaps and holes in their interlanguages, it allows them to generate, test, and revise hypotheses based on subsequent and stored input, and it can trigger feedback and conscious reflection about language form. Finally, research premised on the Interaction Hypothesis shows that interactive practice promotes SLA significantly better than mere exposure to comprehensible input because it triggers developmentally helpful processes involving input, output, and output modifications in response to feedback. More interestingly, this
research has illustrated that it is processed rather merely comprehended input that seems to be crucial for acquisition, and therefore interactive practice per se is not the answer: incidental practice such as casual conversation is a poor context for acquisition, so L2 instruction should be geared around form-focused approaches that allow for a balanced focus on meaning and form.

While many SLA studies have investigated the construct of practice both in terms of its intrinsic and extrinsic factors, very few studies have observed the differential effects of learner’s agency in grammar practice, i.e., being actively engaged in practice versus observing others practicing. Based on classroom observations, a number of early studies in the 1970s and 1980s statistically compared the correlation between different degrees of active participation and L2 grammar development, yielding mixed findings. Seliger (1977) found positive effects of active participation on a written grammar test, whereas Ely (1986) did not. This latter study did however obtain a positive correlation between active participation and oral accuracy, but this finding was counterevidenced by Day, who found no correlation (1984). Besides their mixed findings, these studies must be interpreted with caution for two main reasons: first, they operationalized L2 grammar development holistically, without focusing on a specific linguistic form, and second, they cannot inform whether students, regardless of their personal drive to participate, may differently benefit from being pushed to practice or observe others practicing. These two caveats have been superseded by a reduced number of laboratory-based studies since the late 1990s, although their findings are again mixed. In oral interaction, Mackey (1999) found that students negotiating for meaning in a number of tasks experienced developmental gains in English question formation, whereas students who witnessed
those interactions did not. In contrast, using a much larger sample size Muranoi (2000) found that both conditions learned the English article system statistically similarly. Finally, in the field of CALI the only relevant study to date, Hsieh (2007), obtained mixed findings for Spanish *gustar* constructions, with both conditions experiencing similar gains in written recognition and production, and the practice condition outperforming the observers in oral production.

Beyond grammar acquisition, the first findings in other areas of L2 development such as comprehension and vocabulary learning further illustrate that the role of learner’s agency in practice is not as straightforward as one might think. Using L2 comprehension as her dependent measure, Pica (1992) found that students who interacted with their tutor in an information-gap task performed similarly to their observers. In turn, deHaan et al. (2010) found that players in a music videogame recalled significantly fewer words than students who watched them play.

Clearly, then, the paucity of studies and the contradictory findings with regards to the role of learner’s agency in deliberate practice warrant further investigation to address a fundamental gap in the formulation of the Interaction Hypothesis.

2. Feedback

This section is divided into two parts. Section 2.1 deals with feedback in SLA in general, whereas Section 2.2 focuses on computerized feedback exclusively.
2.1 Feedback in SLA

Since Krashen’s times to this day and age, pedagogical interventions such as the provision of manipulated input, metalinguistic instruction, and feedback have gradually been the focus of extensive research among second language acquisitionists, precisely because they constitute one of the main distinctions between first and second language acquisition (Sanz, 2005: xi).

In the field of first language acquisition many researchers have followed the Chomskyan paradigm, contending that what makes acquisition possible is Universal Grammar, defined as the innately human “system of principles, conditions, and rules that are elements of properties of all human languages” (Chomsky, 1975: 29). According to these researchers, it is exposure to positive evidence that activates the sets of parameters that work for a particular language. Conversely, exposure to negative evidence has a negligible role, since it can at best temporarily change language behavior but cannot lead to changes in interlanguage grammar (Carroll, 1996; Cook, 1991; Schwartz, 1993).

In the realm of second language acquisition, early work by Krashen (e.g., 1982 and elsewhere) and the followers of his Input Hypothesis was also in line with Chomskyan nativists and posited that only implicit processing of comprehensible input can lead to L2 acquisition. For these researchers, any form of explicit linguistic information, be it via pre-emptive metalinguistic instruction or corrective feedback, can only lead to learned knowledge, and learning cannot be converted into acquisition. In contrast, today, after three decades of empirical research there is enough evidence attesting that corrective feedback leads to significantly more learning than positive evidence alone, both in classroom and laboratory contexts, and both with children and adults (Li, 2010; Lyster &
Saito, 2010; Russell & Spada, 2006). This however does not answer the question of whether feedback is necessary for L2 acquisition or is just a booster of the learning process. On this token, it has been posited that the role of feedback may vary for L1 and L2 acquisition. Thus, while many researchers believe that corrective feedback has a negligible role in L1 acquisition, there is increasing support to the thesis that adults cannot acquire native-like competence in an L2 merely based on positive evidence. Specifically, researchers claim that there are maturational constraints on language learning (Hyltenstam, 1988; Newport, 1990, cited in Long & Robinson, 1998), and that transfers from the L1 may negatively affect acquisition when the L1 stands in a superset relation with regards to the L2, making it impossible for learners to retreat from overgeneralizations without the help of corrective feedback or explicit instruction (White, 1989; 1991).

After a series of descriptive studies, research in the late 1990s and the 2000s clearly established a link between feedback and L2 development. Interestingly, this early batch of studies was predominantly focused on implicit types of feedback, particularly recasts, for two main reasons. First, recasts were identified as the most frequently occurring feedback type in and out of the classroom environment (e.g., Lyster & Ranta, 1997; Morris, 2002; Oliver, 2000), and second, they were advocated as optimal pedagogical interventions because they are said to be time-saving, to help maintain a focus on meaning to a greater extent than other techniques, and to be less threatening to student confidence (e.g., Han & Kim, 2008; Loewen & Philp, 2006; Long, 2007; Nicholas et al., 2001). To date, recasts have been extensively proved to facilitate acquisition (see, e.g., Long, 2007 for a review). However, empirical studies differ in terms of the amount of
immediate repair that they generate and the learning gains that they lead to, both in the short and long term, and both independently from and viz à viz more explicit types of feedback. As Nassaji (2007: 520-1) puts it, a number of mediating variables may contribute to explicate these divergences:

For example, in some studies, reformulations that seek confirmation of the learner’s message have been coded as recasts (Lyster, 1998b). In others, similar interactional moves have been coded as negotiations or nonrecasts (Braidi, 2002; Oliver, 1995, 2000). In addition, whereas, in some research, recasts have involved random feedback on any error (e.g., Lyster & Ranta, 1997; Panova & Lyster, 2002), in others, recasts have been provided intensively and systematically on specific errors (e.g., Doughty & Varela, 1998; Han, 2002; Mackey & Philp, 1998). Also, whereas, in some studies, recasts have been used with no additional negotiation signals (e.g., Braidi, 2002; Long et al., 1998), in others, they have been used in conjunction with other interactional features, moves, or signals. For example, in Doughty and Varela’s study, recasts were preceded by repetition of the utterance with rising intonation as well as added stress; in Muranoi’s study (2000), which investigated focus on form in L2 classrooms, they were used in combination with clarification requests; and in Philp’s study (2003), they were associated with additional knocking sounds to cue learners to recall the recast.

Addressing these caveats, recent studies (e.g., Loewen & Philp, 2006; Nassaji, 2007; Sheen, 2006) have started to isolate the factors that contribute to the diversity of recasts in the literature (see Chapter 3), thereby distinguishing between “implicit and explicit recasts” (Sheen, 2006) and interestingly concluding that it is actually the most explicit types (short, declarative, reduced, repeated, with a single error focus) that yield the highest rates of uptake and learning.

After the initial boom of research on recasts, authors set out to compare the relative efficacy of different types of feedback. Early work largely compared feedback types in terms of their “explicitness,” understood as either the overt/covert correction of an error or the provision/absence of metalinguistic rules (see Chapter 3). Studies grew at an exponential rate and so far three meta-analyses have used the explicitness criterion to synthesize their findings (Li, 2010; Mackey & Goo, 2007; Russell & Spada, 2006).
Overall, these meta-analyses corroborate the facilitative role of feedback in SLA; however, they are not unanimous about which types of feedback yield the highest effect sizes. Russell and Spada (2006: 154), concluded that “[m]ore research to directly compare varying degrees of explicit and implicit types of CF on L2 learning is needed” (p. 154). Similarly, Mackey and Goo (2007: 440) stated that “any arguments for the efficacy of one kind of feedback over another [are] premature.” In contrast, in the most comprehensive meta-analysis to date, Li (2010) concluded that “explicit feedback worked better than implicit feedback on immediate and short-delayed posttests [<7 days and 7-29 days, respectively], but on long-delayed posttests [>30 days], implicit feedback was slightly more effective” (p. 35-6). Recently, however, authors such as Ranta and Lyster (2007) have proposed that feedback types should not be classified according to the explicitness criterion, but rather, based on whether or not they encourage learners to repair their errors. Thus, in a recent meta-analysis Lyster and Saito (2010) specifically isolated prompts from recasts and explicit correction and concluded that “the effects of prompts are larger than those of recasts in classroom settings” (p. 294), while “explicit correction did not prove significantly different from either” (p. 282-3). To summarize these four meta-analyses, then, Russell and Spada (2006) and Mackey and Goo (2007: 440) concluded that more studies are needed to decide whether implicit and explicit types of feedback are differently beneficial; Li (2010) concluded that explicit feedback outperforms implicit types in the short- to medium run; and Lyster and Saito (2010) concluded that recasts and explicit correction are not differently beneficial, and that prompts lead the way. Clearly, the debate is as heated as can be.
Arguably, there are several factors that may mediate the effects of feedback and thus contribute to explicate these conflicting findings. For example, as noted by a number of authors, studies have often failed to isolate feedback from other variables such as prior metalinguistic instruction, the type of exposure to the L2 form—e.g., practice vs. exposure to positive evidence—or the practice mode—i.e., input- or output-focused (see Sanz & Morgan-Short, 2005 for more details); also, authors have overlooked the role of the task-features and the learning context—i.e., classroom vs. laboratory settings (more in Gass et al., 2005), the learner’s individual differences—e.g., memory, aptitude, motivation, and developmental level (as noted by e.g., Ammar, 2008; Ammar & Spada, 2006; Mackey & Philp, 1998), and the type of target form (for seminal studies see, e.g., Ellis, 2007; Iwashita, 2003; Long et al., 1998; Nagata, 1993).

More importantly, however, the conflicting findings in the feedback literature may be explicated not only by extrinsic but intrinsic factors. Researchers have long conceived feedback as a one-dimensional unit that may provide more or less explicit negative evidence, neglecting the fact that it can also embed positive evidence (Leeman, 2000) and prompt learners to repair their errors (Ranta & Lyster, 2007). Underscoring this latter component, a recent batch of studies has started to compare the relative effects of reformulations vs. prompts, showing superior results for the latter (see the meta-analysis by Lyster & Saito, 2010). However, as posed by Ammar and Spada (2006: 565), it could be argued that recasts and prompts are two totally different entities and hence they “should not be compared at all”: recasts provide positive evidence and typically preclude or hinder error repair, whereas prompts do not provide positive evidence and encourage repair. From an ecological perspective, these comparisons are greatly needed; from a
theoretical standpoint, however, they do not contribute to determine the relative weight of positive evidence and prompting in L2 development. So far, to my knowledge only two studies have experimentally teased out these components from negative evidence (Leeman, 2003; McDonough, 2005).

In her seminal study, Leeman (2003) investigated the separate contributions of positive and negative evidence in recasts on the development of Spanish gender-number agreement. Specifically, she engaged 74 college students in tutor-learner interaction during two communicative tasks (object placement and catalog-shopping) providing one of four types of input, as per the combination of two variables, i.e., the presence or absence of negative evidence (NE) and positive evidence (PE) with different degrees of enhancement. For example, after saying the ill-formed phrase *la taza *rojo (“the cup-FEM-SING red-MASC-SING”), participants in the Recast group (+NE, +enhanced PE) were told *Um hmm, una taza roja. ¿Qué más? [“Um hmm, a red cup. What else?”]; in turn, participants in the Negative evidence group [+NE, -PE] got *Um hmm, pero tú dijiste ‘una taza *rojo’ ¿Qué más? [“Um hmm, but you said ‘a *red cup.’ What else?”]; conversely, students in the Enhanced salience group [-NE, +enhanced PE] did not get any feedback after errors but heard models with rising intonation on the adjective ending, e.g., *La manzana roja está en la mesa [“The red apple is on the table”]; finally, the control group [-NE, unenhanced PE] did not get any feedback and, like the other groups, received input with standard intonation. The results of the study showed that only the Recast and Enhanced-salience groups outperformed the control group on posttreatment measures, which according to Leeman suggests that the utility of recasts is derived at
least in part from enhanced salience of positive evidence, whereas their implicit negative evidence may not be a crucial factor.

Instead of positive evidence, in another pioneering study McDonough (2005) explored the separate contribution of prompting for repair, or pushed output, versus negative evidence, on the development of English question formation. Participants were 60 Thai college students of ESL who initially ranked stage 4 of question development on Pienemann’s model (Pienemann & Johnston, 1987; Pienemann et al., 1988). To create different opportunities for students to modify their output, McDonough manipulated two independent variables, i.e., provision of clarification requests (e.g., pardon?, what?, huh?), and repetition of learners’ errors, which were aimed at helping learners solve the blame assignment problem. The interaction of these two variables, [+/- error enhancement, E] and [+/- clarification request, C] led to 4 different groups (my own acronyms). In the EC group the tutors responded to learners’ errors by repeating the non-target-like part of their questions with stress and rising intonation, immediately after which they formulated an open-ended clarification request. In the C group, the tutors gave no clues about the sources of the errors and formulated open-ended clarification requests, making a pause to give learners the opportunity to modify their output. In the E group the tutors highlighted the non-target-like sequences and continued talking so that the learners did not have an opportunity to acknowledge the feedback or modify their output. Finally, in the Control group, the NS interactors did not provide any feedback at all when the learners produced non-target-like questions, feigning understanding. For all groups, the treatment involved carrying out a series of communicative tasks (information-exchange and information-gap) with an NS counterpart in three 10-minute sessions over
the course of 8 weeks. The analysis of the results showed that the only significant predictor of English question development was the production of successfully modified output in response to negative feedback. Clarification requests thus seemed to play an indirect rather than direct role in English question development, with their effectiveness being subject to the production of successfully modified output on the part of the learner.

Overall, then, the findings by Leeman and McDonough for recasts and clarification requests suggest that under specific circumstances, positive evidence and prompting for repair may outweigh negative evidence per se. This clearly casts many shadows on the existing literature on type of feedback in oral interaction, which has often made claims on the efficacy of implicit vs. explicit negative evidence without isolating the other two dimensions of feedback moves. Additionally, none of the four published meta-analyses of feedback research to date (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006) has included any study from the CALI strand of research, published or otherwise. Therefore, a separate synthesis of this growing body literature is in order.

2.2 Feedback in CALI

Hitherto, research on CALI feedback has spawned a number of different operationalizations and labels that may lead to erroneous comparisons and deceitful conclusions. The two pioneering studies in the field, Nagata (1993) and Nagata and Swisher (1995), confronted “traditional computer feedback” vs. “intelligent computer feedback.” In turn, Rosa and Leow (2004a; 2004b), Bowles (2005), and Hsieh (2007) opposed “implicit-” to “explicit feedback.” Camblor (2006) additionally introduced a third label, so-called “interactive” feedback. Finally, Sanz and Morgan-Short (2004),
Lado (2008), and Moreno (2007) referred to “more vs. less explicit” types of feedback, a dichotomy which they expressed via the [+/-] mathematical operators and which can also be misread as “presence vs. absence” of explicit feedback. There is, therefore, an urgent need for terminological normalization.

A key terminological observation concerns the occasional use of the word implicit to refer to feedback types that signal an error without explaining its nature. Strictly speaking, no CALI study thus far has used “implicit feedback” proper, as they all rather overtly inform learners of the occurrence of an error (see Chapter 3). Bowles (2005), Hsieh (2007), and Moreno (2007) attempted to provide implicit feedback in a CALI study by embedding it into their tasks; however, they converted implicit feedback into explicit feedback by preliminarily explaining its significance to their participants. For example, Bowles’s (2005: 114) written instructions to the participants in the “implicit feedback” condition read:

“ONLY BY COMBINING THE CORRECT WORDS IN THE CORRECT ORDER WILL YOU BE ABLE TO FIND THE PATH THAT LEADS OUT OF THE MAZE. Your task is to find your way through the maze by choosing the paths to the words you think form the correct sentence in Spanish. If you choose a word that does not grammatically fit in the sentence, you will run into a dead end (a wall) soon after. You will have to go back and choose another word (and another path) each time you make incorrect choices.”

Similarly, Hsieh’s (2007: 114) “implicit feedback” groups received the following instructions:

If you happen to choose the wrong path, your [sic] will see a wall blocking you from going on, then you will need to go back and choose another option. You will be allowed to continue on to the path as long as you choose the correct linguistic item to describe the picture.
Finally, Moreno (2007: 111) stated:

Click on the answer that seems correct to you and see what happens. Do this until you have formed the sentence that best describes the picture. Your goal is to reach the end of the path successfully without bumping into too many dead-end roads.

Moving away from task-embedded feedback to textual feedback, Camblor’s (2006) “implicit feedback group” constitutes at first glance the closest approximation to an instantiation of implicit feedback in CALI, but a closer look at it reveals a number of problems. In her study, targeting Spanish noun-adjective number and gender agreement via a picture-description task, participants had to write sentences containing the name and color of a series of objects and their location in a living room (e.g., *Las plantas anaranjado está sobre la mesa* ['The orange-SINGULAR-MASCULINE plants is on the table.']). Next, the computer bounced back the sentence the student wrote, with the lexical part of the agreeing adjective in capital letters and the grammatically (in)correct gender-number morpheme in standard lowercase (e.g., *Las plantas ANARANJADO está sobre la mesa*). Although this sort of contrastive textual enhancement may be argued to corrupt the intended implicitness of the feedback (see Chapter 3), an extra problem arises when Camblor explicitly directs their students’ attention to the capitalized letters through the preliminary instructions to the task:

After you write each sentence, click the ‘Submit’ button. You will be invited to review your sentence. Pay attention to the CAPITALIZED words and rewrite your sentence, making any necessary changes that you consider appropriate. Then, resubmit your sentence and go on to the next picture.
Although most certainly these instructions do not explicitly tell students that the non-capitalized morphemes following the capitalized lexemes constitute a potential source of error, the idea is indeed in the subtext. Thus, whether this particular instantiation of “implicit” feedback is really implicit is at the very least debatable. Most importantly, however, as pointed out by Lyster and Ranta (1997: 49) à propos of recasts, the juxtaposition of an elicitation move (“re-write your sentence”) after an implicit feedback move corrupts the intended implicitness. In light of this, Camblor’s (2006) “implicit feedback” group is certainly hard to label as “implicit” (more on this in section 2.2.1).

Switching from CALI to CALL, a growing number of CMC studies (where computers are used as media or means of communication rather than tutors), have directly or indirectly studied the role of truly implicit feedback delivered by human interactors. Thus, Kötter (2003) and Morris (2005) observed that implicit feedback occurs naturally in CMC, whereas Choi (2000), Sachs and Suh (2007), and Sagarra (2007) observed the impact of recasts on L2 development.

In conclusion, then, existing literature on computerized feedback (as opposed to human-generated feedback delivered over a computer) has exclusively dealt with different degrees of explicit, rather than implicit, feedback. Additionally, however, the term explicit has often been used to refer to very different operationalizations, involving different amounts of error explanation in different codes (plain language or meta-talk). Conversely, in other cases different terms have been used to refer to rather similar types of feedback. To fill this gap and gain a better understanding of the field, in terms of what has been explored so far and what remains to be explored, section 2.2.1 below re-
organizes the CALI feedback literature based on the feedback taxonomy proposed in Chapter 3. Next, section 2.2.2 presents a summary of the main empirical findings.

### 2.2.1 Re-organizing feedback types in the CALI literature

Chapter 3 presented a comprehensive classification scheme of feedback types based on three basic components, i.e., negative evidence with different degrees of explicitness (implicit, explicit [signal error, spot error, explain error]), positive evidence (+/-PE) and requests for error repair (+/- prompt). Also, long-standing feedback labels such as the ones in Lyster and Ranta (1997) were revisited and assigned a place within this taxonomy. In the following pages the proposed scheme will be used to classify the different feedback types that have been reported thus far in the CALI literature. The sources under consideration include published studies by Nagata (1993) and its revised study by Nagata and Swisher (1995), Rosa and Leow (2004b), and Sanz and Morgan-Short (2004). Also, the following unpublished doctoral dissertations are included: Bowles (2005), Camblor (2006), Hsieh (2007), Lado (2008), and Moreno (2007). The forthcoming discussion is best read in tandem with Table 4.1 below and Appendix 4.1. Table 4.1 assigns a place to each feedback type within the proposed classification scheme, whereas Appendix 4.1 describes the experimental tasks and target forms used in each study and contains illustrative examples of the feedback messages delivered in each experimental condition. In the discussion that follows, each feedback move is tagged in terms of its three components, and, additionally, a happy smiley signals whether learners ultimately received positive evidence before moving on to the next item in the practice.
Less explicit feedback types in CALI

As discussed earlier, Camblor’s (2006) “implicit feedback” may at first glance be considered the most implicit type of feedback in CALI research to date. However, this is debatable for a number of reasons. In Camblor’s study, students were exposed to a number of pictures and were asked to write full sentences describing them, which required elicitation of the targeted form, noun-adjective gender and number agreement. The computer then processed students’ utterances and bounced them back, capitalizing the lexeme of the agreeing adjectives and leaving the morphemes in standard lowercase (e.g., Las plantas *ANARANJADO está sobre la mesa [“The *ORANGE-singular-masculine plants *is on the table”]). Students were then prompted to “pay attention to the CAPITALIZED words” and rewrite their sentence, making any necessary changes that they considered appropriate. This combination of intendedly implicit feedback, textual enhancement, and a request for output modification make Camblor’s “implicit feedback” quite hard to label. In principle, it could be posited that a student may read his/her bounced-back sentence, observe the sequence ANARANJADO, not realize that the lowercase –o ending is the target form, and retype the same phrase. According to this, Camblor’s “implicit feedback” type could be tagged as [implicit, -PE, +Prompt, ☹]. On the other hand, if as discussed in Chapter 3 “implicit feedback” is that particular type of feedback which may go unnoticed as such, this seems unlikely for a number of reasons. First, the use of textual enhancement to highlight the targeted structure may corrupt the implicit nature of the feedback. Second, it could be posited that the juxtaposition of an elicitation move (“Rewrite your sentence in the space below”) right after the feedback interrupts the normal flow of the discourse and corrupts the implicit nature of the
feedback, as argued by Lyster and Ranta (1997: 49). Third, the type of message conveyed in the feedback, i.e., the repetition of the student’s utterance as is, with the targeted form capitalized and an instruction to focus on it, seems to place Camblor’s “implicit feedback” in the explicit category, specifically the type that Lyster and Ranta (1997) called “repetition.” As a matter of fact, as these authors (1997: 48) stated, while teachers are producing this type of feedback they usually “adjust their intonation so as to highlight the error,” which is what Camblor did via text enhancement. Therefore, in this discussion Camblor’s “implicit feedback” is tagged as [explicit (Spot error), -PE, +Prompt, 🙁].

Also using the label “implicit feedback” but focusing on Spanish gustar constructions (A mí me gusta el helado [“I like ice cream”]), Bowles (2005) and Hsieh (2007) attempted to provide implicit feedback in their CALI studies by embedding it into their input-based tasks. In Bowles (2005), participants were to find their way out of a maze by selecting and combining, in the correct order, the Spanish words that translated an English sentence. Similarly, in Hsieh (2007), participants were prompted to navigate a series of forking paths with word options in order to sequentially translate a picture cue. In both studies, participants ran into a dead end whenever they selected an inappropriate word in the sequence. Participants were thus forced to select a different word as many times as necessary until they reached the exit of the maze or the end of the forking paths. Although, as argued by Lyster and Ranta (1997: 49), the inclusion of an elicitation move would suffice to corrupt the intended implicit nature of these feedback moves, the authors further turned them into explicit feedback moves by explaining the significance of a dead end to their participants. In addition, given the sequential structure of their input-focused tasks, learners were also aware of the precise location of their errors in the sentence. They
did not, however, have access to positive evidence right away, as they had a varying number of options to choose from. For all of the above, Bowles’s (2005) and Hsieh’s (2007) “implicit feedback” moves may be tagged as [explicit (Spot error), -PE, +Prompt, ☺], constituting examples of what has traditionally been referred to as “elicitation” feedback.

In Rosa and Leow (2004b) “implicit feedback” was used for the instruction of “if” clauses in Spanish past counterfactual conditional sentences (i.e. *Si hubieras parado de beber cerveza…* [“If you had stopped drinking beer…”]). Their experimental task, an input-focused one, consisted of a battery of cloze sentences where students had to select the protasis or “if” part from four possible options. After each selection, participants received textual feedback that overtly indicated whether the answer was right or wrong, and in the latter case, they were invited to retry until they made the right choice. Since the options available only differed in the tense of the verbs, the location of the error was evident. However, since there were four options at play, the solution to the problem after an error was not straightforward. In sum, Rosa and Leow’s (2004b) feedback move was an instance of “elicitation” [explicit (Spot error), -PE, +Prompt, ☺].

In Nagata (1993) and Nagata and Swisher (1995) students were asked to complete an output-focused task, sentence translation, eliciting Japanese passive constructions. After each sentence was typed, the CALI system processed them and, in case of an error, it returned so-called “traditional CALI feedback.” Specifically, this feedback informed learners about whether a word was missing, incomplete, or not expected to be used. After a maximum of three trials, the computer displayed the correct answer, and the student was asked to reproduce it. Consequently, like Bowles (2005), Hsieh (2007), and Rosa and
Leow (2004b), Nagata’s (1993) and Nagata and Swisher’s (1995) “traditional CALI feedback” could be considered an instance of “elicitation” of the type [explicit (Spot error), -PE, +Prompt, ☺]. It should be noted, however, that this type of feedback move is more explicit in nature than its cellmates in Table 4.1.

In her doctoral dissertation, Moreno (2007) used two different tasks with concurrent “[-explicit feedback]” to instruct Spanish pre-verbal direct object pronouns (La ama Juan [“Juan loves her”]). One of the tasks, similarly to Hsieh (2007), required students to navigate through a forking path to correctly translate a picture cue. However, whereas Hsieh focused on all subcomponents of the gustar structure and required students to consecutively select each subcomponent from an array of two or three options, Moreno focused on direct object pronouns exclusively, limiting the choice to only two options. If the students selected the wrong option, the “trip” stopped, a red traffic light appeared, and they were given an opportunity to retry, presumably selecting the alternative option. In the unlikely event that they clicked again on their original choice, the solution was displayed before proceeding to the next item. The other task in the study was an interpretation task requiring students to describe a picture by choosing one of two possible sentences which only differed in the subject. Again, the correctness of the student’s choice was signaled by a green or red light, and, in the latter case, an invitation to retry followed. In sum, similarly to Bowles (2005), Hsieh (2007), and Rosa and Leow (2004b), Moreno’s (2007) instantiations of “[-explicit feedback]” were explicit in nature, prompted error repair, and spotted the exact error location. However, differently from these studies, here positive evidence was indirectly provided, since there were only two options at play. Thus, while all these studies used an input-focused task, the greater array
of options in Bowles (2005), Hsieh (2007), and Rosa and Leow (2004b) defines their “implicit feedback” as “elicitation,” whereas Moreno’s (2007) “[explicit feedback]” is actually closer to “explicit correction,” defined as [explicit (Spot error), +PE, +Prompt, 😊].

Also focusing on Spanish pre-verbal direct object pronouns, Sanz (2004) and Sanz and Morgan-Short (2004) used a series of interpretation tasks with concurrent “[explicit feedback]”. In one of their tasks, students were presented with a written or oral sentence for which they had to select one of two pictures. In the other task students read a Spanish text and were to select the meaning of its title out of two possible translations. When students clicked on the wrong option, the computer displayed the message “Sorry, try again;” conversely, if the response was correct, the next item was presented. As stated by Sanz and Morgan-Short (2004: 55), the feedback was “provided only when needed.” Since in case of an error learners were explicitly informed of its occurrence, Sanz and Morgan-Short’s “[explicit feedback]” group was explicit in nature. Also, similarly to Moreno (2007), since there were only two options at play, the solution to the problem was indirectly provided, and since these options only differed in terms of their thematic roles (doer vs. doee), the error location (i.e., the subject or the direct object pronoun) was easy to spot. Consequently, Sanz’s (2004) and Sanz and Morgan-Short’s (2004) [explicit feedback] may be tagged as [explicit (Spot error), +PE, +Prompt, 😊]. In oral interaction terms, this feedback type constitutes, along with Moreno (2007), an example of “explicit correction.”

Framed under The Latin Project, Lado (2008) also used [explicit feedback] of the right/wrong type in a wide range of tasks aimed at instructing the assignment of semantic
functions in Latin. Her tasks were administered in both the written and aural modes, and involved either recognition or interpretation of the form among a series of binary choices. Differently from previous studies, both the choices and the feedback were timed, to control for the duration of exposure. Similarly to Sanz and Morgan-Short (2004) and Moreno (2007), the binary nature of Lado’s tasks allowed for easy spotting of the error locus and provided at least indirect access to positive evidence; however, her e-tutor did not prompt for repair. Consequently, her “[explicit feedback]” may be tagged as [explicit (Spot error), +PE, -Prompt, ☺].

To wrap up, then, the following conclusions can be extrapolated from the previous discussion: (1) no CALI study to date has explored the effects of purely implicit feedback, as they all use different types of “Spot error” explicit feedback, i.e., “repetition” (Camblor, 2006), “elicitation” (Bowles, 2005; Hsieh, 2007; Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004b), and “explicit correction” (Lado, 2008; Moreno, 2007; Sanz, 2004; Sanz & Morgan-Short, 2004); (2) no CALI study has explored the potential of oral feedback, as they all presented it visually in text form (Camblor, 2006; Lado, 2008; Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004b; Sanz, 2004; Sanz & Morgan-Short, 2004) or embedded in the task (Bowles, 2005; Hsieh, 2007; Moreno, 2007); (3) almost all CALI studies failed to isolate exposure to positive evidence, which was presented either via the feedback message or implicitly in the task by an elimination process that was possible due to the input-focused nature of many CALI tasks (Bowles, 2005; Hsieh, 2007; Lado, 2008; Moreno, 2007; Rosa & Leow, 2004b; Sanz & Morgan-Short, 2004); (4) only two studies have investigated the effects of CALI feedback with output-focused tasks (Camblor, 2006; and Nagata, 1993;
Nagata & Swisher, 1995); (5) All CALI studies but Lado (2008) prompted learners to repair their errors, although the number of opportunities varied (Camblor, 2006: one re-trial; Moreno, 2007 and Sanz & Morgan-Short, 2004, presumably one re-trial; Hsieh, 2007 and Nagata & Swisher, 1995, up to two re-trials; Rosa & Leow, 2004, unlimited re-trials, but presumably up to three); finally, (6) all CALI tutors but Camblor’s (2006) granted access to positive evidence before moving on to the next practice item.

**More explicit feedback types in CALI**

All of the CALI studies under discussion here confronted less explicit feedback types against more explicit ones providing a rationale for the errors made. For example, in Camblor’s (2006) CALI tutor students in the “explicit feedback” condition received the following message: “Noun-adjective agreement in this sentence is not correct!” In this study, the task was output-based and no grammar lessons were provided, so learners were left to their own devices to come up with the solution. Only one retrial was allowed, and no positive evidence was given if the response was incorrect. Consequently, then, Camblor’s “explicit feedback” can be tagged as [explicit (Explain error), -PE, +Prompt, 😞].

In contrast to Camblor (2006), all the other CALI studies reviewed here directly or indirectly provided positive evidence in addition to negative evidence. As shown by the examples in Appendix 4.1, Bowles (2005), Hsieh (2007), Nagata (1993) and Nagata and Swisher (1995), and Rosa and Leow (2004b) all included the correct form directly in the feedback. However, given the different nature of their tasks using the feedback was not always straightforward. In the studies that used an input-focused task (Bowles, 2005;
Hsieh, 2007; Rosa & Leow, 2004b) the learner’s job after reading the “explicit feedback” was to select the option containing the form disclosed by the feedback message. In Nagata (1993) and Nagata and Swisher (1995), however, given the output-focused nature of the task, students could use the instructions in the feedback to edit their sentences, deleting unexpected words and inserting suggested ones, but they still had to make syntactic choices, such as where to place what. As opposed to these studies, Moreno (2007) and Sanz and Morgan-Short (2004) did not provide the correct form directly in their [+explicit feedback], trying to elicit it on the part of the student (e.g., “You need a feminine direct object pronoun” in Moreno, or “a direct object pronoun is never the subject” in Sanz and Morgan-Short). However, since their tasks were input-based and involved binary choices, the solution was indirectly provided. Despite the differences just mentioned, all the studies above provided feedback types of the sort [explicit (Explain error), +PE, + Prompt, ☺].

In sum, then, as discussed earlier à propos of less explicit feedback types, the different operational definitions of more explicit feedback in the CALI literature have all (1) used the written rather than the oral mode; and (2) failed to isolate negative evidence from positive evidence and prompting for error repair. Table 4.1 below provides a visual representation of this discussion based on the proposed classification scheme.
Table 4.1 Re-organizing feedback types in the CALI literature

<table>
<thead>
<tr>
<th>Negative evidence</th>
<th>PE</th>
<th>Spot error</th>
<th>Explain error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**PE**

- **Repetition:**
  - Camblor’s (2006) “implicit feedback”. (?) ☺
- **Elicitation:**
  - Bowles’s (2005) “implicit feedback” ☺
  - Rosa and Leow’s (2004b) “implicit feedback” ☺
  - Nagata (1993) and Nagata and Swisher’s (1995) “traditional CALI feedback” ☺

**Explicit correction:**

- Lado’s (2008) “[-explicit feedback]” ☺
- Moreno’s (2007) “[-explicit feedback]” ☺
- Sanz and Morgan-Short’s (2004) “[-explicit feedback]” ☺
- Lados’s (2008) “[+explicit feedback]” ☺

**Explicit correction:**

- Camblor’s “explicit feedback” (2006) ☺
- Hsieh’s (2007) “explicit feedback” ☺
- Moreno’s (2007) “[+explicit feedback]” ☺
- Nagata (1993) and Nagata and Swisher’s (1995) “intelligent CALI feedback” ☺
- Rosa and Leow’s (2004b) “explicit feedback” ☺
- Sanz and Morgan-Short’s (2004) “[+explicit feedback]” ☺
2.2.2 Empirical findings on feedback in CALI

Leaving terminological and methodological differences aside, the growing literature on computerized feedback shows an overall positive effect of feedback in L2 development, while the jury is still out as to the relative efficacy of different types of feedback. Existing research paints a rather blurry picture, with some studies showing a significant advantage of the more explicit conditions —see the published studies by Nagata (1993) and Nagata and Swisher (1995), and Rosa and Leow (2004b) and the unpublished doctoral dissertations by Bowles (2005) and Lado (2008)— and a comparable number of studies showing no significant differences between more and less explanatory types of feedback —see the published studies by Sanz (2004), also reported in Sanz & Morgan-Short (2004), and the unpublished doctoral dissertations by Camblor (2006), Moreno (2007), and Hsieh (2007). What follows is a critical review of these studies.

In Nagata (1993) and Nagata and Swisher (1995) 32 college students of L2 Japanese were exposed to Japanese passive constructions via prior explanation plus different types of computerized feedback. After a series of instructional lessons, participants were asked to complete a task where they were to respond, in writing, to a series of Japanese sentences produced by an imaginary partner. During their interactions with the task, the group receiving less explanatory feedback was informed of all missing or unexpected forms in their responses, whereas the group receiving more elaborated feedback was also provided with detailed information about the nature of the errors made. Results of a written translation task showed significantly better performance for the more explanatory
feedback group in both the immediate and the delayed production posttest (3 weeks later).

In Rosa and Leow’s (2004b) study, reporting on Rosa’s (1999) doctoral dissertation, 100 college students of intermediate Spanish were exposed to Spanish counterfactual conditional sentences under one of six conditions, as per the combination of prior grammatical information [±explicit pretask] and feedback with different degrees of explicitness [rule, no rule, no feedback]. The experimental task consisted of a battery of two types of conditional sentences where the verb in the protasis or “if” part had been blanked out, and participants were to fill in the blanks by selecting one of four verb forms. With regards to feedback specifically, the results of this study showed that more explanatory feedback led to significantly better performance in production (of both old and new exemplars) and in recognition of new exemplars. Furthermore, it was found that feedback provided via interactive practice seemed to play a greater role than prior grammatical explanation, since the presence or absence of the latter led to no significantly different gains.

In Bowles’s (2005) doctoral dissertation, 150 college students of introductory Spanish were exposed to the dative experiencer construction with gustar via an interactive task inspired by the videogame Pacman. In particular, participants were to find their way through a series of mazes by selecting and collecting, in the right order, the words that made up the Spanish translations for a battery of English sentences. Participants in the “explicit” conditions got an explanation of the error after each word-choice, while those in the “implicit” conditions ran into a dead end whenever their word-choice was incorrect. Results based on a sentence translation production posttest revealed that,
although all participants significantly improved from the pretest to the immediate posttest, those in the more explicit feedback condition experienced significantly greater gains for both old and new items. This difference, however, did not hold three weeks later.

In Lado’s (2008) doctoral dissertation, 151 college students with different proficiency levels of L2 Spanish (basic, intermediate, advanced, and native-like) completed a computerized treatment using input-based practice and one of two types of explicit feedback (with and without grammatical rules). The targeted form was the assignment of semantic functions in Latin (who did what to whom). Results of a battery of tests showed an edge of the more explicit condition both for interpretation and production in the immediate posttests; however, these results did not hold constant for all tests. In terms of response times (latency), participants in the less explicit conditions were significantly faster.

The four studies reviewed above support the thesis that providing learners with grammatical information about the nature of their errors during the completion of an interactive computerized task is more beneficial than just informing learners about how accurate their answers are. However, this thesis is counterevidenced by four studies that found no significantly different learning outcomes for these two conditions (Camblor, 2006; Hsieh, 2007; Moreno, 2007; Sanz & Morgan-Short, 2004).

Like Rosa and Leow (2004b), but from a processing instruction (PI) approach, Sanz and Morgan-Short (2004) examined the effects on L2 development of grammatical rules presented before and during interactive practice. Participants were 69 students of introductory Spanish, and the targeted form was Spanish preverbal direct-object
pronouns, to which participants were exposed under 1 of 4 conditions, as per the combination of two variables: preliminary explicit rule presentation [±explanation] and explicit negative feedback [±explicit feedback]. The treatment session consisted of a computer-delivered lesson with referential and affective activities, combining oral and written input at both the textual and sentential levels. For every instance, participants were asked to select the correct visual or textual cue out of two possible options. The results of a variety of posttests measuring interpretation and production achievement revealed a significant improvement over time for all conditions, but no differential effects for either prior explanation or feedback. Based on these results, Sanz and Morgan-Short (2004: 71) concluded that “when learners are exposed to positive evidence presented through task-essential practice, the provision of explicit information through explanation or feedback may not produce increases in acquired knowledge.”

Expanding on previous research, Camblor’s (2006) doctoral dissertation introduced a new type of feedback (so-called ‘interactive’ feedback) and compared its relative effects on the acquisition of Spanish gender and number noun-adjective agreement. In her study, 77 students of introductory Spanish were asked to describe, in writing, a series of pictures of objects in different colors, to which the computer responded in one of four ways: the participants in the ‘implicit’ group received feedback in the form of input enhancement (mistakes in the learner’s output were presented is small case); the ‘explicit’ group was systematically provided with a metalinguistic explanation of the mistakes; the ‘interactive’ group received so-called interactive feedback provided via a secondary task embedded within the main picture description task —a sort of step-by-step input-based tutorial; and finally, the control group received no feedback. In line with Sanz and
Morgan-Short (2004), the results of a controlled and a semi-open production posttest revealed that, while feedback had a beneficial effect for all groups, no group experienced statistically significantly higher gains, either for old or new exemplars.

Finally, both Moreno (2007) and Hsieh (2007) used an input-focused task with right/wrong feedback with and without metalinguistic rules. Moreno targeted Spanish pre-verbal direct object pronouns for third person singular and plural, while Hsieh targeted Spanish dative experiencer constructions with *gustar*. Both of them used a “forking paths” task in the treatment, where students gradually selected one of various constituents to compose a sentence describing a picture. Moreno additionally used a recognition task, where students had to match one of two full sentences to a picture. Overall, neither Hsieh nor Moreno found a difference for type of feedback on the immediate posttests. Moreno, however, found that the less explicit feedback group had an edge on her oral delayed posttest.

Altogether, the studies reviewed above present remarkable strengths. For example, from a methodological perspective, most of these studies have boasted tight control over provision of feedback both in terms of type and frequency; furthermore, from a pedagogical perspective, these studies are indicative of a clear trend towards richer, more sophisticated, more learner-centered, and probably also more engaging materials. Nevertheless, these studies also present several important limitations, i.e., (1) some studies did not adequately isolate feedback from prior metalinguistic instruction (Nagata, 1993; Nagata & Swisher, 1995); (2) all studies failed to tease out negative evidence from positive evidence and prompting for error repair, and (3) no study investigated the potential of oral feedback in CALI. These limitations, along with the blatantly
inconsistent findings to date, certainly call for further investigation on computerized feedback in CALI. This research should not only replicate studies on type of feedback but also address possibly intervening variables such as amount of exposure to input, task complexity, and type of grammatical form (Cerezo, 2007; Cerezo & Leow, 2006; Sanz & Morgan-Short, 2004).

Section summary: Feedback

The role of corrective feedback has fed a heated debate in the language acquisition literature. In line with Chomsky, many first language acquisitionists have derided its role, arguing that only exposure to positive evidence activates the sets of UG parameters that work for a particular language (Carroll, 1996; Cook, 1991; Schwartz, 1993). This idea has been supported by second language acquisitionists like Krashen (e.g., 1982 and elsewhere), according to whom only implicit processing of comprehensible input can lead to L2 acquisition. Currently, however, after three decades of empirical research there is substantial evidence attesting that corrective feedback significantly speeds up the learning process, both in classroom and laboratory contexts, and both with children and adults (Li, 2010; Lyster & Saito, 2010; Russell & Spada, 2006). While this does not resolve whether feedback is necessary for L2 acquisition, there is increasing support to the thesis that adults cannot acquire native-like competence in an L2 based on positive evidence alone because there are maturational constraints on language learning (Hyltenstam, 1988; Newport, 1990, cited in Long & Robinson, 1998) and because transfers from the L1 may have a negative impact when the L1 subsumes the L2 (White, 1989; 1991).
Early empirical research on the role of corrective feedback for L2 grammar development focused almost exclusively on recasts, because they were identified as the most frequently occurring and least disruptive type of feedback (e.g., Han & Kim, 2008; Loewen & Philp, 2006; Long, 2007; Lyster & Ranta, 1997; Morris, 2002; Nicholas et al., 2001; Oliver, 2000). Although the facilitative effects of recasts have been largely corroborated (see, e.g., Long, 2007 for a review), there are discrepancies in the learning outcomes that they may lead to, both independently and vis-à-vis more explicit types of feedback, quite possibly as a result of very different operational definitions and implementations (Nassaji, 2007: 520-1). Addressing these caveat, recent studies (e.g., Loewen & Philp, 2006; Nassaji, 2007; Sheen, 2006) have started to isolate the different components of recasts, showing that it is the most explicit types (short, declarative, reduced, repeated, with a single error focus) that yield the highest rates of uptake and learning.

After the initial boom of research on recasts, authors set out to compare the relative efficacy of different types of feedback. Early work largely compared feedback types in terms of their “explicitness,” understood as either the overt/covert correction of an error or the provision/absence of metalinguistic rules. Recently, however, some authors have suggested that feedback types should not be classified based on the explicitness criterion, but rather, on whether or not they encourage learners to repair their errors (Lyster & Saito, 2010; Ranta & Lyster, 2007). Overall, the results of four recent meta-analyses of empirical research show that corrective feedback has clear facilitative effects for SLA (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006); however, there is no consensus as to which type of feedback yields the highest effect.
sizes. Russell and Spada (2006) and Mackey and Goo (2007: 440) concluded that more studies are needed to decide whether implicit and explicit types of feedback are differently beneficial; Li (2010) concluded that explicit feedback outperforms implicit types in the short to medium run; and Lyster and Saito (2010) concluded that recasts and explicit correction are not differently beneficial, and that it is prompts that lead the way.

Arguably, these conflicting findings may be explicated by a lack of control over several external factors, such as learners’ individual differences, the presence of prior metalinguistic instruction, different practice modes, task-features, learning contexts, and/or targeted forms. More importantly, however, the conflict may originate from a lack of control over internal factors. Existing research has mostly examined feedback types in terms of the explicitness of the negative evidence that they provide, overlooking the fact that feedback moves can also encapsulate positive evidence and prompt learners to repair their errors, and failing to address the independent and combined effects of these three components. On this token, the findings of seminal studies by Leeman (2003) and McDonough (2005) suggest that under specific circumstances, positive evidence and prompting for repair may outweigh the impact of negative evidence per se. Clearly, then, the field needs a Copernican twist of sorts.

Additionally, empirical research on CALI feedback has been completely ignored by the four feedback meta-analyses published to date (Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006). Overall, the findings of this field of inquiry mirror those of the human interaction literature: an overall positive effect of feedback in L2 development and no consensus for type of feedback, with some studies showing an advantage of more explicit over less explicit types (Bowles, 2005; Lado,
Using the feedback typology template proposed in Chapter 3, I revisited the feedback types investigated in these studies and drew the following conclusions: (1) CALI studies have addressed different types of explicit feedback, including repetition, elicitation, and explicit correction, but no study has yet explored the effects of purely implicit feedback; (2) with the exception of Camblor (2006), all CALI studies failed to isolate positive evidence from negative evidence, since the correct solutions were presented either in the feedback message or implicitly in the task by an elimination process; and (3) no study isolated the role of pushed repair, as all of them prompted learners to fix their errors. Additionally, (4) only two studies investigated the effects of CALI feedback with output-focused tasks (Camblor, 2006; and Nagata, 1993; Nagata & Swisher, 1995), (5) no CALI study has explored the potential of oral feedback, as they all presented it visually in text form or embedded in the task, and (6) there is a blatant need for studies to address the role of possibly moderating variables such as type of linguistic form. These limitations, in light of the blatantly conflicting findings to date, certainly call for further research addressing the independent and combined contributions of the three components of feedback, the potentially facilitative effects of oral feedback, and the moderating role of variables such as type of grammatical form.
3. Grammatical form

In recent years, several SLA studies have considered how type of linguistic form may mediate the effects of a particular pedagogical intervention. Among others, four different strands of research have addressed this, i.e., research on input processing, input enhancement, processing instruction, and studies addressing the role of explicitness in L2 instruction, either via rule presentation or corrective feedback. Research on input processing (e.g., Greenslade et al., 1999; VanPatten, 1989, 1990; Wong, 2001) has shown that instructing beginning L2 learners to focus their attention on a lexical item in a reading or listening task does not produce detrimental effects on comprehension, whereas instructing them to focus on a morphological form does, even more so if the form being paid attention to has a low communicative value. Research on input enhancement (Leow, 1993; Leow et al., 2003; Shook, 1994) has shown a more complex picture. For example, while Shook (1994) found that input enhancement had a greater positive impact on intake of a more communicative form over a less communicative one, Leow (1993) and Leow et al. (2003) found no significant differences in intake or comprehension of two forms with different degrees of salience. Finally, DeKeyser and Sokalski (1996; revised in 2001) also used two grammatical structures to compare the effects of Processing Instruction (PI) against Output-Based Instruction (OBI). Based on their results, they concluded that when the difficulty of a form resided in its meaning, PI yielded better results in comprehension and OBI in production; however, if the difficulty resided in the form PI was beneficial only in production, while OBI yielded better results in both tasks.

Overall, these studies showed that the type of targeted linguistic form at play can moderate the effects of different form- or meaning-focused pedagogical interventions.
However, even more interesting for the purposes of this dissertation is a batch of studies analyzing the role of linguistic form in moderating the effects of instruction with grammatical rules versus less explicit instructional techniques. In some of these studies grammatical rules were provided as part of input presentations, while in others they were embedded in corrective feedback in the course of grammar practice. These two blocks of studies are separately discussed below, and summarized in tabulated form in Appendix 4.2.

3.1 Rules in the input

To date, several empirical studies have compared the effects of input presentations with and without metalinguistic rules on the development of two or more targeted grammatical structures (syntax or morpho-syntax). Some studies have been conducted in the classroom (e.g., Scott, 1989), while others were conducted in laboratory settings, using both traditional pen and paper materials (e.g., Alanen, 1995) and computerized treatments (e.g., Gass, Svetics, & Lemelin, 2003; DeKeyser, 1995; Robinson, 1996). Some used natural languages such as French (Scott, 1989; Ayoun, 2001), English (Robinson, 1996), and Italian (Gass et al., 2003), while others used partially or totally artificial languages (e.g., manipulated Finnish in Alanen, 1995, and Implexan in DeKeyser, 1995). Below is an account of these studies.

In Scott’s (1989) study, 34 college students of advanced French were exposed to two grammatical structures, relative pronouns and subjunctive, under two different conditions. In the “implicit” condition, students heard a story with the structure embedded in it, whereas in the “explicit condition” students heard an explanation of the
rules governing the use of the structures along with some examples. The treatment for each structure took up the first 10 minutes of class over six consecutive class periods, up to a total duration of 1 hour. Assessment tests consisted of written and oral fill-in-the-blanks. Results showed that for both targeted structures the explicit condition outperformed the implicit condition on the written sections of the test, while both conditions performed similarly in the oral sections. More interestingly however, although not discussed by Scott, the gain score means on the written test were clearly higher for one structure over the other. Thus, out of a maximum of 15 points, students in the explicit condition experienced an average gain of 1.7 for relative pronouns and 5.7 for the subjunctive; similarly, the implicit condition gained .35 for the former and 2.3 for the latter. Clearly, then, both treatments seemed to work better for French subjunctive than relative pronouns, although it remains unknown whether these differences reached significance. Additionally, several caveats urge caution when interpreting these findings. First, the amount of exposure to the two structures was not controlled for, with many more tokens of pronouns than subjunctive verbs, and with participants in the implicit group receiving 5 times more exemplars than those in the explicit group. Second, the demands of the written tests differed for each structure. In the case of relative pronouns students were asked to recognize the correct form out of six possible choices (qui, que, dont, ce qui, ce que, ce dont) while for the subjunctive students were instructed to produce one of two forms, indicative or subjunctive, with no provided options. Third, the scoring criteria were also different. Errors in the use of subjunctive were considered acceptable if there was a clear attempt to use it, while for relative pronouns students had to select the correct form in order to score points.
In Alanen’s (1995) study, 36 college students were instructed different forms of semi-artificial Finnish under four different conditions. The targeted forms comprised two locative suffixes (\(-ssa\), meaning either “on” or “in”, and \(-lla\), meaning “on”) and two types of phonological changes (\(kk \rightarrow k\), \(pp \rightarrow p\), \(tt \rightarrow t\) double consonant reduction and \(t \rightarrow d\) voicing). The treatment involved reading a short text containing the targeted forms during two 15-minute sessions. In the Control group students read the texts as such; in the Enhance group the targeted forms were highlighted in italics; in the Rule group students read metalinguistic explanations for 5 minutes before reading the texts; and in the Rule & Enhance group the forms had been additionally italicized. Learning outcomes were measured via a written fill-in-the-blanks test. Results showed that both for locative suffixes and consonant changes the two rule-based groups performed the best, while they did not differ among each other — nor did the meaning-based groups. More interestingly, however, a more fine-grained analysis revealed that the between-group differences arose only for certain locative suffixes (\(-ssa\) meaning “in”, \(-lla\) meaning “on”) and consonant changes (double consonant reduction). According to Alanen, the lack of contrasts for \(-ssa\) meaning “on” could be due to the fact that meaning-based groups greatly used this suffix across the board while rule-based students tended to overgeneralize \(-lla\) as a result of L1 transfer. In turn, the lack of contrasts for \(t \rightarrow d\) voicing could result from low learning by all groups as a result of its low frequency in the input. These findings must be interpreted with caution due to several caveats, including varying amount of exposure to each form, relatively small cells (N=9), possible moderating effects of concurrent verbalizations, and questionable comparability between morphological and phonological levels.
In a highly sophisticated and tightly controlled experiment, DeKeyser (1995) compared the different effects of explicit-deductive vs. implicit-inductive learning for two types of structures of Implexan, an artificial language with characteristics of a natural language. The structures at hand were number and case marking on the noun and number and gender marking on the verb. For experimental purposes, DeKeyser developed two versions of Implexan, where either of the two structures behaved according to categorical or prototypical rules, in opposition to each other. Categorical rules were operationalized as those which applied systematically with no exceptions, whereas prototypical rules involved allomorphy, i.e., morphemes varied depending on the ending of the lexeme and were sometimes hard to predict. Clearly, then, these latter rules appealed not only to the learner’s ability to apply or infer rules but also to operate by “feel.” For this reason, DeKeyser hypothesized that participants would learn categorical rules better under explicit conditions, while for linguistic prototypes implicit learning would yield better results. The treatment was administered along 20 sessions of 25 minutes. During these sessions, students (N = 44) were shown a picture with its corresponding Implexan sentence. Participants in the explicit condition were additionally presented with grammar rules before the picture-sentence presentation. Throughout the sessions, participants of both conditions were periodically requested to judge whether a sentence represented a picture and they received feedback, but errors were exclusively due to incorrect vocabulary. Learning outcomes were measured via a timed picture description test, where students had to write a sentence with the correct word endings in 30 seconds. In line with DeKeyser’s hypothesizing, results showed that when the grammatical structure followed a categorical rule explicit instruction outperformed implicit instruction, which in turn did
not experience significant learning. Interestingly, these differences arose for new items only, as for old items all students seemed to rely on memory. On the other hand, if the structure behaved prototypically, implicit instruction seemed to yield the best results, although no statistical analysis was performed due to insufficient data.

In another computerized study, Robinson (1996) compared the results of four differently explicit instructional techniques on the development of two grammatical structures rated by language instructors as differently complex. The “easy” structure was subject-verb inversion in locative topicalization (e.g., Into the house ran John) while the “hard” structure was the locative pseudocleft (e.g., Where Mary and John live is in Chicago, not in New York). A total of 104 students of L2 English participated in the study. Participants in the implicit condition read sentences embedding the structures and were asked to memorize them. In the incidental condition, they read sentences for meaning and answered comprehension questions. In the rule-search condition they were instructed to find the rules underlying each structure. Finally, in the instructed condition they had access to the rules prior to and during the presentation of models. After completing two training sessions of 40 exemplars, participants took a grammaticality judgment test. Results showed that the instructed condition achieved the highest accuracy mean scores for both structures. For the simpler structure, the advantage of the instructed condition was significant against all other conditions but the incidental one, where significance was closely approached. For the harder structure the instructed condition significantly outperformed the rule-search condition only. Overall, students achieved significantly higher scores for the easy structure than the harder structure. This applied to all conditions but the implicit one, where a clear trend in this direction was noted.
Using a computerized tutor, Ayoun (2001) observed the facilitative role of prior grammatical explanation versus less explicit positive evidence (in the forms of models) and corrective feedback (in the form of recasts) in the acquisition of the aspectual distinction between two French past tenses, *passé composé* and *imparfait*. A total of 145 students participated in the study, based on a classic pretest, treatment, and posttest design, over the course of 5 weeks. Participants in the Models and Recasts groups were presented with different illustrated stories, and were respectively asked to read model sentences or to describe pictures, receiving recasts after their production, both for 3 seconds. In turn, students in the Grammar group read traditional grammar lessons, took a short practice, and checked out an answer key. The assessment test consisted of writing a composition on a given topic. Results for the two structures combined showed that the Recast group outperformed the Grammar group but did not differ from the Modeling group. However, an independent analysis for each tense revealed a more interesting picture. For the *passé composé*, all three groups experienced gains, and there were no differences between them. For the *imparfait*, however, only the Recast group experienced gains, while the other two actually performed worse than in the pretest. Ayoun attributed this to the higher difficulty of the *imparfait*, since it involves distinguishing between three aspects (imperfective, durative, and iterative) as opposed to only two in the *passé composé*, which only represents one aspect (perfective). Ayoun’s study, however, is not without limitations. First and foremost, it conflated explicitness with timing by manipulating presence or absence of metalinguistic rules in preliminary instruction and recasts, respectively. Also, it did not control for either the quantity or quality of the input. In terms of quantity, the Recasts and Modeling groups presumably practiced with more
items than the Grammar group. In terms of quality, while the former two groups had to use both tenses in an interspersed way, the Grammar group was apparently asked to provide the form after the presentation of each verb tense, hence not being pushed to make form-meaning connections in the same way.

Finally, in another computerized study Gass, Svetics, and Lemelin (2003) compared the differential effects of focused and unfocused attention on the development of three linguistic areas (syntax, morphosyntax, and the lexicon). Participants were 34 college students of L2 Italian. The syntactic structure was sentence-initial question words in biclausal interrogatives (e.g., *Per chi ha detto Jack che voleva comprare i biglietti?* [“For whom did Jack say he wanted to buy tickets?”]); the morphosyntactic structure was pre-verbal direct and indirect object pronouns (e.g., *Glielo ho fatto vedere* [“To her it I showed”]); and the lexicon was represented by words unlikely to be known by the students (e.g., *squalo* [“shark”]). In the [+focused attention] groups, students read a text containing various instances of the targeted structures, which had been underlined for them. Next, they were asked questions about the structures and were provided with rules explaining their use, or instructions on how to guess the meaning of the lexical items. Finally, they completed task-essential practice activities and read the solutions after each response. Conversely, in the [-focused attention] groups students read the same story without input enhancement, were not instructed to pay attention to the forms, answered off-target comprehension questions, did not receive rules or guessing strategies, and completed practice that did not entail production or recognition of the forms. The computerized treatment session took 50 minutes. For morphology and grammar, learning scores were measured via a grammaticality judgment test where students had to correct
incorrect sentences, whereas the test for vocabulary consisted of a sentence translation task. Results showed a reverse pattern for the two learning conditions. The explicit instruction facilitated by the [+focused attention] condition proved most beneficial for syntax and least for vocabulary, with morphology in the middle. Conversely, the purportedly incidental learning facilitated by the [-focused attention] condition yielded the best results for vocabulary, followed by morphology, and syntax. Additionally, proficiency level seemed to play a mediating role: for first year learners focused instruction proved significantly more beneficial than unfocused instruction in all three linguistic areas; for second year students it only proved more beneficial in vocabulary; and for third year students it made no differential contribution. In light of this, Gass et al. concluded that focused attention has the greatest effects in more complex linguistic areas and for less advanced students. However, these findings should be interpreted with caution due to low sample size and use of only one structure per linguistic domain.

3.2 Rules in the feedback

Very few studies have used two or more grammatical structures to compare the relative pedagogical impact of corrective feedback with or without metalinguistic rules. Several studies on recasts have opposed this type of feedback to other types of evidence, but they did not consider explicit feedback with metalinguistic rules. For example, in addition to Ayoun (2001), reviewed earlier in section 3.1, Long et al. (1998) and Iwashita (2003) observed the differential effects of recasts versus models.

Long et al. (1998) targeted two different grammatical structures of Japanese and Spanish to investigate the pedagogical effects of implicit negative feedback (recasts)
against preemptive positive evidence (models). The targeted structures at play were adjective ordering vs. locative constructions in Japanese, and object topicalization vs. adverb placement in Spanish. Participants in the experimental groups completed a series of communicative activities (oral picture descriptions) receiving recasts or models, while a control group performed distracting tasks (kanji writing or casual conversation). In their Japanese experiment (N = 24), Long et al. found that neither models nor recasts led to significant gains in the development of either structure, although both experimental groups outperformed the control group in the development of a particular subcomponent of the adjective ordering structure. In turn, in their Spanish study (N = 30) no group experienced significant learning of object topicalization, while both experimental conditions outperformed the control group for adverb placement, with recasts yielding significantly higher gains than models. Overall, then, the evidence in the Spanish study suggests that type of target form may bias the effects of recasts in L2 development, by themselves and in opposition to models. However, several methodological caveats must be considered, including low sample size, too few assessment items, and insufficient control of prior knowledge.

More naturalistically, Iwashita (2003) compared the developmental effects of freely-occurring implicit feedback vs. models on two different Japanese structures, locative-initial constructions and te-form verbs. In her study, 55 college students were paired up with native speakers of Japanese, either completing three communicative tasks (spot the difference and two picture descriptions) or engaging in casual conversation (control group). Interactional moves were classified as either implicit negative feedback (recasts or negotiation moves) or positive evidence (completion-, translation-, or simple-models).
Assessment tests consisted of an oral picture description task. Iwashita found that the nature of the target form seemed to moderate the type of response provided by native speakers. For example, translation-models were more profuse for te-form verbs than locative-initial constructions. More crucially, the same type of feedback affected the development of each target structure differently. Thus, both implicit negative feedback and positive evidence contributed to short-term gains for te-form verbs, while only positive evidence in the form of simple and completion models was beneficial for locative-initial constructions (albeit for higher level learners only).

Taken together, the studies by Long et al. (1998), Iwashita (2003), and Ayoun (2001), show that the effects of recasts can be moderated by the grammatical structure under instruction. However, they are not informative to the debate on feedback explicitness. In other words, they do not help to elucidate whether recasts are more or less beneficial than other types of more explicit feedback, since the models they opposed recasts to cannot be considered sources of corrective feedback but, rather, preemptive positive evidence.

In contrast, and more interestingly for the purposes of this dissertation, three studies have tested the effects of different types of feedback — with and without rules — on the development of two or more grammatical structures (de Graaff, 1997; Ellis, 2007, and the study reported in Nagata, 1993 and Nagata & Swisher, 1995). Two of these studies provided preliminary metalinguistic presentations in addition to feedback (de Graaff, 1997; Nagata, 1993 and Nagata & Swisher, 1995) while the other excluded it from the design (Ellis, 2007).

In a computerized study, Nagata (1993) and Nagata and Swisher (1995) observed the effects of differently explicit feedback on the development of various constituents of
Japanese passives, including particles and verbs. A total of 32 participants completed a sentence translation task, receiving feedback after errors. In the “traditional CALI feedback” group students were informed of all missing or unexpected forms in their responses, whereas students in the “intelligent CALI feedback” group were provided with further explanations about the nature of their errors. The results of a sentence translation test showed that intelligent CALI outperformed traditional CALI, both immediately and three weeks after the treatment. More interestingly, the advantage of the more explicit feedback arose for the use of Japanese particles rather than for verb conjugations. Based on this, Nagata concluded that traditional feedback may work as well as intelligent feedback for the development of word-level errors (tense, style, conjugation of passive forms), whereas rules may come in handy for sentence-level errors which involve more complex processing (particle errors). In the words of Nagata and Swisher (1995: 341): “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.”

Several caveats, however, urge caution in interpreting these results. First, feedback was not isolated from grammatical explanation prior to practice, so it is impossible to tease out their relative contributions. And second, amount of feedback was not controlled for, so it cannot be determined whether between differences arose as a result of the quantity or quality of the feedback.

Also using a computerized tutor, de Graaff (1997) compared the pedagogical effects of different types of feedback and input on the development of four structures of eXperanto, an artificial language with close resemblance to Spanish morphology and syntax. The structures differed in nature (syntactic, morphological) and/or difficulty (simple,
Following Hulstijn and de Graaff (1994), difficulty was defined as the number of grammatical concepts required to process or produce a structure. The two morphological structures were the inflection of plural nouns and the inflection of imperatives. Plural noun inflection was construed as simple because its realization depended solely on the stressed syllable vowel. Imperative inflection was considered complex because its realization depended on both formality (informal/formal) and affirmativity (affirmative/negative). In turn, the two syntactic structures were the position of negation forms and the position of objects in a sentence. The position of negation forms was hypothesized as simple because word order was only determined by the verb phrase. Conversely, object position was categorized as complex because word order depended on both the type of object (nominal/pronominal) and its weight (stress/topicalization). In this experiment, 54 undergraduate students completed a number of computerized receptive and productive activities during 10 sessions of 1.5 hours. Students in the “explicit” condition received metalinguistic explanations in two ways, as a preliminary grammar presentation and in the form of corrective feedback during practice. Conversely, participants in the “implicit” condition did not receive prior grammatical explanation and got feedback of the right/wrong type, along with the correct response. Assessment tests included sentence judgment, sentence judgment + correction, and controlled written production, and were administered halfway through the treatment, immediately after, and five weeks later. Results showed that when all four structures were taken together, the explicit condition outperformed the implicit condition in all task types at every testing point. However, when the structures were looked at independently, the advantage of the explicit condition arose for only two of them, the simple
morphological structure (plural noun inflection) and the complex syntactic structure (object placement), both immediately after the treatment and five weeks later. This provided mixed evidence to Hulstijn and de Graaff’s (1994) hypothesizing, according to which the advantage of explicit instruction over implicit instruction is more evident for complex than simple structures. As in Nagata (1993), prior metalinguistic explanation and feedback were not isolated as variables. Therefore, caution is urged when extrapolating de Graaff’s findings to the debate on feedback explicitness.

In contrast with Nagata (1993) and de Graaff (1997), Ellis (2007) did not incorporate prior metalinguistic explanation in his research design, where he compared the effects of recasts vs. metalinguistic feedback on the development of two English forms, the regular past tense morpheme (-ed) and the comparative morpheme (-er). He hypothesized that comparatives are more complex because they have a syntactic component (they entail whole clause constructions), are less frequently present in the input, involve later acquired processing operations (as per Pienemann's 1998 Processability Theory), and are governed by a more complex rule (alternation between “more + ADJ” and “ADJ + -er” is subject to the syllabic structure of the adjective). Participants in the study (N = 34) were divided into two experimental groups and a control group that received no instruction. The treatment was administered over two regular classroom sessions for one hour total. To practice the past -ed morpheme, students were divided in triads and completed two story-retelling tasks based upon a set of verbs and pictures. In turn, to practice the comparative, students individually completed two fill-in-the-blank tasks and shared their productions with the class. During task completion, one of the classrooms received partial recasts, i.e., errors were reformulated in isolation (e.g., Student: “Women are kind than
Recasts did not have a differential effect on the acquisition of both structures, while metalinguistic feedback did, but only for ungrammatical items. Specifically, both in the oral imitation and grammaticality judgment test, students got significantly better at identifying a missing morpheme if it was comparative rather than past tense, although some delayed effects were observed for past tense in the oral imitation test. These results, however, must be interpreted with caution due to several limitations. First, students had different amounts of prior knowledge for each structure (there was more room for improvement for comparatives). Second, the tasks used for each structure were substantially different (free vs. controlled production, triad- vs. individual-work). Third,
amount of feedback was not controlled for, with recasts being notably more profuse than metalinguistic feedback for past tense and about the same for comparatives.

**Summary: Grammatical form**

One of the variables that can moderate the effects of a pedagogical intervention is the targeted grammatical form, as shown by a number of empirical studies on form- vs. meaning-focused instruction, input enhancement, or processing instruction. Of special interest for this dissertation are studies addressing how linguistic forms may moderate the effects of differently explicit types of instruction. On this note, there are two main positions. Some scholars contend that the advantages of explicit instruction will arise for easier forms only (e.g., Reber, 1993; Reber et al., 1999), while for others this advantage will precisely surface for harder structures (e.g., Hulstijn & de Graaff, 1994).

The incipient empirical literature on this area of inquiry can be arranged into two main groups. In one block of studies grammatical rules were provided as part of input presentations (Alanen, 1995; Ayoun, 2001; DeKeyser, 1995; Gass et al., 2003; Robinson, 1996; Scott, 1989) while in a second block of studies grammatical rules were embedded in corrective feedback, as in the present dissertation. Currently, it is hard to make any predictions on how the type of grammatical form may moderate the effects of differently explicit types of corrective feedback, given that only three studies have address this (de Graaff, 1997; Ellis, 2007; Nagata & Swisher, 1995) and only one of them (Ellis, 2007) isolated feedback from prior grammatical explanation. Additionally, their findings are mixed. For **simpler forms**, one study (Ellis, 2007) found that more explicit feedback yielded better results, another study found no significant differences (Nagata & Swisher,
1995), and a third study (de Graaff, 1997) contributed mixed evidence, with explicit feedback working better for morphology and not significantly differently for syntax. In turn, for harder forms, both Ellis (2007) and Nagata and Swisher (1995) found that explicit feedback outperformed implicit feedback, while de Graaff (1997) again provided mixed evidence, with explicit feedback working better for syntax and not significantly differently for morphology. These results are summarized in Table 4.2 below.

**Table 4.2 Results for type of Feedback by type of Form**

<table>
<thead>
<tr>
<th>RESULTS OF FEEDBACK BY FORM</th>
<th>SPECIFIC FORM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simpler forms</strong></td>
<td></td>
</tr>
<tr>
<td>- Explicit feedback works better</td>
<td>eXperanto plural noun inflection</td>
</tr>
<tr>
<td>o de Graaff (1997)</td>
<td></td>
</tr>
<tr>
<td>o Ellis (2007)</td>
<td>English past tense morpheme (-ed)</td>
</tr>
<tr>
<td>- No difference</td>
<td></td>
</tr>
<tr>
<td>o de Graaff (1997)</td>
<td>eXperanto position of negative words</td>
</tr>
<tr>
<td><strong>Harder forms</strong></td>
<td></td>
</tr>
<tr>
<td>- Explicit feedback works better</td>
<td>eXperanto position of objects</td>
</tr>
<tr>
<td>o de Graaff (1997)</td>
<td></td>
</tr>
<tr>
<td>o Ellis (2007)</td>
<td>English comparative morpheme (–er)</td>
</tr>
<tr>
<td>- No difference</td>
<td></td>
</tr>
<tr>
<td>o de Graaff (1997)</td>
<td>eXperanto imperative inflection</td>
</tr>
</tbody>
</table>

As for the vaster body of literature on the effects of grammar explanation in the input (as opposed to feedback), the following conclusions can be drawn. For simpler forms (or rules, or linguistic domains — morphology vs. syntax), five studies found an edge for grammar presentation over no presentation (Alanen, 1995; DeKeyser, 1995; Gass et al., 2003; Robinson, 1996; Scott, 1989 — written test), while two studies found no difference
(Ayoun, 2001; Scott, 1989 —oral test). In turn, for harder forms, three studies found an advantage of grammar presentation (Gass et al., 2003; Robinson, 1996; Scott, 1989 —written test), three studies found no difference (Alanen, 1995; Robinson, 1996; Scott, 1989 —oral test), and two studies found an edge for implicit instruction (Ayoun, 2001; DeKeyser, 1995). Again, these results are summarized in Table 4.3 below.
Table 4.3 Results for +/- Grammar presentation by type of Form

<table>
<thead>
<tr>
<th>Simpler structures / rules / domains</th>
<th>Specific Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Explicit works better</td>
<td>Finnish locative suffixes –ssa, -lla</td>
</tr>
<tr>
<td></td>
<td>Consonant reduction kk $\rightarrow$ k, pp $\rightarrow$ p, tt $\rightarrow$ t</td>
</tr>
<tr>
<td></td>
<td>Iplexan categorical rules for noun and verb suffixation</td>
</tr>
<tr>
<td>- Alanen (1995)</td>
<td></td>
</tr>
<tr>
<td>- DeKeyser (1995)</td>
<td></td>
</tr>
<tr>
<td>- Gass et al. (2003)</td>
<td>Italian Pre-verbal direct and indirect object pronouns (in first year students)</td>
</tr>
<tr>
<td>- Scott (1989)</td>
<td>French relative pronouns, written test</td>
</tr>
<tr>
<td>- No difference</td>
<td></td>
</tr>
<tr>
<td>- Ayoun (2001)</td>
<td>French passé composé</td>
</tr>
<tr>
<td>- Scott (1989)</td>
<td>French relative pronouns, oral test</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Harder structures / rules / domains</th>
<th>Specific Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Explicit works better</td>
<td>Sentence-initial question-words in bi-clausal interrogatives (in first year students)</td>
</tr>
<tr>
<td>- Robinson (1996)</td>
<td>Locative pseudoclefts (Instructed &gt; Rule search)</td>
</tr>
<tr>
<td>- Scott (1989)</td>
<td>French subjunctive, written test</td>
</tr>
<tr>
<td>- No difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Consonant voicing t $\rightarrow$ d</td>
</tr>
<tr>
<td>- Robinson (1996)</td>
<td>Locative pseudoclefts (Instructed = Implicit, Incidental)</td>
</tr>
<tr>
<td>- Scott (1989)</td>
<td>French subjunctive, oral test</td>
</tr>
<tr>
<td>- Implicit works better</td>
<td></td>
</tr>
<tr>
<td>- Ayoun (2001)</td>
<td>French imparfait</td>
</tr>
</tbody>
</table>
Overall, then, by merging the findings of the literature on the effects of grammar presentation and type of feedback by type of form, the following main conclusions can be drawn:

(1) For simpler forms, rules, or domains, explicit instruction seems to be more effective than implicit instruction, as shown by seven studies showing an edge for the former (Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; Ellis, 2007; Gass et al., 2003; Robinson, 1996; Scott, 1989 —written test) versus four studies showing no significant difference (Ayoun, 2001; de Graaff, 1997; Nagata & Swisher, 1995; Scott, 1989 —oral test), and no study showing an edge for implicit instruction. Therefore, this seems to lend partial support to Reber’s (1993) claim that the advantages of explicit instruction arise for simpler rules.

(2) For harder forms, rules, or domains, the preliminary picture is a bit more complex, though. Six studies showed an edge for explicit instruction (de Graaff, 1997; Ellis, 2007; Gass et al., 2003; Nagata & Swisher, 1995; Robinson, 1996; Scott, 1989 —written test); four studies yielded no significant differences (Alanen, 1995; de Graaff, 1997; Robinson, 1996; Scott, 1989 —oral test), and two studies showed an advantage of implicit instruction (Ayoun, 2001; DeKeyser, 1995). However, these latter two studies must be interpreted with caution. Ayoun (2001) found that students receiving recasts during practice outperformed students receiving grammar presentation with minimal practice, but this study presents two serious caveats. First, it conflated the timing and explicitness factors (practice with feedback has been proven to yield greater effects than presentation —for the most recent review see Stafford, Bowden, & Sanz, 2011) and second, it did not control for quantity and quality of practice, providing the Grammar group with fewer
items that did not push form-meaning connections (see discussion in section 3.1). In turn, DeKeyser’s (1995) claim of an edge of implicit instruction for prototypical rules was only impressionistic, as no statistical analysis could be conducted due to insufficient data. In conclusion then, when it comes to harder forms, rules, or domains, the balance is tipped in favor of Hulstijn and de Graaff (1994) rather than Reber (1993), with explicit instruction yielding better results, although the smaller ratio of studies supporting this claim and the mentioned presence of counterevidence suggest that the edge of explicit instruction is not as likely to surface as for simpler forms.

Now, a couple of considerations must be made. First, as argued by Ellis (2007: 341), the results of the majority of the empirical studies reviewed above “may simply reflect the fact that the testing instruments only provided measures of explicit knowledge,” and therefore future research should incorporate assessment tasks that also tap on implicit knowledge (by e.g., eliciting timed oral production). And second, caution is urged when extrapolating the conclusions drawn here to future research on the effects of type of form on differently explicit instruction. Specifically, forms were classified here as “simpler” or “harder” (as opposed to “simple” or “hard”) based on their relative difficulty against the structures they were compared to by the researchers, but there is no attempt to make holistic claims or present an absolutistic taxonomy of forms in terms of difficulty. Only extensive research can achieve this ambitious goal, if ever accomplishable. The value of this set of conclusions lies in that they reflect how the same pedagogical materials worked on a very controlled population of learners depending on whether or not they included metalinguistic rules, and hence they minimize the moderating role of other
variables (e.g., type of task, type of learners) to a greater extent than comparisons based on different studies addressing only one type of targeted form.

4. Research questions

In light of the gaps and conflicting findings of the literature reviewed above, the present dissertation investigates the role of three independent variables on L2 grammar development, i.e., type of Agency (the compared effects of Practice versus Exposure to practice), type of Feedback, and type of grammatical Form. Below is the motivation for each variable and each specific research question.

Time x Agency. As argued by, e.g., Nagata (1996), the potential edge of computers against other instructional technologies may lie in their ability to facilitate interactive practice, which, according to Long’s (1996) Interaction Hypothesis, can promote SLA because it allows learners to comprehend and notice input, produce output, receive feedback, revise hypotheses, and produce modified output, all of which have been shown to be developmentally helpful processes (Mackey, 2007). However, few claims have been made as to the role of learner’s Agency in interactive practice, or, in other words, whether learners must practice proactively in order to benefit from interaction or whether mere exposure to practice by others may suffice. Existing empirical research is scant and contradictory, with one study showing a benefit for practice (Mackey, 1999), one study yielding no significant differences between practice and exposure (Muranoi, 2000), and one study yielding mixed evidence, with a benefit for practice in production tests only
(Hsieh, 2007). Given this, the present dissertation seeks to answer the following research question:

- **RQ1.** Does type of learner’s Agency in interactive practice (i.e., performing interactive practice versus observing others practice) have differential effects on L2 grammar development, as measured by three tests (written production, oral production, and grammaticality judgment), immediately after the treatment and two weeks later?

  **Time x Feedback.** Existing literature on feedback, both in oral interaction and CALI shows conflicting findings as to the relative efficacy of different types of feedback, for example, less explicit feedback (e.g., feedback indicating that an error has been made) vs. more explicit feedback (feedback explaining why an error occurred). While these divergences may be due to uncontrolled intervening variables, such as type of linguistic form, type of practice, learning environment, and individual differences, among others, they may be also rooted in the fact that there are great differences in the way researchers define, operationalize, and code feedback moves (e.g., Nassaji, 2007). Specifically, most researchers have failed to isolate the components of feedback moves, i.e., the provision of negative evidence, which can be less or more explicit; the provision of positive evidence, and prompting for error repair. Additionally, existing studies on computerized feedback have only explored written feedback, neglecting the potential of computers to simulate real-life communication (Chapelle, 2004; Underwood, 1984). To address this, the present dissertation isolates two components, negative evidence and prompting, in an attempt to answer the following question:
RQ2. Does type of oral Feedback (with different levels of negative evidence and prompting) have differential effects on L2 grammar development, as measured by three tests (written production, oral production, and grammaticality judgment), immediately after the treatment and two weeks later? If so, which of two particular subcomponents of feedback, explicitness of negative evidence and prompting for repair, contribute to the observed differences?

Time x Form. The effects of linguistic form have been found to mediate the efficacy of incidental learning through exposure, deliberate grammar practice, and feedback. Currently, it is hard to make any predictions on how the type of grammatical form may moderate the effects of differently explicit types of corrective feedback, given that only three studies have address this (de Graaff, 1997; Ellis, 2007; Nagata & Swisher, 1995) and only one of them (Ellis, 2007) isolated feedback from prior grammatical explanation. Additionally, their findings are mixed. Nagata and Swisher’s (1995) study showed an advantage of explicit feedback for easier forms only, in line with, e.g., Reber’s (1993) theorizing; Ellis’s (2007) study showed an advantage of explicit feedback for both easier and harder forms, thus also lending partial support to Hulstijn and de Graaff’s (1994) claim that the edge of explicit instruction surfaces for harder structures. Finally, de Graaff’s (1997) study provided mixed evidence. In this context, the following research question is formulated:

RQ3. Does type of Form (more or less complex) have differential effects on L2 development resulting from a pedagogical intervention, as measured by three tests
(written production, oral production, and grammaticality judgment), immediately after the treatment and two weeks later?

Interaction Time x Form x Agency x Feedback. Finally, the last research question addresses the interaction between the three variables under investigation. Specifically:

- RQ4. Do type of Agency, type of Feedback, and type of Form work in concert towards different learning outcomes, as measured by three tests (written production, oral production, and grammaticality judgment), immediately after the treatment and two weeks later? If so, how do these variables interact?

  (a) Time x Agency x Feedback: Does type of oral Feedback moderate the learning effects of type of learner’s Agency in grammar practice (or vice versa)?

  (b) Time x Form x Agency: Does type of linguistic Form moderate the learning effects of type of learner’s Agency in grammar practice?

  (c) Time x Form x Feedback: Does type of linguistic Form moderate the learning effects of type of oral Feedback?

  (d) Time x Form x Agency x Feedback: Does type of linguistic Form moderate the learning effects of a pedagogical treatment with different types of learners’ Agency and oral Feedback?
CHAPTER 5.
THE EMPIRICAL STUDY

The present chapter is divided into six sections. Section 1 provides information about the targeted pool of participants and the experimental groups. Section 2 discusses the targeted forms under investigation. Section 3 provides details about the experimental materials, including the CALI tutor that was developed for this dissertation, the treatment practice, the treatment feedback, and the assessment tests. Section 4 explains data collection procedures. Section 5 lays out the coding and scoring protocols. Finally, section 6 discusses the proposed procedures for statistical analysis.

1. Participants and experimental groups

Participants in this study consisted of university undergraduate students of intermediate Spanish I and II at Georgetown University, randomly assigned to 1 of 8 groups, as per the combination of two between-subject variables, i.e., Agency and Feedback. Agency had two levels, Practice and Exposure to practice. Practice was operationalized as the completion of an output-focused task, while Exposure to practice entailed witnessing the completion of the task by a peer. The Feedback variable had five levels: No feedback, and four feedback groups, as per the interaction of two feedback components with two levels each, i.e., negative evidence with lesser or greater explicitness (Spot error vs. Explain error) and prompting for error repair or absence thereof (Prompt vs. Continue). In principle, the combination of the Agency and Feedback variables required data collection from ten groups (2 levels of Agency x 5 levels of Feedback). However, due to
restricted availability of participants, the two grayed out cells in Figure 5.1 below were
left out. Consequently, there were 8 experimental groups: Practice with five types of
feedback, i.e., Spot & Continue [Practice SC], Spot & Prompt [Practice SP], Explain &
Continue [Practice EC], Explain & Prompt [Practice EP], and No Feedback [Practice
NF], and Exposure to practice with both Spot and Explain Feedback with Prompts
[Exposure SP], [Exposure EP], and No Feedback [Exposure NF].

**Figure 5.1** Experimental groups

<table>
<thead>
<tr>
<th>[Practice SC]</th>
<th>[Practice SP]</th>
<th>[Practice EC]</th>
<th>[Practice EP]</th>
<th>[Practice NF]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Exposure SP]</td>
<td>[Exposure EP]</td>
<td>[Exposure EP]</td>
<td>[Exposure NF]</td>
<td></td>
</tr>
</tbody>
</table>

The exclusion of the grayed out groups was based on theoretical grounds. Arguably, a
student who witnesses the interaction of another with a tutor as well as his/her attempts at
repairing errors after corrective feedback is exposed to richer interaction episodes than a
student who watches a practicing participant who is deprived from the opportunity to
repair. Thus, if differences between the Practice and Exposure to practice conditions were
found in the Prompt groups, they were expected to be at least equal, if not larger, in the
Continue groups. This of course is an empirical question that must be addressed by
further research.

In addition to the two between-subject variables just mentioned, Agency and
Feedback, this study was designed to investigate the effects of a within-subject variable,
linguistic Form. Specifically, all participants were exposed to two grammatical forms,
Spanish Prepositional relative clauses and Spanish Present subjunctive in adjectival relative clauses (see Section 2).

The initial pool of participants in this study consisted of 218 students. After the pretest session, 73 students were excluded because they either reported prior formal study in a romance language other than Spanish for 2 semesters or longer or because they scored more than 3 points (out of 10) for at least one of the two structures on the written production pretest (see Section 3.4). Out of the 145 remaining students, 17 failed to complete all three experimental sessions. Additionally, one student reported having consulted both grammatical structures outside the laboratory setting at some point and provided the appropriate grammatical rules, so she was excluded from the analysis. As a result, the final pool of students comprised 127 students. However, the number of participants per assessment test (i.e., written production, oral production, grammaticality judgment) ranged from 121 to 126 for several reasons, including external exposure to one of the two targeted structures, as appropriately reported on the debriefing questionnaire, data loss (some oral test files could not be retrieved or were inaudible), and technology problems (the CALI tutor crashed for the Exposure to practice groups in some cases). Out of the 127 participants, 77 were students of intermediate Spanish I and the other 50 were enrolled in intermediate Spanish II. Age ranged between 18 and 25 years old, and the average was 21.3. Participants were compensated with 5 extra points on a test.

Procedures for subject recruitment and compensation were approved by Georgetown University’s IRB (2009-595).
2. Targeted forms

The present study targeted two linguistic forms with allegedly different degrees of
difficulty, Spanish Prepositional relative clauses (e.g., \textit{La persona CON LA QUE VIVO se
llama Pepa} [“The person WHO I LIVE WITH is called Pepa”]) and Present subjunctive
in relative clauses (e.g., \textit{Quiero un apartamento QUE TENGA tres habitaciones} [“I
would like an apartment THAT HAS-SUBJ three bedrooms”]). Sections 2.1 - 2.3 below
discuss these structures and the difficulties they pose for L1 English speakers.

2.1 Spanish prepositional relative clauses

In English, prepositional relative clauses can surface as two main structures, as in (1) and
(2) below:

(1) The person with who(m) I live is named Pepa.

(2) The person (who) I live with is named Pepa.

Generative syntax explains the structure in (1) as the result of a phenomenon known as
\textit{preposition pied-piping}, whereby an entire prepositional phrase moves to the front of the
relative clause. For example, in (3) below, the prepositional phrase \textit{with who(m)},
consisting of the preposition \textit{with} and the relative pronoun \textit{who}, moves to the Spec of the
Complementizer (CP), leaving a trace ($t_i$):

(3) The person $[\text{CP} [\text{with who(m)}]_i [\text{C} \text{ I live } t_i]]$ is named Pepa.
On the other hand, the structure in (2) is explained by a syntactic process known as *preposition stranding*, whereby only the relative pronoun, *who*, is extracted to the beginning of the relative clause, leaving the preposition, *with*, stranded at the end of the clause, as in (4).

(4) The person $[_{CP} \text{who}, [_{C'} \text{I live with } t_i]]$ is named Pepa.

In English, both surface structures, relative clauses with *preposition pied-piping* and *preposition stranding*, are grammatical, although their use is determined by several factors. Specifically, preposition stranding is the preferred choice in informal oral speech, whereas pied-piping is often the preferred construction in more formal settings, although dialectal preferences also apply. As opposed to English, in Spanish only one of these structures, preposition pied-piping, is possible (cf. Campos, 1991, for an interesting squib on counter-arguments). This is illustrated by the grammatical and ungrammatical exemplars in (5) and (6) below:

(5) La persona $[\text{con la que / con la cual / con quien}_i \text{ vivo } t_i]$ se llama Pepa.

(6) *La persona $[\text{la que / la cual / quien}_i \text{ vivo con } t_i]$ se llama Pepa.

As discussed in Chapter 3, according to DeKeyser (2005) the relative difficulty of a grammatical structure from an acquisitional perspective can be assessed in terms of the problems it poses for the acquisition of its *meaning* and *form*, as well as the mapping between these two. In terms of *meaning*, understanding an adjectival relative clause,
prepositional or not, entails understanding the predication relation that exists between an antecedent and the information encapsulated in the relative clause. In this regard, since English has relative clauses too, L1 English speakers should not encounter many problems to understand the meaning of its Spanish counterparts. In terms of form, however, L1 English speakers may face two main obstacles. First, they must know the repertoire of Spanish relative pronouns (que, quien, cual) and how they operate to build relative clauses, and second, they may have a conflict when deciding what syntactic phenomenon, either preposition stranding or pied-piping, applies. The first problem is not relevant in this dissertation because all participants already knew how to formulate Spanish non-prepositional relative clauses at the outset of the study. The second problem, however, is directly related to the object of this research.

According to Perpiñán (2008), L1 English speakers may have trouble acquiring the syntax of Spanish prepositional relative clauses because there is a conflict between the so-called Subset Principle and the Transfer Hypothesis. The first tenant of the Subset Principle is that grammatical structures can relate to each other in terms of a subset/superset relation. Thus, given that all languages that present preposition stranding have preposition pied-piping but the contrary is not necessarily true, preposition pied-piping can be considered a subset of preposition stranding. By the same token, English can be considered a superset language with respect to Spanish, because it accepts both preposition-stranding and pied-piping, whereas Spanish only accepts the latter. Now, according to the Subset Principle, acquisition starts with the narrowest grammar possible and may develop into a larger grammar as a result of exposure to positive evidence. However, this is counterproductive for L1 English speakers learning Spanish because
they already possess the bigger grammar and cannot hypothesize the smaller one. In other words, English speakers will transfer their superset grammar and assume that both pied-piping and preposition-stranding are grammatically correct in Spanish, and no amount of positive evidence can make them reformulate this hypothesis. Here is precisely where negative evidence, or corrective feedback, may have special relevance, as posited by White’s (1989; 1991) seminal work, because learners need to be told that their hypothesis is incorrect. Consequently, since the present dissertation examines the role of corrective feedback in adult L2 acquisition from L1 English to L2 Spanish, prepositional relative clauses constituted an ideal targeted structure.

2.2 Spanish present subjunctive in adjectival relative clauses

Arguably, the acquisition of the Spanish modality, or the distinction between indicative and subjunctive moods, is one of the most challenging aspects for L1 English students. According to Timberlake (2007: 315):

> [m]odality is about alternatives — how we come to know and speak about the world, how the world came to be as it is, whether it might be other than it is, what needs to be done to the world to make it what we want. The alternatives are sorted out and evaluated by some sort of authority, often the speaker, or, if not the speaker, some other participant or even another situation. Modality, then, is consideration of alternative realities mediated by an authority.

Traditionally, grammarians have distinguished between two different moods. Realis is the mood that refers to actual events, whereas irrealis refers to non-actual events. In Spanish, realis and irrealis are basically equivalent to indicative and subjunctive, respectively. Now, as Chung and Timberlake (1985: 241) put it, “[w]hereas there is basically one way for an event to be actual, there are numerous ways that an event can be
less than completely actual.” For that reason discussions on modality largely focus on the irrealis/subjunctive mood.

From a semantic perspective, in Spanish the use of subjunctive seems to be regulated by at least three systems, deontic, epistemic, and epistemological (S. Chung & Timberlake, 1985). **Deontic** modality characterizes an event as non-actual because it is imposed. In Spanish, the use of subjunctive to express deontic modality can be seen in negative commands (e.g., *No fumes* [“Do not smoke”]) and indirect commands, i.e., those expressed through a complement clause headed by a predicate expressing imposition (e.g., *Te prohíbo que fumes* [“I forbid you to smoke”]). Often, two subtypes of deontic moods are distinguished, *permission* (e.g., *Te dejo que fumes* [“I allow you to smoke”]) and *obligation* (e.g., *Te ordeno que fumes* [“I command you to smoke”]).

**Epistemic** modality, on the other hand, characterizes an event with respect to the actual world and its possible alternatives. Here too authors distinguish between two types, *necessity* (when the event belongs to all alternative worlds) and *possibility* (when the event belongs to one alternative world). In Spanish, epistemic modality expressing possibility also triggers the use of subjunctive in complement clauses (e.g., *Puede que llueva mañana* [“It may rain tomorrow”]).

Interestingly, there is a close parallelism between the subtypes of the deontic mood (permission and obligation) and those of the epistemic mood (necessity and possibility). In fact, some English modal verbs can often have both readings, as illustrated by Table 5.1 below.
Table 5.1 Deontic vs. epistemic mood

<table>
<thead>
<tr>
<th></th>
<th>Deontic Mood</th>
<th>Epistemic Mood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permission / Possibility</td>
<td>You may stay</td>
<td>Pepe may be Spanish.</td>
</tr>
<tr>
<td>Obligation / Necessity</td>
<td>You must stay</td>
<td>Pepe must be Spanish.</td>
</tr>
</tbody>
</table>

As Table 5.1 shows, the English modal *may* can have a deontic reading when it invokes the sense of a source giving permission (*You may stay* = *You are allowed to stay*) and an epistemic reading when it expresses the possibility that there is at least one alternative world where the event occurs (*Pepa may be Spanish* = *There is a possibility that Pepa is Spanish, although she could be Italian too*). Similarly, *must* can have a deontic reading when it expresses obligation (*You must stay* = *You have to stay*) and an epistemic reading when the event must hold in all alternative worlds (*Pepa must be Spanish* = *Based on her accent, her traits, etc. I cannot think of Pepa as being a citizen of a country other than Spain*).

Finally, besides the deontic and epistemic moods Spanish boasts a third mood, so-called *epistemological*, whereby the actuality of an event is evaluated with respect to the source orparticipant target. The use of Spanish subjunctive to express epistemological modality can be seen in complement clauses to factive emotive verbs such as *alegrarse* (e.g., *Me alegro de que estés aquí* [“I am glad you are here”]).

Using this tripartite mood framework, Pérez-Leroux (1998: 586) stated that “[i]n adjunct clauses, such as relative clauses and temporal clauses, one may argue that mood selection is governed by the semantics of epistemic modality” because they involve the evaluation of an event with respect to the actual world and its possible alternatives.
Spanish, relative clauses can trigger the use of indicative or subjunctive, depending on whether the event is respectively perceived as actual in all or only some possible worlds, or, alternatively, from a different perspective, whether the speaker points to an individual or a class, as in (1) and (2) below.

(1) Busco a una persona que habla español
   [“I’m looking for a person who speaks-IND Spanish”]
(2) Busco a una persona que hable español
   [“I’m looking for a person who speaks-SUBJ Spanish”]

Noticeably, both (1) and (2) read the same way in English; however, the alternating use of present indicative or subjunctive in Spanish renders very different meanings. In (1) the speaker is aware of the existence of a specific person who speaks Spanish in that particular place and is asking for directions —in semantic terms, the speaker is selecting an individual or perceives the event as actual in all possible worlds. In (2) however, the speaker is asking the hearer whether s/he knows of anyone who speaks Spanish; that is, s/he is referring to a class.

As shown by (1) and (2), the use of subjunctive in Spanish relative clauses is not determined exclusively by distributional associations (i.e., they both co-occur with the exact same antecedent). This differs from other uses of the subjunctive, where deontic verbs (e.g., querer, ordenar, prohibir) and purpose clauses with prepositions such as para trigger the obligatory use of the subjunctive in the complement clause —hence the ungrammaticality of (3) and (4):
(3) *Te ordeno que vienes.
   [“I command you to come-IND”]
(4) *Trabajo duro para que puedas comer.
   [“I work hard so that you can eat”]

In contrast, the subjunctive in relative clauses may or may not be determined by grammatical constraints. As Pérez-Leroux (2001: 78) puts it, indefinite NPs can co-occur with both indicative and subjunctive, as in (5) and (6). Definite NPs, however, do typically preclude subjunctive (7), and hence examples such as (8) are usually rejected by speakers as odd or ungrammatical.

(5) Indefinite/Indicative

Busco una bicicleta que tiene rueditas.
[“I’m looking for a bicycle that has-IND little wheels”]

(6) Indefinite/Subjunctive

Busco una bicicleta que tenga rueditas.
[“I’m looking for a bicycle that has-SUBJ little wheels”]

(7) Definite/Indicative

Busco la bicicleta que tiene rueditas.
[“I’m looking for the bicycle that has-IND little wheels”]

(8) Definite/Subjunctive

*Busco la bicicleta que tenga rueditas.
[“I’m looking for the bicycle that has-SUBJ little wheels”]
Now, from the perspective of first language acquisition, Pérez-Leroux (1998: 591) notes that there is a growing body of empirical research showing that the acquisition of Spanish subjunctive seems to follow a developmental timetable. Studies show that children first master the deontic modality (i.e., negative and indirect commands), continue with the epistemic modality (e.g., relative and adverbial clauses), and finally acquire the epistemological modality (e.g., attitude-based complement clauses). For example, Pérez-Leroux (1993) showed that children who could use the subjunctive with purpose clauses (e.g., para que vengas [“so that you come-SUBJ”]) were still unable to use it with relative clauses. In contrast with L1 acquisition literature, very few efforts have been put into finding out whether there is also a developmental sequence in the acquisition of Spanish as a second language. Curiously, though, in line with the previous discussion, most form-oriented curricula typically present mood alternation in relative clauses after deontic uses of the subjunctive.

According to DeKeyser’s (2005) criteria, L1 English speakers should not encounter many problems of meaning when acquiring Spanish relative clauses because English already possesses such construction. In terms of form, however, the acquisition of relative clauses poses two main problems. First, learners must know the repertoire of Spanish relative pronouns (i.e., que, quien, cual) and how to use them to build relative clauses (e.g., Spanish is not pro-drop, pronoun selection is subject to the human/object nature of the antecedent, etc.) and second, learners must master the morphology of the present indicative and subjunctive, which are formally very similar. Finally, perhaps the most challenging aspect for acquisition purposes is related to problems of form-meaning.
mapping. As explained earlier, the indicative/subjunctive alternation may be determined by syntactic constraints (e.g., definite NPs trigger indicative) or by semantic (and pragmatic) factors, and this alternation is not formally mirrored in English. As a result, even if they notice the alternation learners may think that the indicative and subjunctive co-occur in free distribution, i.e., that their use is random or optional. Due to all these potential problems this structure was an ideal candidate for the purposes of this dissertation. Problems of form were solved by conducting the experiment after participants had learned the verbal morphology and by exposing them to a morphology review session before the pretest. On the other hand, problems of form-meaning mapping were precisely the object of the treatment, where students were pushed to reformulate their hypotheses by being engaged in the completion of an output-focused task with concurrent feedback.

2.3 Present subjunctive vs. Prepositional phrases in adjectival relative clauses

As discussed in sections 2.1 and 2.2, the two structures in this study pose different problems to English L1 speakers. Both structures pose problems of form, as they presuppose knowledge of (1) the repertoire of relative pronouns, (2) how these are used to build non-prepositional relative clauses, and (3) the morphology of present indicative and subjunctive. In this dissertation these problems were overcome by administering the treatments only after students knew how to build non-prepositional relative clauses and had formally studied the present indicative/subjunctive morphology, both in the classroom and via a preliminary review session in the laboratory.
Now, acquiring Spanish Present subjunctive (vs. indicative) in relative clauses poses problems of form-meaning mapping because learners have to correctly infer the semantic/pragmatic information of the utterance (i.e., whether the event is actual in all or some possible worlds). Conversely, Spanish prepositional relative clauses pose problems of form because, first, English is a superset language in relation to Spanish and students may hypothesize that both preposition stranding and pied-piping are grammatical in Spanish, and second, the most frequent structure in English is actually preposition stranding, which is ungrammatical in Spanish.

According to DeKeyser (2005), formal aspects such as word order are typically nonproblematic past the initial stages of acquisition. Conversely, understanding form-meaning mappings in structures that invoke abstract meanings is arguably more challenging. For this reason it is posited here that for L1 English students of Spanish the acquisition of Present subjunctive in relative clauses will be more challenging than the acquisition of Prepositional relative clauses.

3. Materials

The next four sections describe the pedagogical and testing materials used in this study. Section 3.1 explains the nature of the CALI tutor developed for this dissertation. Section 3.2 discusses the type of practice used in the experimental treatment. Section 3.3 focuses on the different types of feedback provided during task completion. Section 3.4 describes the tests used to assess learning achievements.
3.1 CALI tutor

The experimental treatment in this study was presented via *Talking to Avatars*, a computerized tutor that uses audiovisual recordings to simulate a conversation between the learner and a series of “avatars” or pre-filmed human actors. Following is a description of its components.

*The video recordings.* By using Talking to Avatars, learners were immersed in two real-life situations (finding an apartment and a roommate, reporting a theft) during a fictional year-abroad program in Spain. To simulate these interactions, a total of 217 monologues were scripted and video-recorded (see Appendices 5.1 A and B for the complete scripts). In these videos, the avatars looked directly into the camera, the student’s point of view, asking a question or providing feedback in several ways, as appropriate (more details in section 3.3). Since it was of utmost importance that the avatars delivered the scripted questions and feedback messages verbatim, they read their lines off autocues. These autocues were created with Microsoft PowerPoint and were projected onto an HD flat-panel TV that was strategically placed in front of the actors and underneath the camera (see Figure 5.2 below). To simulate continuity between the video transitions, the camera was mounted on a tripod and remained in the same location throughout the entire shooting. Also, special care was taken for lighting and background noise. After the shooting, all videos were captured onto a PC computer using Adobe Premiere Pro CS4. Each video was cut and labeled with a distinctive filename, including the scene number and the nature of the feedback (positive/negative; Spot/Explain; Prompt/Continue). For example, the label “B-9-XC-KO” was used for the video that would be played after any student in the eXplain & Continue feedback group (XC)
produced an incorrect answer (KO) during item 9 of communicative situation B. Finally, in the post-production phase all videos were compressed from the standard AVI format (Audio Video Interleave) into FLV format (Flash Video) to facilitate faster download over the Internet. This file conversion process was done using eRightSoft Super, with a video compression rate of 25 frames per second, an audio sampling frequency of 22050 Hz in mp3 format, and a bitrate of 64 kbps per second.

**Figure 5.2** Avatar reading his lines off autocues during a video shooting session
The software component. To facilitate data access, storage, and sharing, Talking to Avatars was developed as a web-based application, using a combination of PHP and HTML programming codes. The application consists of two components, a static component that contains the program code, and a dynamic component that can be edited by the administrator. There are two different gateways to access the program, one for the students and one for the administrator. Both gateways can be accessed by visiting http://www.talkingtoavatars.com and entering a password. The administrator gateway grants access to two different databases, the Questions database and the Answers database. The Question database contains all the necessary information to be displayed to each student, including which video should be played at each point, the specific task to be completed by the student, a catalog of correct answers, discourse strategies on how to proceed depending on the student’s answer, and some extras (e.g., transcripts, translations). This information can be edited to fine-tune or expand the program. In turn, the Answers database contains a log of the students’ responses, a search engine for selective display of collected data, and an application that generates transcripts of
individual sessions in Adobe Acrobat PDF format. Finally, to analyze the student’s output, Talking to Avatars uses a string-matching algorithm with different specifications for every targeted structure (more details in section 3.3).

3.2 Treatment practice

By using Talking to Avatars participants in this study interacted with prefilmed human actors in two real-life situations (Finding an apartment and a roommate; Reporting a theft) while on a simulated year-abroad experience in Málaga, Spain. Each situation was designed to instruct one targeted form at a time and consisted of an introductory presentation and 15 items or mini-episodes, with 10 of them eliciting the targeted form and the remaining 5 eliciting the non-targeted counterparts (i.e., Present subjunctive vs. indicative in relative clauses; Prepositional vs. non-prepositional relative clauses). In every mini-episode, participants in the Practice condition were orally addressed by an avatar that was looking directly into the camera, the student’s point of view, thus initiating an interaction sequence in four steps, as follows:

1. “Question”: the avatar asked a question to elicit information on the part of the student (e.g., Tengo varios apartamentos… ¿Cuánto dinero puedes pagar? [“I have several apartments… How much are you able to spend?”]);
2. “Activity”: a fill-in-the-blank exercise was displayed on the screen to help the student provide the requested information in a task-essential manner. Specifically, the student was asked to provide a written translation of the targeted form from English into Spanish and to supply the appropriate content information by either
filling in a blank or selecting an option from a drop-down list (e.g., *Busco un apartamento (THAT COSTS) ________ (menos de 500 | de 500 a 1000 | más de 1000) euros al mes.* [“I’m looking for an apartment (THAT COSTS) ________ (less than 500 | from 500 to 1000 | more than 1000) euros a month”]);

(3) “Answer”: the student then provided the requested information, producing an answer (e.g., *Busco un apartamento que *cuesta menos de 500 euros al mes [“I am looking for an apartment that *costs-IND less than 500 euros a month”]);

(4) “Feedback”: the avatar replied with feedback on the content and form, depending on the student’s experimental group (e.g., *Bueno, hay que ahorrar dinero... Una cosita: el verbo COSTAR no es correcto. ¿Puedes corregirlo?* [“Sure, you should try to save some money… One thing, though: the verb COSTAR is not correct. Could you please correct it?”]).

Following is a description of the two communicative situations at play (Finding an apartment and a roommate; Reporting a theft). For the complete script of each situation, see Appendices 5.1 A and B.

*Communicative situation A: Finding an apartment and a roommate.* The participant, an American college student, has just arrived in Málaga, Spain, and needs to find an apartment for the duration of the academic year. S/he visits the local University’s Student Accommodation Services and interacts with a member of the staff to describe the type of accommodation and roommates s/he is ideally looking for. This situation elicits the Present subjunctive (vs. Present indicative) in relative clauses. Figure 5.3 provides an example of a communicative episode.
Figure 5.3 Communicative episode in situation A (Finding an apartment and a roommate)

**AVATAR**  
*Tengo varios apartamentos… ¿Cuánto dinero puedes pagar?*  
[“I have several apartments… How much are you able to spend?”]

**STUDENT**  
*Busco un apartamento que *cuesta menos de 500 euros al mes*  
[“I am looking for an apartment that *costs-IND less than 500 euros a month”]

*Communicative situation B: Reporting a theft.* The student has been robbed in his own apartment, so s/he goes to the police station to file a report. Throughout the process, s/he interacts with a police officer to reconstruct the sequence of events that occurred prior to the theft, in an attempt to discover the identity of the mysterious thief. This situation elicits Prepositional (vs. non-prepositional) relative clauses. Figure 5.4 below provides a sample mini-episode.
Figure 5.4 Communicative episode in situation B (Reporting a theft)

AVATAR Bueno, volviendo a tu compañero de apartamento, ¿qué es lo que hace? ¿Estudia o trabaja?

[Alright, returning to your roommate, what does s/he do? Does s/he work or go to school?]

STUDENT La persona con la que vivo es estudiante.

[“The person I live with is a student.”]

To ensure task-essentialness (Loschky & Bley-Vroman, 1993) or, in other words, trigger the production of the targeted forms, the experimental activity was based on a cloze-sentence structure. Specifically, after listening to the question posed by the avatar, the student was presented with (1) the rubric or instructions of what to do next, (2) a vocabulary list with the key words to be used (for situation A, the corresponding Spanish verb in infinitive form; for situation B, a verb and the Spanish preposition at play); (3) when appropriate, a note on the stem-changing nature of some verbs; and (4) a Spanish sentence with a section blanked out and the English segment to be translated into Spanish.
(the targeted form or its non-targeted counterpart) and, optionally, a scroll-drown menu with options to select from (the content part). Below is an example of a targeted item.

**AVATAR VIDEO**

Vale, comencemos por el apartamento. ¿En qué área de la ciudad te gustaría vivir? ¿En la playa, en el centro de la ciudad o en el campus universitario?

[“Okay, let’s begin with the apartment. Which part of the city would you like to live in? On the beach, downtown, or on the university’s campus?”]

**POP-UP TEXT**

Please answer Pepa’s question by filling in the box and selecting your preferred choice.

HINT: = “TO BE” = “estar”

Quiero un apartamento (THAT IS) ______________ en (la playa / el centro / el campus)

[“I would like an apartment (THAT IS) ______________ (on the beach / downtown / on campus)”]

Arguably, this learning activity has a hybrid nature because it shares qualities that are characteristic of both tasks and exercises. From a structural point of view, the activity is composed by a series of concatenated drills. According to Paulston’s (1970) classification, these drills could be labeled as “semi-meaningful” because although the response is totally controlled, the learner has to understand the input in order to produce the correct L2 form. Additionally, these drills are interconnected, contributing to create a discourse that is oriented towards a real goal, i.e., finding an apartment and a roommate, and filing a police report. Following Ellis’s (2003) taxonomy of task-criterial features, then, this activity could be defined as follows:

(1) Is there a workplan? Yes, finding an apartment, a roommate, and the money that the student is missing.
(2) Is the focus on meaning? Yes, in a large number of items the student has to express his/her own preferences about something (e.g., the preferred location for his/her apartment, whether s/he wants a smoking or non-smoking roommate, his/her hypothesis about the identity of the potential thief).

(3) Does the learner choose the forms? No, the student is given the L1 forms that s/he has to translate into the L2.

(4) Is there situational or interactional authenticity? Yes, there is situational and interactional authenticity because the learner is involved in a real-life situation and has to understand the input, produce output, process feedback, and, in some cases, modify his/her own output to move forward in the discourse.

(5) Are extra-linguistic cognitive processes required? Yes, decision-making, because the student has to make decisions to select his/her preferred choices.

(6) Is there a non-linguistic outcome? Yes, at the end of the activity the student will have found an apartment, a roommate, and the money s/he was missing.

(7) Which language skills are involved? Listening comprehension, reading comprehension, and written production.

In sum, then, the experimental activity in this dissertation can be considered as an exercise that boasts many of the properties that are characteristic of tasks. It is an exercise because the rubric specifies the language forms that the student has to use; however, it is based on a workplan, it involves a focused balance on form and meaning, it has situational and interactional authenticity, it triggers extra-linguistic cognitive processes, and it is oriented towards the achievement of a non-linguistic outcome.
In terms of elicited skills, the interaction between avatars and students is based on a combination of the aural and written modes, due to computational constraints. In the spirit of creating authentic CALI tutors for meaningful communication (e.g., Garrett, 1991; Lavine, 1992; Quinn, 1990; Underwood, 1993) Talking to Avatars aimed to simulate entirely aural interactions. However, the speech recognition technology necessary for that purpose is not yet available. Hitherto, researchers and developers have mostly targeted L1 speakers, and only recently have they started to address the many problems posed by L2 speakers’ speech (see, e.g., Swerts & Paliwal’s in-progress special issue of *Speech Communication* journal: “Non-native speech perception in adverse conditions: imperfect knowledge, imperfect signal”). For that reason, Talking to Avatars adopted a hybrid approach using both the aural and written modes to create authentic pedagogical materials that are computationally possible. Specifically, all interactions from the avatars to the learners occurred in the aural mode. To ensure that learners had no problems understanding the meaning of the avatars’ questions, several options were made available. Specifically, learners could (1) press the Replay button and watch the video again; (2) press the Transcribe button and retrieve a transcript of the avatar’s Spanish utterance, or (3) press the Translate button and get an English translation. Conversely, all interactions from the learners to the avatars were conducted in the written mode. After listening to an avatar question, learners were instructed to type in their answers. Figure 5.5 below provides a screenshot of the software interface in Talking to Avatars.
As opposed to students in the Practice condition, the job of the students in the Exposure to practice condition consisted of watching the interaction mini-episodes as held by the avatar and a practicing peer. In previous related studies, the operational definitions of the Exposure to practice condition have varied. For example, in order to ensure that participants paid attention to the “interactors,” Mackey (1999) asked them to complete the same tasks by drawing pictures or using their L1 instead of the L2. In turn, Murano (2000) did not impose any requirements on his observers other than watching the interactions and acting as consultants if the “performers” needed help, although no
specifics were reported as to whether they eventually acted as such and how frequently. Finally, Hsieh (2007) asked her no-practice participants to think aloud while observing the interactions. Arguably, all these operationalizations have their pros and cons. On the one hand, forcing students to pay attention may subvert the nature of the Exposure to practice condition, where ideally students should be left to their own devices. Also, directing their attention to meaning or form can have differential effects, as shown by a growing body of empirical research initiated by VanPatten (1989). On the other hand, if attention is not controlled for it is impossible to determine whether any observed differences between Practice and Exposure to practice are due to the nature of learner’s Agency in the interaction or differing levels of attention. Finally, asking students to think aloud may throw in another variable, reactivity, which is currently the focus of much research in applied linguistics. In the present dissertation it was decided that in order to be able to make claims about the effects of Agency in interaction attention had to be controlled for. Specifically, participants in the Exposure to practice condition were asked to observe every interaction mini-episode as it occurred and they were periodically asked check-up questions about their peers’ responses. For example, after a student in the Practice condition wrote *Quiero un apartamento que esté en el campus* (“I want an apartment that is-SUBJ on campus”), the observer in the Exposure to practice condition was asked “What is NAME’s preferred location?” where NAME was the actual name of the participant in the Practice condition. Table 5.2 below includes the list of check-up questions and possible answers for each communicative situation, i.e., (A) Finding an apartment and a roommate and (B) Reporting a theft, which respectively elicited the
Present subjunctive versus indicative and Prepositional versus non-prepositional relative clauses.

**Table 5.2** List of check-up questions and answers for Exposure to practice groups

<table>
<thead>
<tr>
<th>Situation</th>
<th>Screen</th>
<th>Question to observer</th>
<th>Possible answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>What's NAME's preferred location?</td>
<td>la playa / el centro / el campus</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>How many rooms is NAME looking for?</td>
<td>una habitación / dos habitaciones</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Does NAME have a pet in Málaga?</td>
<td>sí / no</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Does NAME prefer to share the apartment with a guy or a girl?</td>
<td>chico / chica</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>What kind of roommate does NAME prefer? One who is usually at home, or one who is usually out and about?</td>
<td>en casa / fuera de casa</td>
</tr>
<tr>
<td>B</td>
<td>22</td>
<td>According to NAME, how much money was stolen?</td>
<td>un poco de dinero / mucho dinero</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>What does NAME's roommate do for a living?</td>
<td>es estudiante / trabaja</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>NAME kept the stolen money in a book. What was it next to?</td>
<td>cama / mesa</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>Who does NAME think stole the money?</td>
<td>compañero de cuarto / señora de la limpieza</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>What does NAME think about the police officer's hypothesis?</td>
<td>probable / improbable</td>
</tr>
</tbody>
</table>

### 3.3 Treatment feedback

In Talking to Avatars, when a student in the Practice condition produced an answer, a two-step process was initiated. First, the software string-matching algorithm evaluated
the quality of the student’s response, and second, one of six possible avatar feedback videos was played, based on a feedback delivery protocol. These two processes are explained below.

The string-matching algorithm. To analyze the quality of the students’ responses Talking to Avatars uses a string-matching algorithm that selectively focuses on the target form, ignoring other potential errors. Each response is evaluated according to specific criteria. Observe, for example, the two sample cloze sentences from situations A (Finding an apartment and a roommate) and situation B (Reporting a theft), respectively.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Sample Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td><em>Quiero un apartamento (THAT IS) ________ en</em> (la playa / el centro / el campus)</td>
</tr>
<tr>
<td></td>
<td>[“I would like an apartment (THAT IS) ________ (on the beach / downtown / on campus)”]</td>
</tr>
<tr>
<td>B</td>
<td><em>La persona (THAT I LIVE WITH) ________ (es estudiante/trabaja).</em></td>
</tr>
<tr>
<td></td>
<td>[“The person I live with is a student.”]</td>
</tr>
</tbody>
</table>

For the sample item in situation A, the string-matching algorithm parsed as “correct” any answer containing the third person singular Spanish present subjunctive form esté (or its unaccented variants), and anything else was parsed as “incorrect.” In turn, for the sample item in situation B, answers were considered “correct” if the preposition con was placed before the relative pronoun que (or any of its counterparts quien, el/la cual), and “incorrect” otherwise. It should be noted that in Spanish some prepositions trigger the insertion of an article before the relative pronoun (e.g., *la compañía para LA que trabaja* [“the company for-THE-which she works”]). However, inserting or leaving out an article
when appropriate did not alter the nature of the evaluation, since this was not the focus of
the study.

*Feedback delivery protocol.* During the course of the interaction with the avatars
participants were exposed to one of five different types of corrective feedback, i.e., zero
feedback or a message resulting from the combination of two feedback components,
negative evidence with two different degrees of explicitness (Spot error, Explain error)
and prompting for error repair (Prompt, Continue). All feedback messages were fully
delivered in Spanish. For illustrative purposes, a sample mini-episode from the
experimental materials is reproduced next (see also Appendices 5.1 A and B).
AVATAR’S QUESTION:

Vale, comencemos por el apartamento. ¿En qué área de la ciudad te gustaría vivir? ¿En la playa, en el centro de la ciudad o en el campus universitario?

[“Okay, let’s begin with the apartment. Which part of the city would you like to live in? On the beach, downtown, or on the university’s campus?”]

STUDENT’S ILL-FORMED ANSWER:

Quiero un apartamento QUE *ESTÁ en el campus

[“I want an apartment THAT *IS-INDICATIVE on campus”].

AVATAR’S FEEDBACK:

¡Buena elección! Oye, tienes un problema con el verbo ESTAR, pero te comprendo. Continuemos.

[“Good choice! Hey, you have a problem with the verb TO BE, but I understand what you are saying. Let’s move on.”]

STUDENT’S ILL-FORMED ANSWER:

Quiero un apartamento QUE *ESTÁ en el campus

[“I want an apartment THAT *IS-INDICATIVE on campus”].

AVATAR’S FEEDBACK:

¡Buena elección! Oye, tienes un problema con el verbo ESTAR. ¿Puedes reformular tu frase, por favor?

[“Good choice! Hey, you have a problem with the verb TO BE. Could you reformulate your sentence, please?”]

EXPLAIN & CONTINUE

¡Buena elección! Oye, tienes un problema con el verbo ESTAR. Estamos hablando de un apartamento hipotético, en tu imaginación. Hay muchos apartamentos en el centro, en la playa y en el campus. Por eso, aquí tienes que usar el subjuntivo. Pero no te preocupes, continúa.

[“Good choice! Hey, you have a problem with the verb TO BE. We are talking about a hypothetical apartment here, one in your imagination. There are many apartments downtown, on the beach, and on campus. Consequently, you should use the subjunctive here. But don’t worry, let’s move on.”]

EXPLAIN & PROMPT

¡Buena elección! Oye, tienes un problema con el verbo ESTAR. Estamos hablando de un apartamento hipotético, en tu imaginación. Hay muchos apartamentos en el centro, en la playa y en el campus. Por eso, aquí tienes que usar el subjuntivo. ¿Puedes reformular tu frase, por favor?

[“Good choice! Hey, you have a problem with the verb TO BE. We are talking about a hypothetical apartment here, one in your imagination. There are many apartments downtown, on the beach, and on campus. Consequently, you should use the subjunctive here. Could you reformulate your sentence, please?”]

NO FEEDBACK

[The avatar asked the next question without providing any feedback.]
In order to keep amount of feedback constant, two measures were taken. First, the videos with the avatars’ feedback messages were played only once, as opposed to the videos with the avatars’ questions, which the learner could replay at will. And second, if the student’s answer was correct, positive feedback was provided. In this case, the Prompt/Continue feedback levels were reduced to only Continue, as prompting for error repair did not apply. Below are some illustrative examples.

---

**STUDENT’S CORRECT ANSWER:**

*Quiero un apartamento QUE ESTÉ en el campus*

[“I want an apartment THAT IS-SUBJ on campus”].

---

**AVATAR’S FEEDBACK:**

**SPOT FEEDBACK**

¡Buena elección! Y muy bien, el verbo ESTAR es correcto.

[“Good choice! And well done, the verb TO BE is correct.”]

**EXPLAIN FEEDBACK**

¡Buena elección! Y muy bien, el verbo ESTAR es correcto. Estamos hablando de un apartamento hipotético, en tu imaginación. Hay muchos apartamentos en el centro, en la playa y en el campus. Por eso, usa el subjuntivo.

[“Good choice! And well done, the verb TO BE is correct. We are talking about a hypothetical apartment here, one in your imagination. There are many apartments downtown, on the beach, and on campus. Therefore, use the subjunctive here.”]

---

Since the goal of the study was to examine the potential contribution of oral rather than written feedback in the L2, transcriptions and translations of the feedback messages were not made available, unlike for avatars’ questions. To keep amount of feedback constant several measures were taken. First, only one opportunity for repair was allowed, and second, no feedback was produced after the student’s repair, whether it was
successful or not. This ensured that any potential differences between the Prompt and Continue groups could be fully attributable to the prompting component of feedback rather than the extra amount of negative evidence. Also, positive evidence was not provided anywhere in the task or the feedback messages. Although the metalinguistic explanations of the Explain error groups contained key information to produce the correct form, no models were provided at any moment. Finally, feedback messages were customized for every item, both in terms of content and form. By doing so, the repetitive nature of feedback was lessened and the communicative value of the discourse was increased. Additionally, this allowed students in the Explain error groups to see how particular variants of a rule may govern the use of a structure depending on the context. Three illustrative examples are reproduced below.

*Bueno, hay que ahorrar dinero... Por cierto, ¡buena gramática! El verbo COSTAR es correcto. Estás hablando de un apartamento genérico, en tu cabeza. No sabes exactamente cuánto dinero cuesta. Puede costar 600, 700, 800 euros... Por eso, usa el subjuntivo.*

[“Right, you should try to save some money... By the way, good grammar! The verb TO COST is correct. You’re talking about a generic apartment, one in your head. You don’t know exactly how much it will cost. It may cost 600, 700, 800 euros... Therefore, use the subjunctive.”]

*Qué interesante. Y muy bien, el verbo PERMITIR es correcto. Tu conoces exactamente el apartamento del que estás hablando. Es tu apartamento en América. Por eso, usa indicativo. ¿Sabes? A mí me encantan los animales. ¡Tengo un perro!*  

[“Very interesting. And well done, the verb TO ALLOW is correct. You know very well the apartment you are talking about. It’s your apartment in America. Therefore, use the indicative. You know what? I love animals. I have a dog!”]

*Vale, el verbo BEBER es correcto. Cuando dices lo que piensas sobre el mundo, cuando expresas tu opinión sobre algo, tu percepción de la realidad, tienes que usar el indicativo. Por cierto, ten cuidado, que en España el alcohol es demasiado barato...*

[“Alright, the verb TO DRINK is correct. When you say what you think about the world, when you express your opinion about something, your perception of reality, you have to use the indicative. By the way, watch out, because alcohol in Spain is too cheap...!”]
3.4 Assessment measures

The present study used three different types of assessment tests: *untimed written production*, *timed oral production*, and *untimed grammaticality judgment*. All three tests were delivered through the computer at each of the three time points in the experiment (pretest, posttest, delayed posttest). To minimize test-effects, the oral production test was administered first, followed by the written production test, and finally the grammaticality judgment test. To control for test-retest effects, two different versions were created for all tests (except grammaticality judgment). The A versions were used for the pretests and then again for the delayed posttests, which were administered five weeks later. The items were the same but they were presented in randomized order. In turn, the B versions, containing different items, were used in the immediate posttests.

Notably, different time constraints were set for each type of test. As pointed out by a number of authors (Doughty, 2003; Ellis, 2007; Norris & Ortega, 2000) experimental research in form-focused instruction often relies on assessment tests that call on explicit memory-based performance, seemingly favoring explicit conditions. Consequently, as posed by, e.g., Ammar (2008: 201), it is important to select proper measurement tools “in order to avoid biased findings disfavouring implicit treatments.” Along these lines, it could be argued that untimed tests give learners in the more explicit conditions an added advantage, since they have the opportunity to remember and use the metalinguistic rules that they learned during the treatment. To counterbalance this, the written production and grammaticality judgment tests were untimed, while the oral production test was timed.

Finally, since the present dissertation sought to investigate the separate contributions of negative evidence and prompting for repair in the absence of positive evidence, special
care was taken in ensuring that participants would not receive models of the targeted structures either during the treatment or testing phases. For that reason, multiple choice recognition tests were not included in this design, and the grammaticality judgment test asked students to evaluate the (un)grammaticality of the non-targeted rather than targeted structures (more on this later). A separate description of the nature of each test is provided below.

The *untimed written production test* was created with Blackboard teaching and learning platform (http://www.blackboard.com) and individually administered to all participants through Georgetown University’s gateway (http://campus.georgetown.edu). In this test students were presented with a context cue describing a situation, followed by an unfinished Spanish sentence that students had to complete by translating a capitalized English segment. To avoid vocabulary problems, English glosses of the main content words in the Spanish sentence were made available by hovering the mouse pointer over them. The test contained 30 items total, with 20 of them eliciting the targeted structures (10 items for Present subjunctive in relative clauses and 10 items for Prepositional relative clauses) and 10 fillers eliciting the non-targeted structure counterparts (Non-prepositional relative clauses in Present indicative). No time-constraints were set. Figure 5.6 below provides a snapshot of the test instructions and a sample item. The complete list of items is available in Appendices 5.2 A and B.
Figure 5.6 Untimed written production task

Instructions

In the next pages you will be presented with a number of situations that you may encounter in real life and how you could react to them. The situations are in English, but your answers should be in Spanish. In order to help you, the beginning of each answer is already provided. Simply complete your answers by translating the CAPITALIZED English bits into their Spanish equivalents.

If you don’t know a word, hover your mouse over it and you will get a Spanish gloss. Not all words are glossed. Just the most difficult ones. If you don’t know a word, improvise. Please do not worry if you do not know how to express something in Spanish. I am more interested in your intuitions than anything else. Thanks!

Sample item (Present subjunctive)

You accidentally stained your apartment’s carpet and need a stain remover, so you go to the nearest drug store.

YOU SAY: “Busco un producto THAT REMOVES STAINS FROM CARPETS.”

The timed oral production test was individually administered to every student via an automated Microsoft PowerPoint presentation, and the students’ responses were recorded in MP3 format using QuickTime audio recording software. The test consisted of a sentence-combining task where students had to join two sentences by using a relative clause. Information appeared on the screen gradually, to facilitate processing. To prime usage of the targeted structure, the beginning of the combining sentence was already provided, and students were given 7 seconds to orally complete it. This time limit was decided after a pilot session where, overall, students were unable to complete the sentences in 5 seconds and finished ahead of time if they were given 10. In order to familiarize students with the task, they practiced with two dummy items before the test, and where encouraged to ask the researcher if they needed extra practice. Similarly to the written production test, the oral production test consisted of 20 targeted items (10 per
structure) and 10 fillers. Figure 5.7 shows the instructions provided and a sample item.

Appendices 5.3 A and B include the complete list of items.

**Figure 5.7** Timed oral production task

**Instructions**

In this task you will see a picture of a person or a thing and 2 sentences with information about them.

Your goal is to link these 2 sentences into 1 by using the relative pronoun “QUE” and including all the words in RED color.

State your answers orally by talking clearly into the microphone.

You’ll have 7 seconds to provide an answer. If you are unsure, please don’t remain quiet, just give it your best shot.

**Sample item (Prepositional relative clause)**

*Una PUERTA*

*En la puerta hay un agujero.*  
(There is a peephole in the door)

*Tú PODER VER el otro lado POR el agujero.*  
(You can see the other side through the peephole)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  *En la puerta hay un agujero…*

The *untimed grammaticality judgment* was created and administered using Blackboard online learning platform. In this test students were presented with an English sentence
and its alleged translation into Spanish, where the bits related to the targeted structures had been capitalized (the relative clause verb, in the case of Present subjunctive vs. indicative, or the preposition, in the case of Prepositional vs. non-prepositional relative clauses). Students were then asked to focus on the capitalized bits to decide whether the Spanish translation was grammatically “correct” or “incorrect,” or to click on an “I don’t know” option. Differently from other grammaticality judgment tests, this test did not provide models of the targeted structures anywhere, to preclude exposure to positive evidence. The test contained 20 items. Five of them used the Present indicative where the Present subjunctive should have been used instead. Similarly, other five sentences contained preposition-stranding where preposition pied-piping should have applied. Finally there were ten fillers, five for each of the two non-targeted structures, where these were correctly used. Figure 5.8 shows the instructions provided to the students and a sample of each item type. For a complete list of items, see Appendix 5.4.
Figure 5.8 Untimed grammaticality judgment task

Instructions

Next you will see 20 English sentences and their translations into Spanish. Decide whether the Spanish translations are grammatically correct or incorrect. If you don’t know the answer, select I don’t know. Focus specifically on the CAPITALIZED words.

PLEASE NOTE: You don’t have to judge whether the word choice is correct (e.g., “ser” vs. “estar”). Rather, what we ask you to do here is to determine whether the sentence is grammatical or ungrammatical. For example, in English "I don’t do nothing" is ungrammatical, because you shouldn’t use double negation.

Targeted item (Present subjunctive)

ENGLISH: Carmen wants to meet a man who loves her.
SPANISH: Carmen desea conocer a un hombre que la QUIERE.

Targeted item (Prepositional relative clause)

ENGLISH: Coffee is something that I cannot live without.
SPANISH: El café es una cosa que no puedo vivir SIN.

Filler (Present indicative)

ENGLISH: Antonio bought a car that runs on electricity.
SPANISH: Antonio ha comprado un coche que FUNCIÓN con electricidad.

Filler (Non-Prepositional relative clause)

ENGLISH: I don’t like people who smoke in the workplace.
SPANISH: No me gustan las personas que fuman EN las oficinas.
4. Procedure

This study followed a classic experimental design of pretest (week 1) – treatment and immediate posttest (week 2) – delayed posttest (week 5). During the pretest session, participants completed an interactive review lesson on the morphology of various Spanish tenses (present indicative and subjunctive, preterit and imperfect indicative) and the structure of non-prepositional relative clauses (see Appendix 5.5). Next, students completed all assessment tests, including timed oral production, untimed written production, and untimed grammaticality judgment, in that order. Finally, they answered some biometric questions and reported whether they had been exposed to formal instruction in any Romance language other than Spanish for two semesters or more.

The treatment session took place one week later. Group assignment was totally randomized. Students entered the language laboratory, sat at a computer of their choice, logged onto Talking to Avatars (http://www.talkingtoavatars.com) and the program central database assigned them to the next group in a loop. The 5 groups in the Practice condition were assigned first. Then, as these students finished their sessions, the database created log files with the students’ answers that served to generate video files for the Exposure to practice participants. Whenever there were video files available, participants got assigned to the Exposure to practice condition. Otherwise, they got to practice. All participants received practice or exposure to practice with both linguistic forms during the treatment session. To counterbalance performance effects derived from e.g., familiarity with the task or fatigue, the two communicative situations (i.e., Finding an apartment and a roommate, Reporting a theft) were presented in randomized order. Once
the treatment was over, participants completed the posttests of all three dependent measures.

Finally, during the fifth week of the experiment, participants completed a delayed posttest for all three dependent measures and filled in a debriefing questionnaire. This questionnaire asked students to rate several aspects of the CALI tutor and the experimental sessions, and to explain whether they had received any external exposure to the targeted forms throughout the duration of the experiment (see Appendix 5.6).

5. Coding and scoring

As explained earlier, the present study used oral production, written production, and grammaticality judgment pretests, posttests, and delayed posttests to assess learning of two different targeted structures (Present subjunctive in relative clauses, and Prepositional relative clauses) in contrast to their non-targeted counterparts (Present indicative in relative clauses and Non-prepositional relative clauses). Many previous SLA studies have compared the effects of pedagogical interventions on newly instructed grammatical forms against forms already known by the learner (e.g., subjunctive vs. indicative, perfect vs. simple tense, imperfect vs. preterit, accusative vs. nominative, non-canonical word order vs. SVO); however, in most cases scoring protocols focused exclusively on the newly instructed forms, overlooking the fact that learners often overgeneralize recently acquired knowledge. For example, it is a common occurrence that students in the process of learning Spanish subjunctive extend its use to contexts where they earlier appropriately used the indicative. On that note, in a partial replication of Rosa and Leow (2004b), Cerezo and Leow (2006) found, using the same tests, that
students learning past counterfactual conditional sentences in opposition to already-learned present and future counterfactual conditionals often overgeneralized the former, using it where the latter should have applied. However, because Rosa and Leow’s scoring procedure did not penalize overgeneralizations, they may have inappropriately assigned “learning” to students who failed to correctly discriminate form-meaning mappings by consistently using the targeted structure throughout. To avoid this, the present study used a scoring protocol that not only rewarded correct use of the newly taught structures but also penalized overgeneralizations. Tests contained and equal number of items eliciting the targeted and non-targeted structures. Each correct use of the targeted structure was granted 1 point, each correct use of the non-targeted structure was assigned 0 points, and each overgeneralization of the targeted structure where the non-targeted counterpart applied was penalized with -1 point.

The *written production test* in this study contained 30 items total. Out of these, 10 items elicited Present subjunctive, 10 items elicited pied-piping in Prepositional relative clauses, and 10 items elicited Present indicative in Non-prepositional relative clauses. The minimum score for each targeted structure was -10 points (i.e., the student did not use the targeted structure in any of the targeted items and used it in all the non-targeted items) and the maximum score was 10 points (i.e., the student used the targeted structure in all targeted items and the non-targeted structures in all non-targeted items). If a student used the targeted structures systematically for all items, the final score was zero (10 points for correctly using the targeted structure in all targeted items and -10 points for overgeneralizing it to all non-targeted items). Scoring focused on the targeted/non-targeted alternation exclusively (mood selection and preposition order). Thus, if for
example a learner made a mistake outside the relative clause, e.g. *Yo *querer una casa que tenga dos habitaciones [“I *to want a house that has-SUBJ two bedrooms”] s/he was still awarded one full point. Also, concessions were made if the student produced an inaccurate form but there was a clear attempt to use it. For example, in translating the verb tener (“to have”), some students correctly used the accurate form tiene for the indicative and the inaccurate form *tiena (instead of tenga) for the subjunctive, but they still received full credit because the object of this study for this particular targeted structure was measuring form-meaning connections rather than form accuracy. Similarly, in translating “the company she works for” some students appropriately used preposition pied-piping but exchanged the preposition, e.g., la compañía *por que trabaja instead of la compañía para que trabaja, but they still received full credit. Any other errors were penalized. For example, it was observed that as a result of the treatment a number of students changed from preposition stranding (la persona que vivo con) to defective preposition pied-piping (la persona que con vivo instead of la persona con que vivo), but they were granted zero points. Future re-coding of collected data should observe whether partial credit contributes a different picture.

Similarly to the written production test, the oral production test also contained 30 items, 10 eliciting each targeted structure and 10 fillers eliciting the non-targeted counterparts. Again, the minimum and maximum scores were -10 and 10 points, respectively. The same scoring criteria were applied. The coding process entailed listening to and evaluating a total of 96 hours of recorded audio (128 participants x 3 tests per participant x 15 minutes per test = 5760 minutes = 96 hours). To significantly reduce coding time, audio-files were open in Final Cut Pro video editing suite with the “show
Audio Waveforms” feature enabled (see Figure 5.9 below). This method allowed the researcher to scroll back and forth the audio waveforms at will, easily replaying bits if necessary and skipping the 18 second gap between each student verbalization (information appeared gradually on the screen for 11 seconds, and then students were given 7 seconds to complete the sentence after the sentence cue was displayed). Naturally, audio files were differently problematic (some students mumbled throughout the entire recording time or self-corrected extensively, thus creating a noisier waveform), but overall this method allowed for greater control over the audio files, significantly reducing coding time.

**Figure 5.9 Audio waveform view in Final Cut Pro**

Finally, the grammaticality judgment test contained 20 items, 5 eliciting each targeted structure and 5 fillers for each non-targeted counterpart. The minimum and maximum scores were -5 and 5 points, respectively. Coding was performed by the computer in a two-step process. First, Blackboard generated an Excel spreadsheet where all answers had already been labeled as “correct,” “incorrect,” or “I don’t know” by matching the
students’ answers against the preliminary provided list of correct solutions. Then, a “search and replace” script was run to convert these labels into 0 points, 1, or -1.

6. Proposed statistical analysis

The present dissertation examined the role of type of learners’ Agency (RQ1), type of oral Feedback (RQ2), type of linguistic Form (RQ3), and their combined interactions with Time (RQ4) in L2 development using a pretest, posttest, and delayed posttest design and three dependent measures (written production, oral production, and grammaticality judgment). To that purpose, assessment data for two grammatical Forms (Present subjunctive, Prepositional relative clauses) were collected from eight different groups of students. Specifically, five groups of students were engaged in Practice with four different types of Feedback (Spot & Continue, Spot & Prompt, Explain & Continue, and Explain & Prompt) as well as No Feedback whatsoever. In turn, due to limited availability of participants the number of Exposure to practice groups was reduced from five to three, i.e., the two Prompt groups—which were deemed to warrant exposure to richer interactions than the Continue groups—and the No Feedback group, for control purposes (see Figure 5.1 in Section 1 above). For this reason, the uncounterparted Continue groups were excluded from the analyses involving Agency (RQ1 and RQ4).

RQ1 (Time x Agency). The first research question addressed whether type of learner’s Agency in grammar practice (i.e., participating in interactive grammar practice versus watching others practice) has differential effects on L2 learners’ development of a grammatical structure. To answer this question, two different batteries of analyses were
conducted, using data from the six experimental groups highlighted in Figure 5.10 below. The first battery of analyses addressed the effects of type of Agency across Time holistically for all counterparted experimental groups, regardless of type of Feedback and Form. To that aim, the raw achievement scores of all six groups on each of the three assessment tests were entered into a 3x2x2x3 (Time x Form x Agency x Feedback) repeated-measures ANOVA, with two within-subject factors, i.e., Time (pretest, posttest, delayed) and Form (Prepositional relative clauses, Present subjunctive), and two between-subjects factors, i.e., Agency (Practice, Exposure to practice) and Feedback (Spot & Prompt, Explain & Prompt, No Feedback). In turn, the second battery of analyses zoomed on the effects of Agency more locally within each experimental dyad, by type of Feedback and Form. For each of the three Feedback dyads in Figure 5.10, the raw scores by Form and dependent measure were submitted to separate 3x2 (Time x Agency) repeated-measures ANOVAs, with Time as the within-subject factor and Agency (Practice, Exposure to practice) as the between-subject factor.

**Figure 5.10** Statistical analysis for RQ1 (3x2x2x3 repeated-measures ANOVAs for all six groups and 3x2 repeated-measures ANOVAs for each group dyad and form)

<table>
<thead>
<tr>
<th>Practice SC</th>
<th>Practice SP</th>
<th>Practice EC</th>
<th>Practice EP</th>
<th>Practice NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure SP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Exposure SP]</td>
<td>[Exposure EP]</td>
<td>[Exposure NF]</td>
<td></td>
</tr>
</tbody>
</table>

RQ2 (Time x Feedback). The second research question addressed whether type of oral Feedback provided in response to learners’ productions of a grammatical structure has differential effects on learners’ development of that form, and if so, which of two
particular subcomponents of feedback, explicitness of negative evidence and prompting for repair, may contribute to any observed differences. To answer this research question, two different batteries of analyses were performed, using data from the five Practice groups highlighted in Figure 5.11 below. The first battery addressed the effects of type of Feedback in the development of the two grammatical structures in this design combined, by entering the raw achievement scores of all five Practice groups in each of the three assessment tests into a 3x2x5 (Time x Form x Feedback) repeated-measures ANOVA, with two within-subject factors, i.e., Time (pretest, posttest, delayed) and Form (Prepositional relative clauses, Present subjunctive), and one between-subjects factor, Feedback (Spot & Continue, Spot & Prompt, Explain & Continue, Explain & Prompt, No Feedback). In turn, the second analysis looked at the effects for each form independently, by submitting the raw scores of the five Practice groups for each dependent measure into separate 3x5 repeated-measures ANOVAs, with Time (pretest, posttest, delayed) as the within-subject factor, and Feedback (Spot & Continue, Spot & Prompt, Explain & Continue, Explain & Prompt, No Feedback) as the between-subject factor. These analyses were performed twice, one time for each grammatical structure.

Figure 5.11 Statistical analysis for RQ2 (3x2x5 repeated-measures ANOVAs for both forms and 3x5 repeated-measures ANOVAs for each form independently)
RQ3 (Time x Form). The third research question addressed whether type of linguistic Form (i.e., more vs. less complex) has differential effects on L2 development as a result of a pedagogical intervention. Since in the present dissertation all eight experimental groups received treatment for both targeted structures, Spanish Present subjunctive and Prepositional relative clauses, all groups were considered in the analysis (see Figure 5.12). Two different batteries of analyses were performed. The first battery investigated the effects of type of Form holistically, by submitting the raw scores of all eight groups for each dependent variable to a 3x2 (Time x Form) repeated-measures ANOVA, where Time was entered as the dependent factor and Form was entered as the independent factor. In turn, the second battery zoomed on the effects of type of Form for each experimental group independently, by submitting the raw scores of each group again into a 3x2 (Time x Form) repeated-measures ANOVA.

**Figure 5.12** Statistical analysis for RQ3 (3x2 repeated-measures ANOVA for all eight experimental groups and 3x2 repeated-measures ANOVA for each group independently)

RQ4 (Interactions: Time [Form, Agency, Feedback]). The fourth research question addressed whether type of learner’s Agency, type of oral Feedback, and type of linguistic Form interact with each other towards different learning outcomes. To answer this, the raw scores of the six counterparted cells in this design (see Figure 5.13 below) on all three dependent variables were separately submitted to a 3x2x2x3 (Time x Form x
Agency x Feedback) repeated-measures analysis of variance (ANOVA), using a two within-subject, two between-subject design. The within-subject factors were Time (pretest, immediate posttest, delayed posttest) and Form (Prepositional relative clauses, Present subjunctive), since all participants completed all tests for both targeted structures. In turn, the between-subject factors were Agency (Practice, Exposure to practice) and Feedback. For this latter variable, only three levels out of five, those of the counterparted treatments, were entered (Spot & Prompt, Explain & Prompt, No Feedback). Naturally, this was taken into consideration when discussing the results.

**Figure 5.13** Statistical analysis for RQ4 (3x2x2x3 repeated-measures ANOVAs for all six counterparted groups)
CHAPTER 6.

RESULTS

This chapter reports on the results of the statistical analyses performed. Section 1 presents the results of the reliability analyses for all the assessment tests (written production, oral production, and grammaticality judgment) across time (pretest, posttest, delayed posttest). Section 2 includes various descriptive statistics. First, subsection 2.1 presents the descriptive statistics of the dependent variable, L2 accuracy development, for each test arranged by type of Agency, type of Feedback, and type of Form; next, subsection 2.2 includes the descriptive statistics for time-on-task during the treatment, again arranged by experimental group and grammatical form, while subsection 2.3 presents the descriptive statistics on the attention rates paid by the Exposure to practice groups to the Practice groups as these were completing the treatment. Section 3 presents the results for each of the four research questions under investigation by submitting raw achievement scores to separate batteries of repeated-measures analysis of variance (ANOVA). Finally, section 4 reports the results of the e-tutor evaluation survey by submitting Likert-scale scores to one-way ANOVAs. All analyses were run using the Statistical Package for the Social Sciences (SPSS), with the alpha level set at 0.05 throughout. Effect size was measured via partial eta-squared ($\eta_p^2$, a measure of the portion of variance attributable to a factor), in which small, medium, and large effects were operationalized as .01, .06, and .14, respectively. For greater transparency and comparability purposes, all plotted means are scaled according to the maximum and minimum scores possible on each test.
1. Test reliability analyses

Prior to conducting statistical analyses to answer the research questions, test reliability analyses were performed on all three tests (written production, oral production, and grammaticality judgment) at each time point (pretest, posttest, delayed posttest). As mentioned in Chapter 5, the number of participants across tests varied for several reasons, including data loss (some oral test files could not be retrieved or were blank); technology problems (the e-tutor crashed for the Exposure to practice groups in some cases); and external exposure to the targeted structures (participants reported consulting one or both of the grammatical structures outside the laboratory setting throughout the duration of the study).

For each of the two grammatical structures tested, both written and oral production tests comprised a total of 20 items (10 targeted items and 10 non-targeted items), while grammaticality judgment tests consisted of 10 items (5 targeted items and 5 non-targeted items). No items were removed from the analyses, and the reliability coefficients were computed using Cronbach’s alpha in all cases. Table 6.1 below provides a summary of results for both targeted structures, Spanish Prepositional relative clauses and Present subjunctive in adjectival relative clauses.
Table 6.1 Reliability coefficients for all assessment tests by Form

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prepositional relative clauses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written production (N= 123)</td>
<td>.521</td>
<td>.925</td>
<td>.930</td>
</tr>
<tr>
<td>Oral production (N= 121)</td>
<td>.350</td>
<td>.926</td>
<td>.916</td>
</tr>
<tr>
<td>Grammaticality judgment (N= 123)</td>
<td>.381</td>
<td>.476</td>
<td>.581</td>
</tr>
<tr>
<td><strong>Present subjunctive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written production (N= 126)</td>
<td>.887</td>
<td>.873</td>
<td>.914</td>
</tr>
<tr>
<td>Oral production (N= 123)</td>
<td>.898</td>
<td>.875</td>
<td>.892</td>
</tr>
<tr>
<td>Grammaticality judgment (N= 126)</td>
<td>.436</td>
<td>.637</td>
<td>.706</td>
</tr>
</tbody>
</table>

As Table 6.1 illustrates, similar patterns can be observed for both grammatical structures. Both written and oral production immediate and delayed posttests yielded very high alphas (ranging from .916 to .930 for Prepositional relative clauses and from .873 to .914 for Present subjunctive), while grammaticality judgment tests produced notably lower alphas throughout the entire timeline (.381 to .581 for Prepositional relative clauses and .436 to .706 for Present subjunctive). Also, in the particular case of Prepositional relative clauses, pretest alphas were notably lower than posttest alphas across all three dependent measures, a pattern not observed for Spanish Present subjunctive.

Typically, standards for interpreting reliability coefficients state that a value of .60 is enough for exploratory research, while basic research should require alphas of .80 or higher and research leading to important decision-making calls for coefficients of minimally .90 (Nunnally & Bernstein, 1994). As table 6.1 illustrates, the obtained alphas
reach ideal standards for most of the written and oral production tests. The lower coefficients obtained for Prepositional relative clauses pretests, however, have a logical explanation that dissipates any potential doubts on their reliability. Participants across the board had full knowledge of the non-targeted structure (non-prepositional relative clauses), as opposed to zero or minimal prior knowledge of Prepositional relative clauses, and variance across participants was minimal. For example, in the oral pretest, while all 100% of the students got all 10 non-targeted items right, only less than 10% of the students got at least one of the 10 targeted items right, and among these only 5 of those 10 items were correctly answered. As a result, 15 out of the 20 items on the pretest were excluded due to zero variance and the reliability analysis was computed on a total of 5 items only. This explains the low coefficient obtained, since in shorter tests any given item measurement error constitutes a large percentage of students’ scores (Traub & Rowley, 1991). Arguably, the reliability of the pretest items is corroborated by the high coefficients obtained for the delayed posttest. Both tests, pretest and delayed posttest, contained the exact same items, only their relative order was different.

On the other hand, the low to medium alphas obtained for the grammaticality judgment tests for both structures also call for some considerations. First, unlike the written and oral production tests, which had a free-response format, grammaticality judgment tests followed a multiple-choice format, which arguably allowed for considerably higher guessing. Second, the nature of the test and the format of the instructions may have led students to interpret the test goal differently. Specifically, the text enhancement used to highlight the focus of the judgments may have influenced students to evaluate the capitalized words (i.e., the conjugated verb or the preposition) in
an isolated manner, without considering the context. Third, this test contained half the items of written and oral production tests (5 targeted items and 5 non-targeted items), and as mentioned earlier shorter tests typically produce lower reliability coefficients. Finally, grammaticality judgment tests were administered last, so students may have rushed their way to the end of the session after a lengthy battery of tests.

2. Descriptive statistics

This section includes various descriptive statistics. Subsection 2.1 presents the descriptive statistics of the dependent variable, L2 accuracy development, for each test. Subsection 2.2 includes the descriptive statistics for time-on-task during the treatment. Subsection 2.3 presents the descriptive statistics on the attention rates paid by the Exposure to practice groups.

2.1 L2 accuracy

Appendices 6.1 through 6.6 report the means and standard deviations for all experimental treatments by type of Agency (Practice, Exposure to practice) and type of Feedback (Spot & Continue, Spot & Prompt, Explain & Continue, Explain & Prompt, and No Feedback) for each targeted structure (Spanish Prepositional relative clauses and Present subjunctive in adjectival relative clauses). As stated in Chapter 5, the highest and lowest possible scores on both written and oral production tests were 10 positive or negative points (1 positive point for each correct targeted item and zero or minus 1 point for each non-targeted item, depending on whether they were answered correctly or incorrectly). In turn, with half the number of targeted and non-targeted items, the maximum and
minimum possible scores on the grammaticality judgment tests were positive and negative 5.

Preliminary observation of group means across time suggested that, overall, participants experienced learning gains for both grammatical structures. To identify which groups experienced significant gains, a battery of paired \( t \)-tests for each group at each time-interval, Time 1-2, Time 1-3, and Time 2-3, was computed. Significant gains for each time interval are indicated with superscripted letters on pretest and immediate posttest means in Appendices 6.1 through 6.6. Based on these results, a number of observations can be made for each targeted structure.

*Prepositional relative clauses.* Learning outcomes were highly consistent for all three dependent measures. Participants in the control groups, [Practice NF] and [Exposure NF], did not experience any significant learning. Conversely, most groups with Feedback experienced learning. Within the Practice condition, all groups learned. The highest gainer was [Practice EC], with immediate gains of 64% for written production (6.41 points out of 10), 34% for oral production (3.47 out of 10), and 57% for grammaticality judgment (2.88 out of 5). On the opposite end of the Practice condition, the lowest gainer was [Practice SC]. Switching to the Exposure condition, [Exposure EP] experienced learning across all three measures, while [Exposure SP] did not show any significant learning in any test.

*Present subjunctive.* Similarly to Prepositional relative clauses, the participants in the No Feedback groups, either [Practice NF] or [Exposure NF], did not experience any significant learning in any of the three tests. On the other hand, most of the experimental groups with Feedback showed learning gains on at least written and oral production tests.
Generally, however, gains seemed to be more moderate than for Prepositional relative clauses. In the Practice condition, the winning group was again [Practice EC], with immediate gains of 40% and 28% on written and oral production. On the other extreme was [Practice SC]—the only group in the Practice condition that did not experience any significant learning. In the Exposure to practice condition, both [Exposure SP] and [Exposure EP] learned significantly, with the latter experiencing greater gains. Overall, the grammaticality judgment test yielded the poorest results, with only two Practice groups, [Practice SP] and [Practice EC], experiencing significant learning.

Overall, the descriptive statistics obtained for both structures are pretty consistent with each other. The control groups, [Practice NF] and [Exposure NF], experienced no significant learning whatsoever. In the Practice condition, all groups showed significant gains in written and oral production of both structures, except [Practice SC], that only showed improvements for Prepositional relative clauses. The highest gainer was [Practice EC]. On grammaticality judgment tests, all groups showed learning of Prepositional relative clauses, while for Present subjunctive only [Practice SP] and [Practice EC] experienced significant learning. As for the Exposure to practice condition, only the [Exposure EP] group experienced learning of Prepositional relative clauses across all three tests, while for Present subjunctive both groups showed improvements in written and oral production.

2.2 Time on task

Table 6.2 below reports the average number of minutes spent by the Practice groups to complete the treatment. Due to a programming error, no time logs for the Exposure to
practice groups were recorded. However, since the role of the participants in these groups consisted of watching a video with the performance of their dyad partners and answering periodical check-up questions, the duration of their treatment is deemed to have been only slightly higher in each case.

Table 6.2 Time-on-task by experimental group (in minutes)

<table>
<thead>
<tr>
<th></th>
<th>Situation A</th>
<th>Situation B</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Practice SC]</td>
<td>16.16</td>
<td>18.72</td>
<td>34.88</td>
</tr>
<tr>
<td>[Practice SP]</td>
<td>13.70</td>
<td>19.11</td>
<td>32.82</td>
</tr>
<tr>
<td>[Practice EC]</td>
<td>18.76</td>
<td>21.52</td>
<td>40.29</td>
</tr>
<tr>
<td>[Practice EP]</td>
<td>14.88</td>
<td>20.00</td>
<td>34.88</td>
</tr>
<tr>
<td>[Practice NF]</td>
<td>10.50</td>
<td>14.56</td>
<td>25.06</td>
</tr>
</tbody>
</table>

*Note.* Situation A = Finding an apartment and a roommate (Present subjunctive); Situation B = Reporting a theft (Prepositional relative clauses)

As Table 6.2 illustrates, the fastest group to complete the treatment was [Practice NF], with an average of 25.06 minutes. On the opposite end was [Practice EC], with 40.29 minutes on average. Finally, the other three groups, [Practice SC], [Practice SP], and [Practice EP] clustered in the middle, with 34.88, 32.82, and 34.88 minutes, respectively. In terms of the two communicative situations, all groups consistently spent more time to complete situation B, “Reporting a theft,” which elicited Prepositional relative clauses.
2.3 Attention to practice

Throughout the treatment, participants in the Exposure to practice condition were asked five check-up questions per communicative situation to control that they paid attention to their partners’ as these completed the practice tasks (see Table 5.2 in Chapter 5 for a complete list of check-up questions and possible answers). Table 6.3 below presents the attention rates to each communicative situation by experimental condition.

Table 6.3 Attention to practice in Exposure to practice conditions by type of Form

<table>
<thead>
<tr>
<th></th>
<th>Situation A</th>
<th>Situation B</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Exposure SP]</td>
<td>91.07%</td>
<td>94.64%</td>
<td>92.85%</td>
</tr>
<tr>
<td>[Exposure EP]</td>
<td>90.38%</td>
<td>92.30%</td>
<td>91.34%</td>
</tr>
<tr>
<td>[Exposure NF]</td>
<td>96.37%</td>
<td>98.05%</td>
<td>97.21%</td>
</tr>
</tbody>
</table>

Note. Situation A = Finding an apartment and a roommate (Present subjunctive); Situation B = Reporting a theft (Prepositional relative clauses)

Since there were 5 check-up questions in each communicative situation, a ratio of 4/5 correct answers would translate into 80% accuracy. Clearly, since participants in all three Exposure to practice groups scored 90% or higher for each of the two structures, they demonstrated high levels of attention to task completion. For both targeted structures combined, participants answered at least 9 out of the 10 questions correctly. Looking at each group separately, it becomes clear that participants in the two groups with exposure to some type of feedback, [Exposure SP] and [Exposure EP], demonstrated similar levels of attention, with rates of 92.85% and 91.34%, respectively. In turn, participants in the [Exposure NF] condition scored even higher, with 97.21% accuracy in their responses. This may be due to the fact that, since no feedback videos were projected between the
performers’ answers and the check-up questions, memory decay played a lesser role. Overall, then, the results of these analyses indicate that any potential differences in learning outcomes between the Practice and Exposure to practice conditions cannot be attributed to a potential lack of attention to task completion by the latter.

3. Research questions

The next sections present the results of the statistical analyses for each research question. Sections 3.1 through 3.3 address the effects of type of learner’s Agency, type of corrective oral Feedback, and type of grammatical Form, in L2 development, while section 3.4 addresses their combined interactions.

3.1 Research question #1. Type of learner’s Agency and L2 development

The first research question addressed whether type of learner’s Agency in grammar practice (i.e., participating in interactive grammar practice versus watching others practice) has differential effects on L2 learners’ development of a grammatical structure. To investigate this, developmental data for two different grammatical Forms were collected from dyads of practicing learners and their observers under three different Feedback conditions, i.e., Spot & Prompt, Explain & Prompt, and No Feedback (see Figure 6.1 below).
To answer this research question, two separate batteries of analyses were performed. The first battery of analyses addressed the effects of type of Agency holistically for all the groups, regardless of type of Feedback and Form, while the second analysis looked at the effects of Agency for each Form and Feedback dyad independently. Following is a report of the first analysis (the two forms and the three feedback types combined), while sections 3.1 through 3.6 report on the independent analyses conducted by type of feedback and form.

To investigate the effects of type of Agency for the two forms and three feedback types combined, three steps were taken. First, individual participants’ means for the two structures at each of the three testing points were calculated. Second, one-way ANOVAs on the pretest mean scores were performed to identify possible between-group differences at the outset of the study. Third, the raw achievement scores of all six experimental groups on each of the three assessment tests were entered into a 3x2x2x3 (Time x Form x Agency x Feedback) repeated-measures ANOVA, with two within-subject factors, i.e., Time (pretest, posttest, delayed) and Form (Prepositional relative clauses, Present subjunctive), and two between-subjects factors, Agency (Practice, Exposure to practice) and Feedback (Spot & Prompt, Explain & Prompt, No Feedback).

The results of the one-way ANOVAs conducted on the pretests revealed no significant difference for either written production, $F(1, 90) = 1.387, p = .242$, oral
production, $F(1, 89) = 1.287, p = .260$, or grammaticality judgment, $F(1, 90) = .088, p = .768$. Clearly, then, any potential subsequent differences between groups can be fully attributable to the effects of the treatment rather than differences in prior knowledge.

The results of the 3x2x2x3 repeated-measures ANOVA on the *written production* tests yielded a significant main effect for Time, $F(2, 166) = 45.554, p = .000, \eta^2_p = .354$, no significant main effect for Agency, $F(1, 83) = 1.117, p = .294, \eta^2_p = .013$, and no significant interaction Time x Agency, $F(2, 166) = .549, p = .578, \eta^2_p = .007$ (see plotted means in Figure 6.2 below and summary of results in Appendix 6.18). For *oral production*, the ANOVA yielded a significant main effect for Time, $F(2, 162) = 29.136, p = .000, \eta^2_p = .265$, no significant main effect for Agency, $F(1, 81) = .016, p = .901, \eta^2_p = .000$, and no significant interaction Time x Agency, $F(2, 162) = .083, p = .920, \eta^2_p = .001$ (see Figure 6.3 and Appendix 6.19). Finally, for *grammaticality judgment*, the ANOVA yielded a significant main effect for Time, $F(2, 166) = 16.462, p = .000, \eta^2_p = .166$, no significant main effect for Agency, $F(1, 83) = .306, p = .581, \eta^2_p = .004$, and no significant interaction Time x Agency, $F(2, 166) = .150, p = .861, \eta^2_p = .002$ (see Figure 6.4 and Appendix 6.20).

The obtained main effects for Time on all three dependent measures indicate that, when the two forms, the two agency levels, and the three feedback types were considered together, learning was observed. In contrast, the lack of main effects for Agency reveal that when time, type of feedback, and type of form were left out of the equation there were no differences between the two agency levels. Finally, the no significant interactions Time x Agency indicate that when the different types of feedback and form were combined, the learning gains experienced by the two agency levels were
comparable. In other words, participants who practiced and participants who observed others practicing learned the same. This overall learning is low because it comprises the means of the control groups, [Practice NF] and [Exposure NF], which experienced zero learning.

**Figure 6.2** Written production accuracy for both forms in all six counterparted groups: Time by Agency
Figure 6.3 Oral production accuracy for both forms in all six counterparted groups: Time by Agency
The following sections present the results for each of the three group dyads and each form type separately. In all cases, one-way ANOVAs were performed on the pretests to identify possible between-group differences at the outset of the study. Next, a series of 3x2 (Time x Agency) repeated-measures ANOVAs were conducted to compare participants’ learning evolution. Agency (Practice, Exposure to practice) was entered as the between-subject factor, and Time (pretest, posttest, delayed) was entered as the
within-subject factor. Analyses were performed separately for each grammatical structure.

3.1.1 Prepositional relative clauses: Spot & Prompt treatment

The results of the one-way ANOVAs conducted on the pretests revealed a significant difference between the [Practice SP] and [Exposure SP] groups for written production, $F(1, 28) = 6.401, p = .017$, and no significant differences for either oral production, $F(1, 27) = .429, p = .518$, or grammaticality judgment, $F(1, 28) = .364, p = .551$. Although the difference in the written production pretest needs to be considered when interpreting the results, the repeated-measures ANOVA is still a valid procedure to measure the effects of Agency across time. Following are the results of this analysis by dependent measure.

**Written production.** The 3x2 repeated-measures ANOVA performed on the achievement scores of the written production task yielded a significant main effect for Time, $F(2, 56) = 11.098, p = .000, \eta_p^2 = .284$, a significant main effect for Agency, $F(1, 28) = 7.333, p = .011, \eta_p^2 = .208$, and a significant interaction Time x Agency, $F(2, 56) = 3.722, p = .030, \eta_p^2 = .117$. The $\eta_p^2$ figures here suggest that the effect sizes of the treatment are medium to large, since they account for 11% to 28% of the total accompanying variance (i.e., effect + error). These results, tabulated in Appendix 6.7, are best interpreted based on the plotted means in Figure 6.5 below.
The main effect for Time indicates that regardless of Agency, participants learned as a result of the treatments. In turn, the main effect for Agency reveals that, regardless of Time, both groups performed differently from each other. Finally, the significant interaction Time x Agency shows that both groups experienced significantly different learning gains across time. As mentioned earlier, both groups scored statistically differently on the pretest, so this must be considered when interpreting the results. In order to determine whether the groups also performed statistically differently on the
immediate and delayed posttests, two separate one-way ANOVAs using gain scores were computed. Results revealed that the difference was significant on the immediate posttest, $F(1, 28) = 7.098, p = .013$, but not two weeks later, $F(1, 28) = 2.255, p = .144$.

Furthermore, as reported in Appendix 6.1, a battery of paired sample $t$-tests for each group at each time interval revealed that the [Practice SP] group learned from pretest to posttest and maintained those gains two weeks later, while the [Exposure SP] group experienced no significant gains at any point.

In sum, participants in the [Practice SP] group experienced significant learning gains in written production of Spanish Prepositional relative clauses, both immediately and two weeks later. On the immediate posttest, they outperformed the participants that observed them, who in turn did not derive any significant learning from the experience.

**Oral production.** The 3x2 repeated-measures ANOVA performed on the achievement scores of the oral production task yielded a significant main effect for Time, $F(2, 52) = 7.545, p = .001, \eta_p^2 = .225$, no significant main effect for Agency, $F(1, 26) = 3.161, p = .087, \eta_p^2 = .108$, and a significant interaction Time x Agency, $F(2, 52) = 4.035, p = .023, \eta_p^2 = .134$ (see Appendix 6.7). Again, these results are best interpreted based upon the plotted means in Figure 6.6.
Figure 6.6 Oral production accuracy for Prepositional relative clauses in Spot & Prompt treatment: Time by Agency

The significant main effect for Time indicates that, regardless of Agency, participants learned across time. Furthermore, the significant interaction Time x Agency reveals that the [Practice SP] group learned significantly more than the [Exposure SP] group. The $\eta^2_p$ figures for this interaction suggest that the effect size is medium to high, since it accounts for almost 14% of the total accompanying variance. Given that both groups performed statistically similarly on the pretest, two separate one-way ANOVAs using gain scores were computed to determine whether the groups performed differently on both immediate
and delayed posttests. Results revealed that the difference was not significant on the immediate posttest, $F(1, 26) = 1.942, p = .175$, but it reached significance on the delayed posttest, $F(1, 26) = 5.229, p = .030$. Moreover, a battery of paired sample $t$-tests for every time-interval revealed that the [Exposure SP] group did not experience any significant learning gains at any point, whereas the [Practice SP] group experienced learning at Time 2 and maintained these gains at Time 3 (see Appendix 6.2 for a summary).

In sum, participants in the [Practice SP] group significantly improved their oral production of Spanish Prepositional relative clauses, both immediately and two weeks later. When the entire timeline was considered, they outperformed the participants that observed them, who did not experience any learning gains.

*Grammaticality judgment.* The 3x2 repeated-measures ANOVA performed on the achievement scores of the grammaticality judgment task yielded a significant main effect for Time, $F(2, 56) = 3.539, p = .036, \eta^2_p = .112$, a significant main effect for Agency, $F(1, 28) = 5.781, p = .023, \eta^2_p = .171$, and no significant interaction Time x Agency, $F(2, 56) = 1.712, p = .190, \eta^2_p = .058$ (see Appendix 6.7 for a summary of results). Once again these results are best interpreted based upon the plotted means in Figure 6.7.
The main effect for Time indicates that regardless of Agency, participants experienced learning after the pretest. To determine whether this was the case for both groups, a series of paired sample \( t \)-tests on each group scores measured the significance of change in accuracy for each time interval. Results revealed that only the [Practice SP] group experienced significant learning, from pretest to posttest, although significance was lost at the time of the delayed posttest (see Appendix 6.3). The obtained main effect for Agency indicates that the two groups performed differently from each other when considering their combined means regardless of Time. Clearly, according to the graph the
[Practice SP] group performed the best. At Time 2, results from a one-way ANOVA on gain scores revealed that the [Practice SP] approached significance over the [Exposure SP] group, $F(1, 28) = 3.680, p = .065$, while no significant difference was detected two weeks later, $F(1, 28) = 1.333, p = .258$. The lack of a significant interaction Time x Agency here means that both groups did not experience significantly different learning gains; however, the $\eta^2_p$ figures (.058) indicate a nearly medium rather than trivial effect size of the treatment, thus suggesting that a significant difference may surface after replication with a larger sample size.

3.1.2 Prepositional relative clauses: Explain & Prompt treatment

The results of the one-way ANOVAs conducted on the pretests revealed a significant difference between the [Practice EP] and [Exposure EP] groups for written production, $F(1, 28) = 6.827, p = .014$, and no significant differences for either oral production, $F(1, 28) = 1.356, p = .254$, or grammaticality judgment, $F(1, 28) = 1.534, p = .226$. Like in the previous treatment comparison, the observed difference on the written production pretest should be considered when interpreting the results. However, a repeated-measures ANOVA is still a valid procedure to measure the effects of Agency over time. The results of this analysis are separately presented below for each dependent measure.

Written production. The 3x2 repeated-measures ANOVA performed on the achievement scores of the written production task yielded a significant main effect for Time, $F(2, 56) = 21.225, p = .000, \eta^2_p = .431$, no significant main effect for Agency, $F(1, 28) = .060, p = .808, \eta^2_p = .002$, and no significant interaction Time x Agency, $F(2, 56) = .171, p = .844$,
$$\eta^2_p = .006$$ (see Appendix 6.8). These results are best interpreted based on the plotted means in Figure 6.8.

**Figure 6.8** Written production accuracy for Prepositional relative clauses in Explain & Prompt treatment: Time by Agency

The main effect for Time indicates that regardless of Agency, participants learned across time. In turn, the no significant effect for Agency, the no significant interaction Time x Agency, and the overlapping plotted means in Figure 6.8 clearly indicate that both [Practice EP] and [Exposure EP] experienced statistically similar gains throughout
the entire timeline. A battery of paired sample *t*-tests on each group indicated that both
groups learned at Time 2 and maintained those gains at Time 3 (see Appendix 6.1).

Finally, observation of the $\eta^2_p$ figures revealed interesting conclusions. On the one hand,
the effect size for Time was very large, accounting for over 40% of the total
accompanying variance, which indicated that the pedagogical effects of the treatment
were solid. Conversely, the effect sizes for Agency and the interaction Time x Agency
were minimal, accounting for barely .2% and .6% of the total variance. Consequently, it
can be concluded that no between-group differences are likely to arise upon further data
collection, which means that whenever some students observe other students engaging in
practice with very explicit feedback, both are bound to learn, and they are bound to learn
the same.

*Oral production.* The 3x2 repeated-measures ANOVA performed on the achievement
scores of the oral production task yielded a significant main effect for Time, $F(2, 56) =
14.301, p = .000, \eta^2_p = .338$, no significant main effect for Agency, $F(1, 28) = .353, p =
.557, \eta^2_p = .012$, and no significant interaction Time x Agency, $F(2, 56) = .580, p = .563,
\eta^2_p = .02$ (see Appendix 6.8). The plotted means in Figure 6.9 below provide a quick
snapshot of each group performance.
The main effect for Time reveals that participants learned regardless of Agency. In addition, the lack of a significant main effect for Agency and the lack of a significant interaction Time x Agency specify that both groups experienced similar gains over time. Results from a battery of paired $t$-tests for each time interval revealed that the [Practice EP] group significantly learned from pretest to immediate posttest and maintained these gains two weeks later. On the other hand, the [Exposure EP] group only reached significance at Time 3 (see Appendix 6.2). Similarly to written production, the large
effect sizes obtained for Time and the low effect sizes obtained for Agency and the interaction Time x Agency clearly indicate that the compared effects of the treatments are trivial. In other words, Practice with Explain & Prompt Feedback is conducive to learning, and the Agency of the learners (i.e., whether they practice or observe) does not make a difference.

**Grammaticality judgment.** The 3x2 repeated-measures ANOVA performed on the achievement scores of the grammaticality judgment task yielded a significant main effect for Time, $F(2, 56) = 20.994, p = .000, \eta^2_p = .429$, an almost significant main effect for Agency, $F(1, 28) = 3.970, p = .056, \eta^2_p = .124$, and no significant interaction Time x Agency, $F(2, 56) = .208, p = .813, \eta^2_p = .007$ (see Appendix 6.8 for a summary). The plotted means in Figure 6.10 below facilitate a better interpretation of these results.
The main effect for Time indicates that overall, participants experienced learning after the pretest. The lack of a significant interaction Time x Agency, however, reveals that both groups learned similarly. As the plotted means show, both groups evolved parallel to each other. The non significant difference at the outset of the study, repeated at Times 2 and 3, is responsible for the almost significant main effect for Agency. Results from a battery of paired *t*-tests on each time interval revealed that for both groups, learning occurred from pretest to posttest, and gains were maintained two weeks later.
(see Appendix 6.3). Again, the large effect size for Time and minimal effect size for the interaction Time x Agency reveal that both treatments were conducive to learning and that potential replication of this research is not bound to yield significant differences between the two groups.

3.1.3 Prepositional relative clauses: No Feedback treatment

Analyses for the No Feedback groups yielded highly consistent results for all three dependent measures. The one-way ANOVAs on the pretests did not yield any significant differences between groups for either written production, $F(1, 28) = .431, p = .517$, oral production, $F(1, 28) = .350, p = .559$, or grammaticality judgment, $F(1, 28) = .665, p = .422$. In turn, for each of these measures respectively, the results of the 3x2 repeated-measures ANOVA yielded no significant main effects for Time [$F(2, 56) = 2.206, p = .120, \eta^2_p = .073; F(2, 56) = 1.333, p = .272, \eta^2_p = .045; F(2, 56) = 1.188, p = .312, \eta^2_p = .041$], no significant main effects for Agency [$F(1, 28) = .355, p = .556, \eta^2_p = .013; F(1, 28) = 1.099, p = .303, \eta^2_p = 0.38; F(1, 28) = .569, p = .457, \eta^2_p = .02$], and no significant interactions Time x Agency [$F(2, 56) = .233, p = .793, \eta^2_p = .008; F(2, 56) = 1.778, p = .178, \eta^2_p = .06; F(2, 56) = .610, p = .547, \eta^2_p = .021$] (see Appendix 6.9 for a summary). Clearly, these results, supported by the plotted means in Figures 6.11, 6.12, and 6.13, reveal that there were no contrasts between the [Practice NF] and the [Exposure NF] groups because neither of them experienced any significant learning gains in written production, oral production, or grammaticality judgment of Spanish Prepositional relative clauses.
Figure 6.11 Written production accuracy for Prepositional relative clauses in No Feedback treatment: Time by Agency
Figure 6.12 Oral production accuracy for Prepositional relative clauses in No Feedback treatment: Time by Agency
3.1.4 Present subjunctive: Spot & Prompt treatment

The results of the one-way ANOVAs conducted on the pretests revealed a significant difference between the [Practice SP] and [Exposure SP] groups for written production, \( F(1, 29) = 5.034, p = .033 \), and no significant differences for either oral production, \( F(1, 27) = 2.236, p = .146 \), or grammaticality judgment, \( F(1, 29) = .438, p = .513 \). Although the difference in the written production pretest needs to be considered when interpreting the results, the 3x2 repeated-measures ANOVA is still a valid procedure to measure the
effects of Agency across time. This analysis yielded analogous results across all three 
dependent measures. For written production, the ANOVA yielded a significant main 
effect for Time, $F(2, 58) = 11.329, p = .000, \eta^2_p = .281$, no significant main effect for 
Agency, $F(1, 29) = .642, p = .430, \eta^2_p = .022$, and no significant interaction Time x 
Agency, $F(2, 58) = .194, p = .824, \eta^2_p = .007$. For oral production, the ANOVA produced 
a significant main effect for Time, $F(2, 54) = 8.946, p = .000, \eta^2_p = .249$, no significant 
main effect for Agency, $F(1, 27) = .404, p = .530, \eta^2_p = .015$, and no significant 
interaction Time x Agency, $F(2, 54) = .337, p = .716, \eta^2_p = .012$. Finally, for 
grammaticality judgment, the ANOVA showed a main effect for Time, $F(2, 58) = 4.361, 
p = .017, \eta^2_p = .131$, no significant main effect for Agency, $F(1, 29) = .665, p = .422, \eta^2_p 
= .022$, and no significant interaction Time x Agency, $F(2, 58) = .159, p = .853, \eta^2_p = 005$ 
(see Appendix 6.10 for a summary).

Overall, the obtained significant main effects for Time, the lack of significant main 
effects for Agency, and the no significant interactions Time x Agency show that 
participants experienced learning and that neither group outperformed the other (see 
plotted means in Figures 6.14, 6.15, and 6.16). When each group was analyzed 
independently, a battery of paired sample $t$-tests confirmed that both groups experienced 
gains in written and oral production; however, for grammaticality judgment only the 
[Practice SP] group showed gains, although these gains were not statistically superior to 
the [Exposure SP] group (see Appendices 6.4, 6.5, and 6.6). Closer examination of $\eta^2_p$ 
figures revealed that the effect sizes for Time were large for written and oral production 
and medium for grammaticality judgment, accounting for 28%, 25%, and 13% of the 
variance. However, the effect sizes for the no significant main effect for Agency and the
no significant interaction Time x Agency were very small, revealing that further replication of this study under comparable circumstances is not likely to yield significant differences between groups. These findings contrast with those for the Spot & Prompt treatment for Prepositional relative clauses.

**Figure 6.14** Written production accuracy for Present subjunctive in Spot & Prompt treatment: Time by Agency
**Figure 6.15** Oral production accuracy for Present subjunctive in Spot & Prompt treatment: Time by Agency
Figure 6.16 Grammaticality judgment accuracy for Present subjunctive in Spot & Prompt treatment: Time by Agency

3.1.5 Present subjunctive: Explain & Prompt treatment

The results of the one-way ANOVAs conducted on the pretests revealed no significant difference between the [Practice EP] and [Exposure EP] groups for any of the dependent measures, written production, $F(1,28) = 2.912, p = .099$, oral production, $F(1, 28) = .093, p = .763$, or grammaticality judgment, $F(1, 28) = .116, p = .736$. The 3x2 repeated measures ANOVA yielded results similar to those obtained for the previous treatment comparison, Spot & Prompt, on all three dependent measures. For written production, the
ANOVA yielded a significant main effect for Time, $F(2, 56) = 14.077$, $p = .000$, $\eta_p^2 = .335$, no significant main effect for Agency, $F(1, 28) = .094$, $p = .761$, $\eta_p^2 = .003$, and no significant interaction Time x Agency, $F(2, 56) = 1.346$, $p = .269$, $\eta_p^2 = .046$. For oral production, the ANOVA yielded a significant main effect for Time, $F(2, 56) = 9.627$, $p = .000$, $\eta_p^2 = .256$, no significant main effect for Agency, $F(1, 28) = .618$, $p = .438$, $\eta_p^2 = .022$, and no significant interaction Time x Agency, $F(2, 56) = 1.132$, $p = .330$, $\eta_p^2 = .039$. Finally, for grammaticality judgment the ANOVA showed a significant main effect for Time, $F(2, 56) = 3.243$, $p = .047$, $\eta_p^2 = 104$, no significant main effect for Agency, $F(1, 28) = 1.885$, $p = .181$, $\eta_p^2 = .063$, and no significant interaction Time x Agency, $F(2, 56) = .878$, $p = .421$, $\eta_p^2 = .03$ (see Appendix 6.11 for a summary).

Overall, the obtained significant main effects for Time, the lack of significant main effects for Agency, and the no significant interactions Time x Agency show that participants experienced learning and that there was no ‘winning’ group (see Figures 6.17, 6.18, and 6.19). Looking at each group independently, a battery of paired sample $t$-tests confirmed that both [Practice EP] and [Exposure EP] experienced gains in written and oral production; however, none of them experienced significant gains for grammaticality judgment, although the two together contributed to the found main effect for Time (see Appendices 6.4, 6.5, and 6.6). Again, the low effect sizes obtained for the non-significant interactions Time x Agency do not suggest that any potential between-group differences may surface as a result of further data collection.
Figure 6.17 Written production accuracy for Present subjunctive in Explain & Prompt treatment: Time by Agency
Figure 6.18 Oral production accuracy for Present subjunctive in Explain & Prompt treatment: Time by Agency
3.1.6 Present subjunctive: No Feedback treatment

The results of the one-way ANOVAs conducted on the pretests revealed no significant difference between the [Practice NF] and [Exposure NF] groups for any of the dependent measures, written production, $F(1, 28) = 1.149$, $p = .293$, oral production, $F(1, 28) = .901$, $p = .351$, and grammaticality judgment, $F(1, 28) = .179$, $p = .676$. Again, like for Prepositional relative clauses, the results of the 3x2 repeated measures ANOVAs
revealed that these two control groups experienced no learning on any dependent measure. Specifically, for written production the analysis yielded no significant main effect for Time, $F(2, 56) = 2.482, p = .093, \eta_p^2 = .081$, no significant main effect for Agency, $F(1, 28) = .443, p = .511, \eta_p^2 = .016$, and no significant interaction Time x Agency, $F(2, 56) = 1.586, p = .214, \eta_p^2 = .054$. Similarly, for oral production, the ANOVA showed no significant main effect for Time, $F(2, 56) = .472, p = .626, \eta_p^2 = .017$, no significant main effect for Agency, $F(1, 28) = 1.865, p = .205, \eta_p^2 = .057$, and no significant interaction Time x Agency, $F(2, 56) = 2.614, p = .082, \eta_p^2 = .085$. Finally, the same results were obtained for grammaticality judgment, with no significant main effect for Time, $F(2, 56) = 1.161, p = .321, \eta_p^2 = .04$, no significant main effect for Agency, $F(1, 28) = .127, p = .725, \eta_p^2 = .005$, and no significant interaction Time x Agency, $F(2, 56) = .075, p = .928, \eta_p^2 = .003$ (see Appendix 6.12 for a summary of results and Figures 6.20, 6.21, and 6.22 for a visual representation of mean scores).
Figure 6.20 Written production accuracy for Present subjunctive in No Feedback treatment: Time by Agency
Figure 6.21 Oral production accuracy for Present subjunctive in No Feedback treatment: Time by Agency
**Summary of results for RQ1**

The first research question addressed whether type of learner’s Agency in grammar practice (i.e., participating in interactive grammar practice versus watching others practice) has differential effects on L2 learners’ development of a grammatical structure. Results showed that:

1. If type of Feedback and type of Form are left out of the equation (all three counterparted Feedback groups —Spot & Prompt, Explain & Prompt, No Feedback—
and all two grammatical forms—Prepositional relative clauses, Present subjunctive—are taken together) type of Agency plays no role in L2 development for any of the dependent measures (either written production, oral production, or grammaticality judgment).

However, separate analyses for each of the three Agency dyads in this design revealed that the effects of Agency depended on the nature of the Feedback and the targeted Form (see Table 6.4 below for a summary of significant gains for each group at each point in the timeline, and observed interactions Time x Agency). Specifically:

2. When the learning condition included No Feedback, Agency did not play a role for any structure. Practice by itself did not lead to any learning of any structure on any dependent measure, and Exposure to practice did not either.


4. Finally, when the learning condition included Spot & Prompt Feedback, Agency played a role for Prepositional relative clauses only. [Practice SP] led to significant gains on all dependent measures, whereas [Exposure SP] did not learn at all. The difference was statistically significant for written and oral production, and a similar trend was observed for grammaticality judgment. For Present subjunctive, however, both [Practice SP] and [Exposure SP] learned significantly on all dependent measures
(except [Exposure SP] on grammaticality judgment), and both conditions performed similarly.
Table 6.4 Significant learning gains by group and observed interactions Time x Agency

<table>
<thead>
<tr>
<th></th>
<th>SPOT &amp; PROMPT</th>
<th>EXPLAIN &amp; PROMPT</th>
<th>NO FEEDBACK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pr</td>
<td>Ex</td>
<td>T x A</td>
</tr>
<tr>
<td><strong>Prepositional relative clauses</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written production</td>
<td>A,B</td>
<td>Pr&gt;Ex [T2]</td>
<td>A,B</td>
</tr>
<tr>
<td>Oral production</td>
<td>A,B</td>
<td>Pr&gt;Ex [T3]</td>
<td>A,B</td>
</tr>
<tr>
<td>GJT</td>
<td>A</td>
<td>~ Pr&gt;Ex [T2]</td>
<td>A,B</td>
</tr>
<tr>
<td><strong>Present subjunctive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written production</td>
<td>A,B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Oral production</td>
<td>A</td>
<td>A,B</td>
<td>A,B</td>
</tr>
<tr>
<td>GJT</td>
<td>A,B</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Pr = Practice; Ex = Exposure; T x A = Interaction Time x Agency; A = significant learning from pre to post; B = significant learning from pre to delayed; [T2] = Time 2; [T3] = Time 3; ~ = Trend.*
3.2 Research question #2. Type of Feedback and L2 development

The second research question addressed whether type of oral Feedback provided in response to learners’ productions of a grammatical structure has differential effects on learners’ development of that form, and if so, which of two particular subcomponents of feedback, explicitness of negative evidence and prompting for error repair, may contribute to any observed differences. To investigate this, the present dissertation engaged students in grammar practice with four different types of feedback, as per the combination of two levels of explicitness of negative evidence (whether the error was spotted or explained) and prompting for error repair (whether or not learners were pushed to reformulate their errors), as well as no feedback whatsoever. Figure 6.23 below provides a graphic representation of all experimental groups, with the five groups under examination in frame.

Figure 6.23 Experimental groups under analysis in RQ2

<table>
<thead>
<tr>
<th>Practice SC</th>
<th>Practice SP</th>
<th>Practice EC</th>
<th>Practice EP</th>
<th>Practice NF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exposure SP</td>
<td></td>
<td>Exposure EP</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exposure NF</td>
</tr>
</tbody>
</table>

To answer this research question, two different types of analyses were performed. The first analysis addressed the effects of type of Feedback for the development of the two grammatical structures in this design combined, while the second analysis looked at the effects for each form independently. Following is a report of the first analysis (both forms combined), while the independent analyses for each form are presented in sections 4.1 and 4.2 below.
To investigate the effects of type of Feedback for the development of both grammatical structures combined, four steps were taken. First, the means of the two structures at each of the three time-points were calculated for each participant. Second, one-way ANOVAs were performed on the pretests to identify possible between-group differences at the outset of the study. Third, the raw achievement scores of all five Feedback groups on each of the three assessment tests were entered into a 3x2x5 (Time x Form x Feedback) repeated-measures ANOVA, with two within-subject factors, i.e., Time (pretest, posttest, delayed) and Form (Prepositional relative clauses, Present subjunctive), and one between-subjects factor, Feedback (Spot & Continue, Spot & Prompt, Explain & Continue, Explain & Prompt, No Feedback). Fourth, additional post-hoc Scheffé tests on gain scores were conducted to identify between-group significant differences at each testing moment, where appropriate.

The results of the one-way ANOVAs conducted on the pretests revealed a significant difference between two of the five Feedback groups for written production, $F(4, 80) = 2.925, p = .026$ ([Practice SP] and [Practice NF]), while no significant differences were obtained for either oral production, $F(4, 79) = .647, p = .631$, or grammaticality judgment, $F(4, 80) = .413, p = .799$. Despite the observed differences for written production, the 3x2x5 repeated-measures ANOVA is still a valid procedure to address this research question. The results of this analysis are separately reported below for each dependent measure.

Written production. The 3x2x5 repeated-measures ANOVA performed on the achievement scores of the written production task yielded a significant main effect for
Time, $F(2, 152) = 70.607, p = .000, \eta_p^2 = .482$, a significant main effect for Feedback, $F(4, 76) = 6.010, p = .000, \eta_p^2 = .240$, and a significant interaction Time x Feedback, $F(8, 152) = 4.840, p = .000, \eta_p^2 = .203$ (see Appendix 6.13 for a summary of results). These results, discussed below, are best understood in line with the plotted means in Figure 6.24.

**Figure 6.24** Written production accuracy for both Forms combined: Time by Feedback

The significant main effect for Time indicates that participants learned across time when all Feedback groups were considered together; the significant main effect for
Feedback indicates that group performance was different when Time was left out of the equation; and the significant interaction Time x Feedback shows that groups experienced different learning gains. According to the plot in Figure 6.24, it seems that the [Practice EC] group reached the highest level of achievement, the [Practice NF] group the lowest, and the other three groups performed relatively similarly to each other, in between these two levels of achievement. To investigate whether group performance was significantly different at the time of the immediate and delayed posttests, separate one-way ANOVAS on gain scores were computed. The results revealed a significant difference among groups on both the immediate posttest, $F(4, 80) = 6.987, p = .000$, and the delayed posttest, $F(4, 80) = 4.828, p = .002$. Next, post-hoc Scheffé tests on gain scores were performed at each time point to identify the significant differences between groups. As Table 6.5 shows, when all three time-points (pre, post, delayed) were considered, the [Practice EC] group performed differently from the [Practice NF] and [Practice SC] groups. When only Time 2 was considered, the [Practice EC] group outperformed all others but [Practice EP] (see Table 6.6). At Time 3, the [Practice EC] group outperformed [Practice NF] and [Practice SC] only (see Table 6.7).
Table 6.5 Post-hoc Scheffé on Written production overall scores for both forms by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>2.21</td>
<td>2.21</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>2.50</td>
<td>2.50</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>3.60</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.6 Post-hoc Scheffé on Written production immediate posttest scores for both forms by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>16</td>
<td>.7188</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>18</td>
<td>2.0556</td>
<td></td>
</tr>
<tr>
<td>Practice SP</td>
<td>17</td>
<td>2.3529</td>
<td></td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>3.1471</td>
<td>3.1471</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>5.1765</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.7 Post-hoc Scheffé on Written production delayed posttest scores for both forms by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>16</td>
<td>.8438</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>18</td>
<td>1.6667</td>
<td></td>
</tr>
<tr>
<td>Practice SP</td>
<td>17</td>
<td>2.1471</td>
<td>2.1471</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>2.9412</td>
<td>2.9412</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td></td>
<td>4.6471</td>
</tr>
</tbody>
</table>

*Subset for alpha = 0.05*

Oral production. The 3x2x5 repeated-measures ANOVA performed on the achievement scores of the oral production task yielded a significant main effect for Time, $F(2, 148) = 37.757, p = .000, \eta^2_p = .338$, a significant main effect for Feedback, $F(4, 74) = 3.775, p = .008, \eta^2_p = .169$, and a significant interaction Time x Feedback, $F(8, 148) = 4.684, p = .000, \eta^2_p = .202$ (see Appendix 6.13). The plotted means in Figure 6.25 aid in interpreting these results.
Like for written production, the obtained significant interaction Time x Feedback indicates that there was a between-group difference in terms of learning gains. The plotted means in Figure 6.25 again suggests three levels of achievement, with [Practice EC] and [Practice NF] at the top and bottom and the other three groups in between. The results of two one-way ANOVAs on gain scores revealed a significant difference among groups at both Time 2, $F(4, 78) = 5.628, p = .000$, and Time 3, $F(4, 78) = 4.609, p = .002$. Post-hoc Scheffé tests revealed that when the entire timeline was considered, the [Practice EC] group outperformed the [Practice NF] group (Table 6.8); at Time 2, it also
outperformed [Practice SC] (Table 6.9); and at Time 3 it only outperformed [Practice NF] again (Table 6.10).

**Table 6.8** Post-hoc Scheffé on Oral production overall scores for both forms by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>Practice SP</td>
<td>14</td>
<td>1.36</td>
<td>1.36</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.47</td>
<td>1.47</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>2.34</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.9** Post-hoc Scheffé on Oral production immediate posttest scores for both forms by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>16</td>
<td>.0313</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>18</td>
<td>.7222</td>
<td></td>
</tr>
<tr>
<td>Practice SP</td>
<td>15</td>
<td>1.4000</td>
<td>1.4000</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.6765</td>
<td>1.6765</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>3.1471</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.10 Post-hoc Scheffé on Oral production delayed posttest scores for both forms by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>Subset for alpha = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Practice NF</td>
<td>16</td>
<td>.4375</td>
</tr>
<tr>
<td>Practice SC</td>
<td>17</td>
<td>1.1471</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>1.5313</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.7647</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>3.4412</td>
</tr>
</tbody>
</table>

Grammaticality judgment. The 3x2x5 repeated-measures ANOVA performed on the achievement scores of the grammaticality judgment task yielded a significant main effect for Time, $F(2, 152) = 23.275, p = .000, \eta^2_p = .234$, a significant main effect for Feedback, $F(4, 76) = 7.883, p = .000, \eta^2_p = .293$, and no significant interaction Time x Feedback, $F(8, 152) = 1.885, p = .066, \eta^2_p = .090$ (see Appendix 6.13). Once again, these results are best interpreted in light of the plotted means in Figure 6.26.
The obtained significant main effects for Time and Feedback clearly reveal that there were between differences when feedback and time were excluded from the equation; in other words, learning was observed, and there were differences between feedback groups. However, there was no significant interaction Time x Feedback, although significance was closely approached \( (p = .066) \). The effect size was medium rather than trivial \( (\eta_p^2 = .090) \), indicating that the differences in performance suggested by the plotted means may reach significance upon further data collection.
The next two sections report the results of the analyses conducted for each grammatical form independently. To that aim, the raw achievement scores of all five Feedback groups on each of the three tests were entered into a 3x5 repeated-measures ANOVA, with Time as the within-subject factor and Feedback as the between-subject factor, once for each of grammatical structure under investigation.

### 3.2.1 Prepositional relative clauses

The results of the one-way ANOVAs conducted on the pretests revealed no significant differences between the five Feedback groups for any of the dependent measures, i.e., written production, $F(4, 76) = 2.275, p = .069$, oral production, $F(4, 75) = 1.084, p = .371$, and grammaticality judgment, $F(4, 76) = .279, p = .891$. Consequently, it can be assumed that any gains in scores from pretest to posttest were due to the treatment, rather than to preexisting differences between groups. Following is a separate report of group performance across time for each dependent measure.

*Written production.* The 3x5 repeated-measures ANOVA performed on the achievement scores of the written production task yielded a significant main effect for Time, $F(2, 152) = 60.733, p = .000, \eta^2_p = .444$, a significant main effect for Feedback, $F(4, 76) = 5.891, p = .000, \eta^2_p = .237$, and a significant interaction Time x Feedback, $F(8, 152) = 5.081, p = .000, \eta^2_p = .211$ (see Appendix 6.14 for a summary of results). These results, discussed below, are best understood in line with the plotted means in Figure 6.27.
The main effect for Time indicates that, regardless of Feedback type, participants learned across time. To further elucidate whether or not this was the case for each group independently, a battery of paired sample $t$-tests was computed to measure the progress for Time 1-2, 1-3, and 2-3. Results showed that except the [Practice NF] group, which experienced no gains at any point, all experimental conditions experienced gains from pretest to posttest, and that these gains held up after two weeks in all cases, with no significant increase or loss (see Appendix 6.1).
The obtained significant main effect for Feedback indicates that regardless of time, there were between-group differences. In turn, the obtained significant interaction Time x Feedback shows that there were significantly different learning gains among groups. The plot in Figure 6.27 seems to suggest that the [Practice EC] group outperformed all other groups across time, that the remaining three groups with Feedback performed similarly one level below, and that the [Practice NF] group performed significantly worse than all the other groups. Since all groups were statistically similar at the outset of the study, to investigate group performance on the immediate and delayed posttests two separate one-way ANOVAs on gain scores were conducted. The results revealed a significant difference among groups on both the immediate posttest, $F(4, 76) = 6.570, p = .000$, and the delayed posttest, $F(4, 76) = 4.844, p = .002$. Next, post-hoc Scheffé tests were performed at each time point to identify the significant differences between groups. As Table 6.11 shows, when all three time-points (pre, post, delayed) were considered, only the [Practice EC] group performed differently from the [Practice NF] group. However, when only Time 2 was considered, the two groups with more explicit feedback, [Practice EC] and [Practice EP], outperformed the [Practice NF] group (see Table 6.12). These results did not hold two weeks after the treatment, where again only the [Practice EC] group performed significantly better than the [Practice NF] group (see Table 6.13).
**Table 6.11** Post-hoc Scheffé on Written production overall scores for Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>.31</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>2.04</td>
<td>2.04</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>2.75</td>
<td>2.75</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>2.88</td>
<td></td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td></td>
<td>4.47</td>
</tr>
</tbody>
</table>

**Table 6.12** Post-hoc Scheffé on Written production immediate posttest scores for Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>3.06</td>
<td>3.06</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>3.25</td>
<td>3.25</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td></td>
<td>4.24</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td></td>
<td>6.41</td>
</tr>
</tbody>
</table>
Table 6.13 Post-hoc Scheffé on Written production delayed posttest scores for
Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>2.69</td>
<td>2.69</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>2.94</td>
<td>2.94</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>4.24</td>
<td>4.24</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>5.94</td>
<td></td>
</tr>
</tbody>
</table>

Oral production. The 3x5 repeated-measures ANOVA performed on the achievement scores of the oral production task yielded a significant main effect for Time, $F(2, 148) = 33.636, p = .000, \eta_p^2 = .312$, a significant main effect for Feedback, $F(4, 74) = 4.291, p = .004, \eta_p^2 = .188$, and a significant interaction Time x Feedback, $F(8, 148) = 4.076, p = .000, \eta_p^2 = .181$ (see Appendix 6.14). The plotted means in Figure 6.28 aid to interpret these results.
The main effect found for Time revealed that there was learning for all Feedback groups taken together. However, the results of a battery of paired sample $t$-tests for each group showed that the [Practice NF] group did not experience any learning, that the [Practice SC] group only learned significantly from immediate to delayed posttest, and that all other groups with Feedback experienced learning from pretest to immediate posttest and maintained those gains two weeks after the treatment (see Appendix 6.2 for a summary).
The main effect obtained for Feedback means that there was a difference between groups when considering pretest, posttest, and delayed posttest means altogether. Additionally, the obtained interaction Time x Feedback indicates that there was a between-group difference in terms of learning gains. Figure 6.28 suggests that similarly to the results for written production, the [Practice EC] group performed the best, all other Feedback groups performed similarly one level below, and the [Practice NF] group performed the worst. Again, since there were no differences between groups on the pretest, two one-way ANOVAs were conducted on gain scores to measure possible differences on the immediate and delayed posttests. The results revealed a significant difference among groups on both tests, immediately after the treatment, \( F(4, 74) = 4.610, p = .002 \), and two weeks later, \( F(4, 74) = 4.490, p = .003 \). Next, post-hoc Scheffé tests were performed to identify the significant differences between groups. As Tables 6.14, 6.15, and 6.16 show, only the [Practice EC] group performed statistically differently from the [Practice NF] group, both when considering the entire timeline or Times 2 and 3 independently.
Table 6.14 Post-hoc Scheffé on Oral production overall scores for Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>Practice SP</td>
<td>14</td>
<td>1.21</td>
<td>1.21</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.47</td>
<td>1.47</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td></td>
<td>2.59</td>
</tr>
</tbody>
</table>

Table 6.15 Post-hoc Scheffé on Oral production immediate posttest scores for Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>-.13</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>1.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Practice SP</td>
<td>14</td>
<td>1.43</td>
<td>1.43</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.88</td>
<td>1.88</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td></td>
<td>3.47</td>
</tr>
</tbody>
</table>
Table 6.16 Post-hoc Scheffé on Oral production delayed posttest scores for Prepositional
relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>-.07</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>1.94</td>
<td>1.94</td>
</tr>
<tr>
<td>Practice SP</td>
<td>14</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>2.18</td>
<td>2.18</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td></td>
<td>4.29</td>
</tr>
</tbody>
</table>

Grammaticality judgment. The 3x5 repeated-measures ANOVA performed on the
achievement scores of the grammaticality judgment task yielded a significant main effect
for Time, \(F(2, 152) = 22.055, p = .000, \eta^2_p = .225\), a significant main effect for Feedback,
\(F(4, 76) = 6.999, p = .000, \eta^2_p = .269\), and a significant interaction Time x Feedback, \(F(8, 152) = 2.809, p = .006, \eta^2_p = .129\) (see Appendix 6.14). Once again, these results are best
interpreted in light of the plotted means in Figure 6.29.
The main effect obtained for Time reveals that when taking together the means of all groups, learning was observed. To refine this finding, a series of paired sample $t$-tests for each time interval showed that the two groups with Explain Feedback, either [Practice EP] or [Practice EC], experienced gains on the immediate posttest and maintained them on the delayed posttest. On the other hand, the two groups with Spot Feedback, [Practice SC] and [Practice SP], learned from pretest to posttest, but the differences between
pretest and delayed posttest lost significance. Finally, the [Practice NF] group experienced no gain at any point (see Appendix 6.3 for a summary).

The main effect obtained for Feedback, which indicates between-group differences when combining means regardless of time, is best interpreted based on the obtained significant interaction Time x Feedback. The plot in Figure 6.29 suggests that similarly to oral and written tests, the [Practice EC] group performed the best, all other Feedback groups performed similarly to each other one level below, and the [Practice NF] group performed the worst. To find out whether these differences were statistically significant at the time of the posttest and delayed posttest, two one-way ANOVAs on gain scores were conducted. Results showed a significant difference among groups on the immediate posttest, $F(4, 76) = 3.718, p = .008$ and delayed posttest, $F(4, 76) = 2.505, p = .049$. Subsequent post-hoc Scheffé tests indicated that over the entire timeline three groups performed significantly differently from the [Practice NF] group, namely, the two Prompt groups plus [Practice EC] (see Table 6.17). However, independent post-hoc Scheffé tests on gain scores revealed that only the [Practice EC] group significantly outperformed the [Practice NF] group at Time 2 (see Table 6.18), and approached significance at Time 3.
Table 6.17 Post-hoc Scheffé on Grammaticality judgment overall scores for Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>Subset for alpha = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>-.51</td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>.94  .94</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>1.10</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.18</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>2.04</td>
</tr>
</tbody>
</table>

Table 6.18 Post-hoc Scheffé on Grammaticality judgment immediate posttest scores for Prepositional relative clauses by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>Subset for alpha = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>-.60</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>1.75  1.75</td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>2.00</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>2.06</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>2.88</td>
</tr>
</tbody>
</table>

4.2.2 Present subjunctive

The results of the one-way ANOVAs conducted on the pretests revealed no significant differences between the five groups for any of the dependent measures, i.e., written production, $F(4, 80) = 1.951, p = .110$, oral production, $F(4, 79) = .419, p = .794$, and grammaticality judgment, $F(4, 80) = 1.427, p = .233$. Consequently, it can be assumed that any gains in scores from pretest to posttest were due to the treatment, rather than to
preexisting differences between groups. Following are the results of the 3x5 repeated measures ANOVA for each dependent measure.

**Written production.** The 3x5 repeated-measures ANOVA performed on the achievement scores of the written production task yielded a significant main effect for Time, $F(2, 160) = 28.841, p = .000, \eta^2_p = .265$, a significant main effect for Feedback, $F(4, 80) = 2.882, p = .028, \eta^2_p = .126$, and no significant interaction Time x Feedback, $F(8, 160) = 1.779, p = .085, \eta^2_p = .082$ (see Appendix 6.15). These results are best interpreted based on the plotted means in Figure 6.30.
The main effect for Time means that overall, participants learned. To probe deeper into this finding, a battery of paired sample $t$-tests for each group at each time-interval was conducted. Results revealed that both the No Feedback group and the Spot & Continue groups experienced no learning whatsoever. On the other hand, all other groups learned at Time 2 and held those gains at Time 3, except the [Practice EP] group (see Appendix 6.4). Additionally, the observed main effect for Feedback indicates that, when the means of each group for the entire timeline are collapsed, between-group differences are observed. On this note, the plotted means in Figure 6.30 suggest that the [Practice
EC] group performed better than [Practice SC] and [Practice NF]. However, the lack of a significant interaction Time x Feedback indicates that learning gains were not significantly different. This is quite possibly due to the low sample size, since the obtained $\eta^2_p$ figures for the non-significant interaction Time x Feedback suggest a medium rather than trivial effect size of the treatment. Interestingly, when a less conservative scoring protocol is applied and participants are not penalized for using the subjunctive in non-targeted items, a significant interaction Time x Feedback is observed, $F(8, 160) = 2.044, p = .045$, and the [Practice EC] group approaches a significant difference from the [Practice NF] and [Practice SC] groups in a post-hoc Scheffé.

**Oral production.** The 3x5 repeated-measures ANOVA performed on the achievement scores of the oral production task yielded a significant main effect for Time, $F(2, 154) = 16.900, p = .000, \eta^2_p = .180$, no significant main effect for Feedback, $F(4, 77) = 1.437, p = .230, \eta^2_p = .069$, and a significant interaction Time x Feedback, $F(8, 154) = 2.185, p = .031, \eta^2_p = .102$ (see Appendix 6.15). Once again the plotted means in Figure 6.31 aid in the interpretation of these findings.
The main effect for Time indicates that when type of Feedback is left out of the equation, participants learned as a result of the treatments. As shown in Appendix 6.5, the results of a battery of paired \( t \)-tests revealed that participants in the [Practice NF] and [Practice SC] groups experienced no learning, whereas all others did. Additionally, the significant interaction Time x Feedback indicates a difference in the learning gains experienced by the groups. Since all groups performed statistically similarly on the pretest, two separate one-way ANOVAs using gain scores were computed to determine whether the groups performed differently on both immediate and delayed posttests.
Results revealed that the difference was significant on the immediate posttest, $F(4, 77) = 3.005, p = .023$, rather than the delayed posttest, $F(4, 77) = 2.326, p = .064$. Post-hoc Scheffé tests revealed that the difference surfaced at Time 2, with [Practice EC] outperforming the [Practice NF] group (see Table 6.19).

Table 6.19 Post-hoc Scheffé on Oral production immediate posttest scores for Present subjunctive by Feedback

<table>
<thead>
<tr>
<th>Feedback</th>
<th>N</th>
<th>Subset for alpha = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice NF</td>
<td>15</td>
<td>-0.06</td>
</tr>
<tr>
<td>Practice SC</td>
<td>17</td>
<td>0.76</td>
</tr>
<tr>
<td>Practice SP</td>
<td>16</td>
<td>1.40</td>
</tr>
<tr>
<td>Practice EP</td>
<td>16</td>
<td>1.47</td>
</tr>
<tr>
<td>Practice EC</td>
<td>17</td>
<td>2.82</td>
</tr>
</tbody>
</table>

**Grammaticality judgment.** The 3x5 repeated-measures ANOVA performed on the achievement scores of the grammaticality judgment task yielded a significant main effect for Time, $F(2,160) = 7.221, p = .001, \eta_p^2 = .083$, a significant main effect for Feedback, $F(4, 80) = 2.855, p = .029, \eta_p^2 = .125$, and no significant interaction Time x Feedback, $F(8,160) = .468, p = .878, \eta_p^2 = .023$ (see Appendix 6.15). The plotted means in Figure 6.32 help interpret these findings.
The obtained main effect for Time implies that there was learning regardless of Feedback type. However, paired $t$-tests for each time interval revealed that the only groups that experienced significant learning were [Practice SP] and [Practice EC] (see Appendix 6.6). The obtained main effect for Feedback reveals that regardless of time, groups performed differently from each other. As the graph shows, this may be the result of carry-over effects arising from non significant differences on the pretest. The no
significant interaction Time x Feedback elucidates that there were no significant differences between groups in terms of learning gains.

**Summary of results for RQ2**

The second research question addressed whether type of oral Feedback provided in response to learners’ productions of a grammatical structure has differential effects on learners’ development of that form, and if so, which of two particular subcomponents of feedback, explicitness of negative evidence and prompting for error repair, may contribute to any observed differences. Results showed that:

1. When both forms were considered together, a significant interaction Time x Feedback was obtained for written and oral production. In both tests, the [Practice EC] and [Practice NF] groups were the highest and lowest achievers, respectively. For written production, [Practice EC] outperformed all other groups but [Practice EP] on the immediate posttest, while two weeks later it only outperformed [Practice NF] and [Practice SC]. For oral production, [Practice EC] outperformed [Practice NF] and [Practice SC] on the immediate posttest, while it only outperformed [Practice NF] two weeks later. For grammaticality judgment, a significant interaction Time x Feedback was approached, with a clear trend towards an edge of [Practice EC] over [Practice NF].

Independent analyses by type of Form revealed that the edge of the [Practice EC] group appeared for Prepositional relative clauses more clearly than for the Present subjunctive (see Table 6.20 for a summary of significant gains for each group at any point in the timeline, and observed interactions Time x Feedback). Specifically:
2. For Prepositional relative clauses, the [Practice NF] group did not experience any learning, while all groups with Feedback did. [Practice EC] outperformed the [Practice NF] group on all three dependent measures at all test times, except for the delayed grammaticality judgment posttest, where significance was approached. Only one additional significant difference was found, with [Practice EP] outperforming [Practice NF] on the immediate written production posttest.

3. For Present subjunctive, the [Practice NF] and [Practice SC] groups experienced no significant learning, while all others did. However, the only observed difference was between the [Practice EC] and [Practice NF] groups, with the former outperforming the latter on oral production and approaching significance on written production.
Table 6.20 Significant learning gains by group and observed interactions Time x Feedback

<table>
<thead>
<tr>
<th></th>
<th>Practice</th>
<th>Practice</th>
<th>Practice</th>
<th>Practice</th>
<th>Practice</th>
<th>T x F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SC</td>
<td>SP</td>
<td>EC</td>
<td>EP</td>
<td>NF</td>
<td></td>
</tr>
</tbody>
</table>

Prepositional RCs


Present subjunctive

| Written production    | A,B       | A,B       | A         |          |          | ~[T2] EC > NF, SC             |
| Oral production       | A         | A,B       | A,B       |          |          | [T2] EC > NF                  |
| GJT                   | A,B       | A         |           |          |          |                              |

Note. TxA = Interaction Time x Feedback; [T2] = Time 2; [T3] = Time 3; ~ = Trend.

3.3 Research question #3. Type of grammatical form and L2 development

The third research question addressed whether type of grammatical Form (i.e., more vs. less complex) has differential effects on L2 development as a result of a pedagogical intervention. Since in the present dissertation all eight experimental groups received treatment for both targeted structures, Spanish Present subjunctive and Prepositional relative clauses, all groups were considered in the analysis (see Figure 6.33).
Specifically, the raw scores of all eight groups for the three dependent variables, written production, oral production, and grammaticality judgment, were submitted to a 3x2 (Time x Form) repeated-measures ANOVA, where Time (pretest, immediate posttest, delayed posttest) was entered as the dependent factor and Form was entered as the independent factor (see Appendix 6.16 for a summary of results).

Results for the written production scores revealed a significant main effect for Time, $F(2, 494) = 96.260, p = .000, \eta_p^2 = .280$, no significant main effect for Form, $F(1, 247) = 3.557, p = .060, \eta_p^2 = .014$, and a significant interaction Time x Form, $F(2, 494) = 4.011, p = .019, \eta_p^2 = .016$. This latter significant interaction indicates that overall, when all groups were considered (including the No Feedback groups, which experienced zero learning of either structure) students achieved higher learning gains for one of the two structures. As Figure 6.34 below shows, the winning structure was Prepositional relative clauses.
In turn, results of the ANOVA for oral production tests showed a significant main effect for Time, $F(2, 484) = 58.887, p = .000$, $\eta^2 = .196$, no significant main effect for Form, $F(1, 242) = .559, p = .455$, $\eta^2 = .002$, and no significant interaction Time x Form, $F(2, 484) = 1.589, p = .205$, $\eta^2 = .007$. These results contrast with those obtained for written production, since in this case students from all eight groups achieved similar overall learning gains for both targeted structures (see plotted means in Figure 6.35).
Finally, results for the grammaticality judgment tests showed a significant main effect for Time, $F(2, 494) = 34.618$, $p = .000$, $\eta^2_p = .123$, no significant main effect for Form, $F(1, 247) = 1.881$, $p = .171$, $\eta^2_p = .008$, and a significant interaction Time x Form, $F(2, 494) = 3.327$, $p = .037$, $\eta^2_p = .013$. In line with the results for written production, this significant interaction indicates that students as a whole learned Prepositional relative clauses significantly better than Present subjunctive (see Figure 6.36 for plotted means).
In sum, then, the results of the 3x2 repeated-measures ANOVA showed that when all eight experimental groups were taken together, students experienced significantly higher learning of Prepositional relative clauses than Present subjunctive on written production and grammaticality judgment measures, whereas for oral production the observed learning of the two structures was statistically similar. Now, two points should be considered. First, the observed mean scores for all three tests are lowered as a result of the fact that this analysis included the two control groups, [Practice NF] and [Exposure
NF]; and second, this analysis does not show whether the observed differences occurred across various experimental groups or were circumscribed to a particular group.

To refine these findings, independent analyses for every group were conducted, by submitting the raw scores of each group to separate 3x2 (Time x Form) repeated-measures ANOVAs, with Time as the dependent factor and Form as the independent factor (see Appendix 6.17). These analyses revealed that the observed interactions Time x Form occurred only in the two Practice groups that received Explain feedback, [Practice EC] and [Practice EP].

Specifically, the ANOVA for [Practice EC] yielded a significant interaction Time x Form for two dependent measures, written production, $F(2, 64) = 3.528, p = .035, \eta^2_p = .099$ (see Figure 6.37) and grammaticality judgment, $F(2, 64) = 3.321, p = .042, \eta^2_p = .094$ (see Figure 6.38). As for the [Practice EP] group, the ANOVA revealed an almost significant interaction Time x Form for written production, $F(2, 64) = 3.116, p = .051, \eta^2_p = .089$ (see Figure 6.39), and a significant interaction for grammaticality judgment, $F(2, 64) = 3.805, p = .027, \eta^2_p = .106$ (see Figure 6.40). As the plotted means below show, in all cases participants’ learning gains for Prepositional relative clauses were significantly higher than for Present subjunctive.
Figure 6.37 Written production accuracy in Practice EC: Time by Form
Figure 6.38 Grammaticality judgment accuracy in Practice EC: Time by Form
Figure 6.39 Written production accuracy in Practice EP: Time by Form
Although no significant interaction Time x Form was found in the Practice groups with Spot feedback, as opposed to Explain feedback, similar trends were detected for both written production and grammaticality judgment tasks. Specifically, the [Practice SC] group experienced significant learning of Prepositional relative clauses but no significant learning gains were observed for Present subjunctive (see Figures 6.41 and 6.42). In turn, the [Practice SP] group experienced learning gains for both structures on both dependent measures, but again the means were higher for Prepositional relative clauses (see Figures 6.43 and 6.44).
Figure 6.41 Written production accuracy in Practice SC: Time by Form
Figure 6.42 Grammaticality judgment accuracy in Practice SC: Time by Form
**Figure 6.43** Written production accuracy in Practice SP: Time by Form

![Graph showing written production accuracy over time for different forms.](image-url)
Finally, in the Exposure to practice groups there were no clear trends for written production and grammaticality judgment. The only significant learning gains in the [Exposure SP] group occurred for Present subjunctive on the written production task, and in both measures the means for this structure were slightly higher than those obtained for Prepositional relative clauses (see Figures 6.45 and 6.46). In turn, the [Exposure EP] group experienced similar gains for both structures on written production, but again in line with the Practice groups the gains for Prepositional relative clauses were higher than for Present subjunctive (Figures 6.47 and 6.48).
Figure 6.45 Written production accuracy in Exposure SP: Time by Form
Figure 6.46 Grammaticality judgment accuracy in Exposure SP: Time by Form
Figure 6.47 Written production accuracy in Exposure EP: Time by Form
Summary of results for RQ3

The third research question addressed whether type of linguistic Form (i.e., more vs. less complex) has differential effects on L2 development as a result of a pedagogical intervention. Results showed that:

1. When all eight experimental groups were taken together, students experienced significantly higher learning of Prepositional relative clauses than Present subjunctive on written production and grammaticality judgment measures, whereas for oral production the observed learning of the two structures was statistically similar.
Subsequent independent analyses performed on each group provided more refined results (see Table 6.21 below for a summary). Specifically:

2. The obtained interaction Time x Form originated in the Practice groups with Explain Feedback, both [Practice EC] and [Practice EP], on both written production and grammaticality judgment. A similar trend, however, was observed for the other Practice groups with Spot Feedback, where students also experienced higher learning gains for Prepositional relative clauses on written production and grammaticality judgment.

3. The Exposure to practice groups did not follow this trend. [Exposure SP] experienced no significant learning except for Present subjunctive on written production, and the obtained means for this structure were slightly higher. In turn, [Exposure EP] obtained comparable gains for both structures on written production, whereas in grammaticality judgment a trend towards higher learning for Prepositional relative clauses was observed.

4. Finally, the No Feedback groups experienced no learning for either structure.
## Table 6.21 Significant learning gains by group and observed interactions Time x Form

<table>
<thead>
<tr>
<th>Written production</th>
<th>Oral production</th>
<th>Grammaticality judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>S</td>
<td>T x F</td>
</tr>
<tr>
<td>Practice SC</td>
<td>A,B</td>
<td></td>
</tr>
<tr>
<td>Practice SP</td>
<td>A,B</td>
<td>A,B</td>
</tr>
<tr>
<td>Practice EC</td>
<td>A,B</td>
<td>A,B</td>
</tr>
<tr>
<td>Practice EP</td>
<td>A,B</td>
<td>A</td>
</tr>
<tr>
<td>Practice NF</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure SP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure EP</td>
<td>A,B</td>
<td>A,B</td>
</tr>
<tr>
<td>Exposure NF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* P = Prepositional relative clauses; S = Present subjunctive; T x F = Interaction Time x Form; A = significant learning from pre to post; B = significant learning from pre to delayed; [T2] = Time 2; [T3] = Time 3.
3.4 Research question #4. Interactions between Agency, Feedback, Form, and Time

The fourth research question addressed whether the variables investigated in this dissertation — type of learner’s Agency, type of oral Feedback, and type of grammatical Form — work in concert towards different L2 learning outcomes, as measured by three tests (written production, oral production, and grammaticality judgment), immediately after the treatment and two weeks later.

To answer this question, the raw scores of the six counterparted cells in this design (see Figure 6.49 below) on all three dependent variables were separately submitted to a 3x2x2x3 (Time x Form x Agency x Feedback) repeated-measures analysis of variance (ANOVA), using a two within-subject, two between-subject design. The within-subject factors were Time (pretest, immediate posttest, delayed posttest) and Form (Prepositional relative clauses, Present subjunctive), since all participants completed all tests for both targeted structures. In turn, the between-subject factors were Agency (Practice, Exposure to practice) and Feedback. For this latter variable, only three levels out of five, those of the counterparted treatments, were entered (Spot & Prompt, Explain & Prompt, No Feedback). Naturally, this must be taken into consideration when discussing the results.

**Figure 6.49** Experimental groups under analysis in RQ4

<table>
<thead>
<tr>
<th>Practice SC</th>
<th>Practice SP</th>
<th>Practice EC</th>
<th>Practice EP</th>
<th>Practice NF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure SP</td>
<td>Exposure SP</td>
<td>Exposure EP</td>
<td>Exposure EP</td>
<td>Exposure NF</td>
</tr>
</tbody>
</table>

The ANOVA performed on the achievement scores of the *written production task* yielded a significant main effect for Time, $F(2, 166) = 45.554, p = .000, \eta^2 = .354,$ a
significant main effect for Feedback, $F(2, 83) = 11.602, p = .000, \eta^2 = .218$, a significant interaction Time x Feedback, $F(4, 166) = 7.337, p = .000, \eta^2 = .150$, and a significant quadruple interaction Time x Form x Agency x Feedback, $F(4, 166) = 2.493, p = .045, \eta^2 = .057$ (see Appendix 6.18 for a summary). This quadruple interaction indicates that among the six experimental groups compared by the ANOVA, between-group differences arose for particular combinations of Form (either Prepositional relative clauses or Present subjunctive), Agency (either Practice or Exposure), and Feedback (Spot & Prompt, Explain & Prompt, No Feedback).

The ANOVA for the oral production task confirmed these results, yielding also a significant main effect for Time, $F(2, 162) = 29.136, p = .000, \eta^2 = .265$, a significant main effect for Feedback, $F(2, 81) = 9.420, p = .000, \eta^2 = .189$, a significant interaction Time x Feedback, $F(4, 162) = 7.083, p = .000, \eta^2 = .149$, and a significant quadruple interaction Time x Form x Agency x Feedback, $F(4, 162) = 2.509, p = .044, \eta^2 = .058$ (see Appendix 6.19 for a summary of results).

Finally, the ANOVA on the grammaticality judgment task yielded a significant main effect for Time, $F(2, 166) = 16.462, p = .000, \eta^2 = .166$, a significant main effect for Feedback, $F(2, 83) = 17.045, p = .000, \eta^2 = .291$, a significant interaction Agency x Feedback, $F(2, 83) = 3.986, p = .022, \eta^2 = .088$, a significant interaction Time x Feedback, $F(4, 166) = 5.531, p = .000, \eta^2 = .118$, and a significant triple interaction Time x Form x Feedback, $F(4, 166) = 4.229, p = .003, \eta^2 = .092$ (see Appendix 6.20 for a summary of results). This latter triple interaction indicates that regardless of Agency there were significantly different learning gains among the Feedback groups, but only for one of the grammatical Forms. Also, the fact that no quadruple interaction was obtained
here, unlike for written and oral production, reveals that for the grammaticality judgment task the two levels of Agency did not contribute to significantly different learning gains for any combination of Form and Feedback.

To sum up, then, the answer to RQ4 reads as follows:

RQ4(a) Time x Agency x Feedback: Does type of oral Feedback moderate the effects of type of learner’s Agency on L2 development (or vice versa)? No, Feedback and Agency did not interact with Time in the present research design when the two grammatical Forms under examination were considered together.

RQ4(b) Time x Form x Agency: Does type of grammatical Form moderate the effects of type of learner’s Agency on L2 development? No, Form and Agency did not interact with Time when all three Feedback levels in this analysis (Spot & Prompt, Explain & Prompt, and No Feedback) were considered together.

RQ4(c) Time x Form x Feedback: Does type of grammatical Form moderate the effects of type of oral Feedback on L2 development? Yes, Form and Feedback interacted with Time when all two Agency levels (Practice, Exposure) were considered together, but only in the grammaticality judgment task.

RQ4(d) Time x Form x Agency x Feedback: Does type of grammatical Form moderate the effects of a pedagogical treatment with different types of learners’ Agency and oral Feedback on L2 development? Yes, Form, Agency, and Feedback interacted with Time in both written and oral production tasks, but not in the grammaticality judgment task.

In conclusion, the significant quadruple interaction Time x Form x Agency x Feedback obtained for both written and oral production tasks reveals that at least one
particular combination of Form (Prepositional relative clauses, Present subjunctive), Agency (Practice, Exposure to practice), and Feedback (Spot & Prompt, Explain & Prompt, No Feedback) yielded higher learning gains than another. As seen in section 3, this was the case of the Spot & Prompt treatment for Prepositional relative clauses, were Practice yielded significantly greater gains than Exposure to practice.

4. Exit questionnaire

The present section reports the results of the statistical analyses performed on the Likert-scale questions of the exit questionnaire, where participants evaluated different aspects of the e-tutor, Talking to Avatars (see Appendix 6.21). All analyses were performed on 118 out of a total of 128 collected questionnaires, since 10 of them were lost and could not be retrieved. For all items, the minimum and maximum possible scores were 1 and 5 points, respectively. Appendix 6.22 summarizes the means and standard deviations for each question by group.

To analyze the data, questions were categorized into two blocks. Questions 8, 11, 12, 14-17, and 23-26 addressed general aspects of the software (e.g., overall evaluation, user-friendliness, amount of practice, content ecology, rating of the e-tutor vis-à-vis classroom based instruction, etc.), and consequently all eight experimental groups were considered in the analysis. On the other hand, questions 18-21, and 27 specifically addressed several aspects of the feedback received (comprehensibility, usefulness, adequacy of explicitness, and adequacy of modality), hence only the four Practice groups with Feedback were included in the analysis. In all cases, a one-way ANOVA was conducted, with participants’ Likert-scale raw scores as the dependent factor and experimental group as
the independent factor. As shown in Appendix 6.22, the results of the one-way ANOVA yielded a significant $p$ value for four questions, Question 14, $F(7, 110) = 2.553, p = .018$, Question 19, $F(3, 55) = 3.287, p = .027$, Question 20, $F(3, 53) = 3.146, p = .033$, and Question 26, $F(7, 110) = 7.054, p = .000$.

Question 14 asked students whether they thought they had learned some Spanish by taking part in the experiment, rating their perceptions on a scale of 1 to 5 (respectively, “I didn’t learn anything at all” and “I learned a lot”). The average score of all eight groups was 3.14, and the highest and lowest scores were 3.75 and 2.86, respectively (see Appendix 6.22 for descriptive statistics and Figure 6.50 below for a means bar-plot). To determine which groups were responsible for the significant ANOVA, $F(7, 110) = 2.553, p = .018$, a post-hoc Scheffé was conducted. This Scheffé, however, did not yield any significant contrasts, likely due to the fact that when comparisons are established between two groups only, power is reduced. The closest significant values, however, were obtained for the comparisons between the group with the highest mean, Practice with Explain & Continue, and those with the lowest, Practice with Explain & Prompt ($p = .204$) and Practice with No Feedback ($p = .229$). These results are interesting because they reveal that the group that reported the highest perception of learning was the actual top gainer in terms of L2 development, Practice with Explain & Continue. Additionally, the fact that Explain & Prompt reported an almost significantly lower perception of learning is congruent with its actual learning outcomes. This group did not parallel the means of its Explain sibling and actually performed on par with the Spot feedback groups.
Question 19 asked students in the Practice with Feedback condition whether they thought that the corrective feedback they had received was useful for learning, on a scale of 1 to 5 (respectively, “It wasn’t useful at all” and “It was extremely useful”). The average score of the four Practice groups was 3.90, thus approaching point 4 on the scale, defined as “It was pretty useful” (see Figure 6.51 below for a means bar-plot). A post-hoc Scheffé yielded a significant contrast ($p = .042$) between the highest and lowest scorers, Explain & Continue and Explain & Prompt. Interestingly, once again the highest achievers also ranked highest in their perceptions of how useful the feedback was, with a
mean score of 4.60, close to point 5 on the scale, defined as “It was extremely useful.” On the other hand, the students in the Explain & Prompt group produced a mean score of 3.5, between “It was ok” and “It was pretty useful.” This is interesting because both groups received essentially the same feedback in terms of explicitness, and the only difference was the absence or presence of Prompts.

Figure 6.51 If you received feedback, do you think it was useful for learning?
À propos of explicitness, Question 20, $F(3, 53) = 3.146, p = .033$, asked students whether they were satisfied with the type of feedback they received in terms of its level of detail, or whether they would have preferred to receive a different type of feedback, with 1 being “I was very unsatisfied with the level of detail” and 5 being “I loved the level of detail.” On average, the four Practice groups produced a mean score of 3.18 (“It was ok”) (see Figure 6.52 below for a means bar-plot). The highest scorer was once again Explain & Continue (3.73), and the lowest scorers were Explain & Prompt and Spot & Continue (2.87 each). Although the ANOVA yielded a significant $p$ value, a post-hoc Scheffé only approached two significant contrasts, precisely between the highest and two lowest scorers ($p = .084$ in both cases). These findings evidence once more that the Explain & Prompt group seemed to be less satisfied with the feedback received than the Explain & Continue group, even if it was the exact same feedback in terms of explicitness. Furthermore, the Prompt group reported the exact same level of satisfaction than the Spot & Continue group, which only got errors pinpointed but never got an explanation of their nature.
Finally, in relation to the issue of Agency, Question 26, $F(7, 110) = 7.054, p = .000$ asked students to answer whether they liked their group or, rather, they would have preferred to switch to a different one, based on the fact that some students got to interact with the avatars while other students got to watch those interactions. All Practice groups rated above the middle point of 3, while all Exposure groups scored 3 or less. The Explain & Continue group was again the most satisfied, with a mean of 4.13 (“No, I prefer the group I fell in”), while the lowest mean was for the Exposure with Explain & Prompt feedback group (see Appendix 6.20 for descriptive statistics and Figure 6.53).
below for a means bar-plot). The results of a post-hoc Scheffé showed many contrasts (see Table 6.22 below). Specifically, all Practice groups except Explain & Prompt reported significantly higher satisfaction with their assigned Agency than two of the Exposure groups, Explain & Prompt and No Feedback. Additionally, the ‘winning’ group, Practice with Explain & Continue, approached a third significant contrast with the remaining Exposure group, Spot & Continue.

Figure 6.53 How did you like your group in terms of Agency?
Table 6.22 Post-hoc Scheffé on Question 26

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure EP</td>
<td>12</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td>Exposure NF</td>
<td>13</td>
<td>2.69</td>
<td></td>
</tr>
<tr>
<td>Exposure SP</td>
<td>15</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Practice NF</td>
<td>14</td>
<td>3.43</td>
<td>3.43</td>
</tr>
<tr>
<td>Practice EP</td>
<td>17</td>
<td>3.65</td>
<td>3.65</td>
</tr>
<tr>
<td>Practice SP</td>
<td>15</td>
<td>3.93</td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>16</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>Practice EC</td>
<td>16</td>
<td>4.13</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 7.
DISCUSSION

The present chapter is organized as follows. Sections 1 to 4 discuss the results for each of the four research questions under investigation, i.e., the incidence of type of learner’s Agency, type of corrective oral Feedback, type of grammatical Form, and their combined interactions, in L2 development. Section 5 presents the conclusion of this dissertation and its implications. Section 6 discusses the study’s limitations and makes suggestions for future research.

1. Research question #1. Type of learner’s Agency and L2 development
The first research question addressed whether type of learner’s Agency in grammar practice (i.e., participating in interactive grammar practice versus watching others practice) has differential effects on L2 learners’ development of a grammatical structure. Results showed that when all two grammatical forms in this study (Present subjunctive and Prepositional relative clauses) and all three counterparted Feedback groups (Spot & Prompt, Explain & Prompt, No Feedback) were considered, significant L2 development was observed for all three dependent measures (written production, oral production, grammaticality judgment), but Agency played no role in any case. In other words, the learning gains experienced by participants who engaged in grammar practice and those who observed them were comparable. However, independent analyses for each of the three Feedback dyads by type of Form investigated here revealed a more complex picture. The results of these analyses are discussed below in separate sections.
Agency in Spot & Prompt treatments

When the learning condition included Spot & Prompt Feedback, learner’s Agency played a role in L2 development, but only for Prepositional relative clauses. Specifically, Practice led to significant learning of Prepositional relative clauses in all dependent measures, whereas Exposure to practice did not. This difference was significant for written and oral production, and a similar trend was observed for grammaticality judgment. For Present subjunctive, however, Agency played no role in L2 development. Participants who practiced learned significantly in all dependent measures, but their observers learned too (except for grammaticality judgment), and there was no statistically significant difference between the two conditions. The results for each grammatical structure are separately discussed below.

To gain a better insight into the edge of Practice for Prepositional relative clauses, the learning gains of each dyad member in the immediate written production posttests were re-coded. Learning was operationalized as gaining 3 or more items, and the written production test was chosen for illustrative purposes over the other two because it was the one that most closely reproduced the nature of the treatment task. As Table 7.1 below illustrates, in 14 Performer-Observer (P-O) dyads, only 1 observer derived learning from watching a performer who did (there were 8 of them). This is particularly striking because some performers provided many instances of positive evidence and successful reformulations in the treatment dialogs, and still observers did not seem to benefit from it despite high rates of attention to task completion (92.85% for the two structures combined, that is, 9/10 questions answered correctly). For example, one performer provided the correct targeted form for Prepositional relative clauses in 8 out of 10
obligatory contexts in the treatment, either on her first attempt or after a correct
reformulation. This participant scored 10%, 90%, and 80% accuracy on the written
pretest, immediate-, and delayed posttest, respectively. In turn, her observer, despite
answering all 10 check-up questions correctly, did not improve his initial pretest score of
zero in any of the tests.

**Table 7.1** Written production gains for Prepositional RCs in Spot & Prompt dyads

<table>
<thead>
<tr>
<th>Form</th>
<th>P learned, O did not</th>
<th>P learned, O learned</th>
<th>P did not learn, O learned</th>
<th>No-one learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepositional RCs</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

*Note. P = Performer; O = Observer*

The obtained edge of Practice over Exposure to practice for Prepositional relative
clauses is in line with Mackey (1999), who found that, in treatments without provision of
grammatical explanation, participants who had an opportunity to negotiate for meaning
with a native speaker advanced significantly in the developmental sequence of English
question formation, while their observers did not experience any significant learning
despite attentional levels of 98%. Additional support for these findings is provided by
Hsieh (2007), where Practice outperformed Exposure to practice in oral production. In
her study, however, the observers did experience significant learning—they just learned
significantly less. On the other hand, counterevidence for the edge of Practice is found in
both Muranoi (2000) and Hsieh (2007). Muranoi found that even when grammar was not
explained, observers were able to derive learning, to a comparable level, from watching
their peers completing the treatment. Similarly, Hsieh found that observers learned significantly as much as performers in written recognition and production.

The fact that the observers in the less explicit condition in Hsieh’s (2007) study did experience significant learning for all dependent measures, unlike Mackey (1999) and the present study for Prepositional relative clauses, can be likely due to the different nature of her task. In the present study, participants who practiced had to write prepositional relative clauses in a very controlled activity (constrained selected response), and in Mackey (1999) they had to produce English questions orally under less controlled conditions. That is, despite the differences in modality and task-essentialness, the tasks in these two studies were output-focused. In contrast, Hsieh used an input-focused task in the written mode, where students had to select options from two or three choices displayed on the computer screen. Arguably, in output-focused tasks the “experiential” gap between performing and observing is wider than in input-focused tasks. In the former, performers have to select a target form from many possible choices —as many as their interlingua permits—, and the same applies when they reformulate their ill-formed productions in response to corrective feedback. From the exposure side, any observer who decides to complete the task subvocally may choose different targeted structures, and the feedback provided to the performer s/he observes may or may not be informative to his or her choice. Conversely, in input-focused tasks the options are given, and they are typically not more than two or three, as in Hsieh’s treatment. The observer can thus tacitly complete the task along with the performer, and indirectly benefit from the feedback to a greater extent: the solution to the problem can be inferred by an elimination process.
On the other hand, the counterevidence provided by Muranoi (2000), where observers in the less explicit condition (IEM) learned significantly as much as the performers they observed, cannot be explained in terms of task mode, since his treatment task was output-focused, as in Mackey (1999) and the present study. One factor that may have contributed to this is the fact that Muranoi’s IEM treatment was actually rather explicit, because in addition to prompts it provided positive evidence through recasts and repetition. When making errors, the participants in Muranoi’s study were prompted to reformulate; if they modified their output correctly, Muranoi repeated the modified output to confirm that the new hypothesis was valid, or if the reformulation was incorrect, he provided a recast, thus supplying positive evidence. In contrast, in Mackey’s study positive evidence was provided indirectly via modified input; and in the present study, participants were prompted to repair, but they did not get a confirmation after their reformulation, nor did they receive positive evidence at any point. Clearly, then, from the perspective of the observer, the experience in Muranoi was more explicit, and did not require an active engagement on the part of the observers to notice their gaps and fill them in, as opposed to the treatments in Mackey (1999) and the present study.

Switching now to the other targeted structure, Spanish Present subjunctive, there are two main factors that may contribute to explain the lack of a gap between performers and observers (both of them learned comparably). The first factor is that, as hypothesized in Chapter 5, Present subjunctive proved to be a harder structure to learn than Prepositional relative clauses —the learning experienced by performers was lower, and as a result, there was less room for differences by type of Agency. For example, for written production the [Practice SP] group experienced immediate gains of 17% (Present
subjunctive) vs. 32% (Prepositional relative clauses). Similarly, for grammaticality judgment the ratio was 16% to 35%. This difference did not occur for oral production, though, where [Practice SP] experienced gains of 14% for both structures, most likely due to the more cognitively demanding nature of this test as a result of time constraints.

The second factor that may explain the lacking role of Agency for Present subjunctive is the fact that, albeit low, observers did actually experience some significant learning (in written and oral production). Specifically, [Exposure SP] obtained immediate gains of 20%, 17%, and 10% in written production, oral production, and grammaticality judgment, in contrast to Prepositional relative clauses, where gains were 6%, 5%, and 3%. Also, as Table 7.2 below illustrates, overall, there was a higher number of observers that derived learning from the experience than for Prepositional relative clauses. Arguably, the higher gains of [Exposure SP] for Present subjunctive cannot be explained by a greater exposure to positive evidence, since performers did worse during the treatment and immediate posttests for Present subjunctive than for Prepositional relative clauses. More likely, then, these gains are explained by activation or generalization of prior knowledge. While participants had not received instruction in the use of Present subjunctive vs. indicative in the particular context of adjectival relative clauses, they had received instruction and practice with Present subjunctive in nominal clauses. Therefore, this prior experience may have helped them to generalize to this new context the rules that they had learned before. Also, they may have stored some items as un-analyzed chunks, retrieving them later in the posttests. Future analyses discriminating between new and old items in the present study may help to elucidate whether or not this was the case.
Table 7.2 Written production gains for both forms in Spot & Prompt dyads

<table>
<thead>
<tr>
<th>Form</th>
<th>P learned, O did not</th>
<th>P learned, O learned</th>
<th>P did not learn, O learned</th>
<th>No-one learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepositional RCs</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Present subjunctive</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

*Note. P = Performer; O = Observer*

Overall, then, the results of the present study suggest that Practice may have an edge over Exposure to practice when the following two conditions are met: (a) the learning condition is more demanding (e.g., when grammatical rules are not explained and positive evidence is not provided, or is provided indirectly via modified input instead of a combination of recasts, prompts, and repetitions, and when the task is output-focused rather than input-focused), and (b) when the grammatical structure is simpler. If the grammatical structure is harder, performers will experience lower gains and these may not be significantly different from their observers’. The question that arises, then, is: What happens, though, when the learning condition is less demanding (e.g., grammatical rules are explained)? The answer to this question is provided in the next section.

**Agency in Explain & Prompt treatments**

When the learning condition included Explain & Prompt Feedback, Agency did not play a role in L2 development. Participants who engaged in controlled production practice with corrective feedback that included grammatical rules and requests for error repair did experience significant learning of both structures on all three dependent measures (except for Present subjunctive in the grammaticality judgment test). However, Agency again
played no role in L2 development because participants who merely observed the completion of grammar practice learned significantly the same as the performers in all cases, for both structures in all dependent measures.

To probe deeper into these findings, the number of observers who experienced learning of 3 or more items in the immediate written production posttests was calculated, for each grammatical structure separately. Again, this test was chosen for illustrative purposes due to its close resemblance to the treatment task. As Table 7.3 below illustrates, out of a total of 13 dyads of performers and observers (P-O), there was only one case in which the observer did not experience learning despite observing a performer who did. In all other cases, when the performer experienced learning the observer did too. Additionally, for Present subjunctive, three observers were able to experience learning even when the participants they observed did not learn themselves. Clearly, this means that for this later structure they were able to use the explicit feedback and (indirect) practice more productively than their peers.

### Table 7.3 Compared written production gains by Form in Explain & Prompt dyads

<table>
<thead>
<tr>
<th>Form</th>
<th>P learned, O did not</th>
<th>P learned, O learned</th>
<th>P did not learn, O learned</th>
<th>No-one learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepositional RCs</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Present subjunctive</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

*Note. P = Performer; O = Observer*

Out of the three empirical studies that have observed the effects of Agency in L2 grammar development (Hsieh, 2007; Mackey, 1999; Muranoi, 2000), only two, Muranoi
(2000) and Hsieh (2007), included a treatment condition that provided the grammatical rule for the targeted form at play. In Murano’s (2000) study, 10 students in the IEF group (interaction enhancement plus formal debriefing) participated in dyadic interaction with the teacher, receiving recasts and prompts in response to errors in their use of English articles, while 20 other students observed their interactions. At the end of each session, all students received a “formal debriefing” where the teacher provided a grammatical explanation on the targeted form. In Hsieh’s (2007) study, 15 participants in the PREFB group [+practice, +explicit feedback] completed a computerized task to sequentially translate a picture cue into its equivalent Spanish *gustar* structure, receiving grammatical explanations after each click. Meanwhile, 7 participants in the EXPEFB group [-practice, +explicit feedback] watched task-completion from a different computer using screen sharing software. In line with this dissertation, Murano (2000) found that performers and observers learned significantly the same. Similarly, Hsieh (2007) found no significant role of type of agency for written recognition and production. However, in contrast to the present study she did find an edge of practice over exposure to practice for oral production.

Some considerations must be made when interpreting the findings by Murano and Hsieh. First, Murano’s analysis was performed on gain scores for all dependent measures combined (grammaticality judgment, written production, and two measures of oral production), so it is hard to assess whether or not there were any differences in terms of type of test. Second, Hsieh had a very low number of participants in her exposure condition (7 and 8 in the explicit and implicit treatments, respectively), although the effect size ($\eta_p^2 = 0.09$) for the obtained significant interaction Time x Agency for oral
production does warrant further data collection. Third, neither Muranoi nor Hsieh performed independent analyses for the role of Agency by type of treatment (more explicit, less explicit, no feedback). Thus, although they did not get a significant interaction Agency x Treatment, or Time x Agency x Treatment, it is hard to assess if, as illustrated in Chapter 6, their holistic analysis may have obscured further differences as a result of the lack of a gap between the two agency levels in the control conditions.

On this note, when looking at learning gains by the explicit conditions in Muranoi (2000) and Hsieh (2007), performers always had higher gains than observers. In the present dissertation, however, the differences were minimal, and in some cases observers experienced higher gains. This may be because attention was more tightly enforced in this study. In Muranoi, observers were told to act as consultants if their fellow performers needed help, but it is not clear how they did so or if they ever did. In Hsieh, observers were not given any special instructions other than to observe task-completion and verbalize whatever it was they were thinking at the moment, like the performers. In the present study, however, observers were told that they would be asked check-up questions on the answers typed in by their performing peers, and they answered them with high rates of success (minimally, 9 out of 10 questions on average). This may explain why the performers consistently got higher gains than the observers in Muranoi’s and Hsieh’s studies, but not here, and why Hsieh reached a significant edge for Practice in oral production.

To summarize, then, the incipient literature on this area of inquiry suggests that in explicit conditions with provision of metalinguistic information, type of Agency (Practice versus Exposure to practice) plays at best a very low role in L2 development, even more
so if attention is tightly controlled for. This is tangentially consistent with claims that explicit instruction may mitigate the effects of type of practice and even learner’s individual differences (see, e.g., Sanz & Morgan-Short, 2005). For example, Erlam (2005) found that deductive instruction (i.e., provision of metalinguistic rules) that gives students opportunities to produce language output and to receive negative evidence (all three conditions shared by the Explain & Prompt groups in this study), may neutralize individual differences in language aptitude. Similarly, in a recent computerized study, Lado (2008: 245) found that when learners with different levels of bilingualism completed practice with feedback to learn a third language “more explicit treatments level[led] the field, eliminating the power of individual differences.”

**Agency in No Feedback treatments**

When the learning condition included No Feedback, Agency did not play a role in L2 development. Participants who completed L2 practice without receiving corrective feedback on their grammatical accuracy did not experience any significant learning of either of the two targeted structures in any of the three dependent measures, either immediately or two weeks later, and neither did the participants who observed them.

These findings are not surprising; however, it may be interesting to discuss them from a psycholinguistic perspective. Arguably, empirical literature on the role of grammar practice in SLA has seldom isolated practice from feedback, very possibly because as posed by Loschky and Bley-Vroman (1993) the presence or absence of feedback is what usually distinguishes tasks for learning and testing purposes. However, as posed by Hsieh (2007), it is necessary to isolate these two variables to investigate their separate
contributions. Drawing on the Interaction Approach (Gass & Mackey, 2006),
participation in interaction, or interactive practice, can be beneficial for SLA because it
may trigger developmentally helpful processes such as comprehending and processing
input or positive evidence, producing output, processing modified input or corrective
feedback in response to errors or communication problems, and producing modified
output. In the present study, however, participants in the [Practice NF] group did not have
access to most of these processes. They did not have the chance to comprehend and/or
process input containing the targeted structures, since every effort was made to exclude
exposure to positive evidence via the treatment and testing materials; they did not receive
feedback, and they were not pushed to modify their output. Consequently, if this
experimental group had experienced any potential learning, it could only have resulted
from producing output.

According to Swain’s Output Hypothesis, output practice can not only increase
fluency and automatization of already acquired-knowledge but also bring about mental
processes that both directly and indirectly improve accuracy and affect acquisition: the
noticing function, the metalinguistic function, and the hypothesis-testing function. This
latter function refers to the fact that, when learners produce an ill-formed utterance, they
may trigger feedback that can cause them to “reprocess” their hypotheses, leading them
to produce modified output. Since in this study the [Practice NF] group did not receive
any negative evidence or prompts for error repair, the only potential sources for learning,
then, were the noticing and metalinguistic functions.

The noticing function refers to the fact that, in the process of producing speech,
learners are confronted with what they want to say and what they can actually say, and as
a result they may notice what they do not know at all or what they know only partially. In turn, the *metalinguistic function* refers to the fact that certain tasks may elicit both production for communication purposes and conscious reflection about language form. The occurrence of these two functions has been documented in the literature, with studies showing that output practice can trigger noticing of problems in vocabulary (e.g., Bialystok, 1990; Faerch & Kasper, 1983; Kellerman, 1991; Tarone, 1977) and grammar (e.g., Swain & Lapkin, 1995), and promote conscious reflection on language form (Donato, 1994; Kowal & Swain, 1997; LaPierre, 1994; Swain, 1998; Swain & Lapkin, 1998, 2002). However, it is still unclear whether these two functions of output practice can constitute a *direct* path to L2 grammar development.

There are two main positions in this debate. On the one hand, researchers such as Krashen (e.g., 1982), Izumi (2002), VanPatten (2004a), and Farley (2001a) claim that output practice can only help acquisition indirectly, by helping learners to detect their gaps and holes and sensitizing them to pay focused attention to subsequent input, or by serving as self-input of already acquired knowledge or as input to other learners. In contrast, authors like Swain and Lapkin (1995), Toth (2006), and McDonough (2005) illustrate that learners do not necessarily have to resort to positive evidence in the ambient input to fill their gaps and holes. Toth and McDonough demonstrated that students can successfully modify their ill-formed production of grammar without processing positive evidence in subsequent input, but rather, by processing tutor’s feedback with metalinguistic explanation or prompts. These findings, however, do not apply to the [Practice NF] group because, again, this group was deprived from negative evidence or prompts for repair. Beyond the hypothesis testing function, Swain and
Lapkin (1995) illustrated that output practice can trigger metalinguistic processing in and by itself. However, ultimately the learner always needs to access positive evidence to corroborate his or her hypotheses, which can be done prospectively, in subsequent input, or retrospectively, in stored input.

If the noticing and metalinguistic functions of output practice were the only potential sources of learning for the [Practice NF] group, the question that arises then is whether or not they took place. In questions 2 and 3 of the exit questionnaire (see Appendix 5.6), participants were asked if they had identified the purpose of the study and the targeted vocabulary or grammatical forms. As shown in Table 7.4 below, out of 13 retrieved testimonials, three participants in the [Practice NF] group (P1, P2, and P3) were able to at least partially identify one or two of the targeted structures.

**Table 7.4 Answers to Q2 and Q3 in exit questionnaire by participants in [Practice NF]**

<table>
<thead>
<tr>
<th>Q2. Purpose of the study?</th>
<th>Q3. Targeted structures/items?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Preposition placement with dependent clauses...?</td>
<td>See Question 2; vocabulary--practical words, particularly useful when traveling abroad</td>
</tr>
<tr>
<td>P2 I don't think so, but possibly having to do with learning about prepositions at the end of sentences and learning that concept by verbal and written responses.</td>
<td>Prepositions at the end of sentences. Use of que. Subjunctive. Getting an apartment</td>
</tr>
<tr>
<td>P3 i think the purpose has to do with repetition and learning from prior tries as well as learning to properly conjugate verbs with pronouns etc.</td>
<td>The author was trying to teach present tense as well as pronouns and connecting words placement.</td>
</tr>
</tbody>
</table>

Close examination of the treatment transcripts showed that only 3 participants out of 15 in the [Practice NF] group had typed in any of the two targeted structures 3 times or more while interacting with the avatars, either correctly or incorrectly. One of them had used the Present subjunctive throughout, clearly overgeneralizing it, quite likely as a
result of the preliminary morphology review session, which focused extensively on this
tense (see Appendix 5.5). Interestingly, the other two participants were precisely 2 of the
3 ones that had reported noticing the targeted structures in the exit questionnaire.
Participant P3 used the Present subjunctive four times, and both she and P1 correctly
used pied-piping for Prepositional relative clauses during the treatment, in 6 and 9 out of
10 obligatory contexts, respectively. Therefore, beyond noticing her interlanguage gap, it
seems like P3 was definitely testing hypotheses, even in the absence of feedback. On the
other hand, P1 was clearly already one step beyond, given his minimal error rate. On
questions 4-6 of the exit questionnaire (Appendix 5.6) both P1 and P3 answered that they
had not consulted any external source of information throughout the experiment.
Question 4 read: “Did you review/go over any of these structures between the first and
last day of this experiment? Please be honest, no matter what you did. This is important
to interpret whether any potential learning was exclusively due to the software used in
this experiment or was aided by other factors.” Consequently, it seems like these two
participants, which represented 13% of the total pool of students (N = 15) were able to
derive learning from the pretest and treatment materials, perhaps by activation of latent
prior knowledge that was not reflected on the pretests. In fact, these two participants were
the group’s highest achievers on the immediate written production posttest, with scores of
10/10 (P1) and 4/10 (P3). Clearly, however, they were outliers in this group, where mean
gain scores from pretests to posttests were negligible and not significant (see Appendices
6.1-6.6). Overall, then, controlled written production practice without positive evidence,
negative evidence, and prompts for error repair only triggered the facilitative functions
posited by Swain in a minimal number of participants.
As a result of the above, participants in the [Exposure NF] group had very few 
opportunities for learning, as corroborated by their posttest scores (see Appendices 6.1-
6.6). However, since P1 and P3 did provide many instances of positive evidence for 
Prepositional relative clauses, it is interesting to see if their observers, O1 and O3 
benefited from it. This was not the case at all for the P1-O1 dyad. As Table 7.5 below 
shows, P1 got 9 out of 10 targeted items right in the treatment (left column) and scored a 
full 10/10 in the immediate written posttest (central column), while his observer, O1 did 
not get a single item right in the immediate written posttest (right column), despite 
answering correctly all of the check-up questions that controlled for attention. 
Conversely, in the P3-O3 dyad, exposure to positive evidence did translate into learning. 
During the treatment, P3 correctly used preposition pied-piping in 6/10 items, and later 
scored 4/10 in her written production posttest. In turn, O3 scored 3/10 in his posttest, 
having scored 0/10 on the pretest (see Table 7.6). When observing their answers to 
questions Q2 and Q3 in the exit questionnaire, both O1 and O3 reported noticing relative 
clauses, but they did not mention prepositions, although lack of verbalization does not 
necessarily entail lack of noticing (see Table 7.7). As for Present subjunctive, neither of 
the two observers that were exposed to positive evidence in the treatment (with 11 and 4 
instances for the 15 items of the communicative situation) derived any learning from it. 

Consequently, given the different learning gains experienced by O1 and O3 for 
Prepositional relative clauses, it seems clear that the pedagogical effects of exposure to 
positive evidence embedded in practice are mediated by variables other than attention to 
task-completion, such as focalized attention to meaning versus form, different levels of
attention and awareness, individual differences in cognitive capacity, or different levels of latent prior knowledge.

**Table 7.5** Accuracy for Prepositional relative clauses in treatment (P1) and immediate written posttest (P1, O1)

<table>
<thead>
<tr>
<th><strong>P1 in treatment</strong></th>
<th><strong>P1 in written posttest</strong></th>
<th><strong>O1 in written posttest</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>X que estoy aqui</td>
<td>contra que todos los paises deben luchar.</td>
<td>X que todos los paises deben luchar contra</td>
</tr>
<tr>
<td>con que hablo</td>
<td>por que estoy aqui es...</td>
<td>X que esta aqui por es</td>
</tr>
<tr>
<td>con que vivo</td>
<td>con que vivo es un estudiante tambien.</td>
<td>X que vivo con es un estudiante tambien</td>
</tr>
<tr>
<td>para que trabaja.</td>
<td>para que trabaja.</td>
<td>X que ella trabaja por</td>
</tr>
<tr>
<td>en que guarda mi dinero.</td>
<td>de que estaban llamando.</td>
<td>X que ellos esta llamando desde</td>
</tr>
<tr>
<td>por que se ve.</td>
<td>sobre que discutan.</td>
<td>X que ellos descuta por</td>
</tr>
<tr>
<td>de que habla</td>
<td>de que esta hablando es imposable.</td>
<td>X que tu es habla de es imposable</td>
</tr>
<tr>
<td>en que pongo.</td>
<td>bajo que viven.</td>
<td>X que ellos vive bajo</td>
</tr>
<tr>
<td>desde que estaba llamando.</td>
<td>a que este telefono pertenece?</td>
<td>X que es el telefono pertenece a</td>
</tr>
<tr>
<td>con que tengo un contrato telefonico</td>
<td>En que vivimos.</td>
<td>X que nosotros vivimos en</td>
</tr>
</tbody>
</table>

*Note.* P = performer; O = observer; X = incorrect answer.
Table 7.6 Accuracy for Prepositional relative clauses in treatment (P3) and immediate written posttest (P3, O3)

<table>
<thead>
<tr>
<th></th>
<th>P3 in treatment</th>
<th>P3 in written posttest</th>
<th>O3 in written posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>X que estoy aqui por</td>
<td>X</td>
<td>que todas las paises lucha contra.</td>
<td>X</td>
</tr>
<tr>
<td>X que hablo con</td>
<td>X</td>
<td>que estoy aqui es...</td>
<td>X</td>
</tr>
<tr>
<td>X que vivo con</td>
<td></td>
<td>con que vivo es un estudiante tambien.</td>
<td>X</td>
</tr>
<tr>
<td>para que ella trabaja</td>
<td></td>
<td>que por trabaja</td>
<td></td>
</tr>
<tr>
<td>en que guardo mi dinero</td>
<td></td>
<td>de que ellos llaman.</td>
<td></td>
</tr>
<tr>
<td>X que viste por</td>
<td>sobre que ellos discutieron</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de que hables</td>
<td>X</td>
<td>que sobre hablas es improbable.</td>
<td>X</td>
</tr>
<tr>
<td>en que lo pongo</td>
<td></td>
<td>que ellos viven bajo.</td>
<td>X</td>
</tr>
<tr>
<td>desde que esta llamando</td>
<td></td>
<td>que eso telefono pertenece a</td>
<td></td>
</tr>
<tr>
<td>con que tengo contrato telefonico</td>
<td></td>
<td>en que vivamos.</td>
<td></td>
</tr>
</tbody>
</table>

Note. P = practicing participant; O = observer; X = incorrect answer.

Table 7.7 Answers to Q2 and Q3 in exit questionnaire by O1 and O3

<table>
<thead>
<tr>
<th></th>
<th>Q2. Purpose of the study?</th>
<th>Q3. Targeted structures/items?</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>its purpose is to see how we improve week by week, and if repetition helps</td>
<td>he is trying to reinforce using que (that) and using sentence structures</td>
</tr>
<tr>
<td>O3</td>
<td>To determine how the repetition of tasks influences one's learning of the Spanish Language, and how a student's interactions with native speakers influence their learning.</td>
<td>Combining sentences using the word “que”; Present, Preterite, &amp; Subjunctive Verb Forms</td>
</tr>
</tbody>
</table>
To sum up, then, output practice without positive evidence, negative evidence, and prompting for error repair led to no significant learning of either grammatical structure for the [Practice NF] group. However, anecdotal evidence of learning was found for 2 of the 15 students for one of the two structures, Spanish prepositional relative clauses. These students reported noticing of Spanish prepositional relative clauses and demonstrated learning. Since they denied having consulted this structure outside the laboratory setting for the duration of the study, learning may have resulted from treatment and/or test effects and possible activation of latent prior knowledge not evidenced on the pretests. In turn, exposure to practice including positive evidence only facilitated learning for one of two participants for Prepositional relative clauses and zero of two participants for Present subjunctive. This is interesting because the observers were exposed to practice that was very controlled and clearly isolated the targeted form (constrained selected response), involved written rather than oral comprehension, which is supposed to enhance noticing, and demonstrated full attention to the practice episodes. In line with White (1989), then, and in contrast to Krashen (e.g., 1984), this anecdotal evidence illustrates that exposure to positive evidence alone may not be sufficient to learn structures that involve retreating from overgeneralizations caused by an L1 superset grammar.

2. Research question #2. Type of Feedback and L2 development

The second research question addressed whether type of oral Feedback provided in response to learners’ productions of a grammatical structure has differential effects on learners’ development of that form, and if so, which of two particular subcomponents of feedback, explicitness of negative evidence and prompting for error repair, may
contribute to any observed differences. Results showed that for both grammatical forms and all five Feedback groups combined, significant L2 development was observed for all three dependent measures. Furthermore, the [Practice EC] group consistently outperformed the control group [Practice NF] in written and oral production, both immediately and two weeks later, and a similar trend was observed for grammaticality judgment. The [Practice EC] group also had an edge over the two less explicit Feedback groups ([Practice SC] and [Practice SP]) on the immediate written posttest, and over the [Practice SC] group on the immediate oral posttest. These differences, however, washed out after two weeks. Independent analyses by type of Form revealed that the edge of the [Practice EC] group appeared for Prepositional relative clauses more clearly than for the Present subjunctive. The results for each form are separately discussed below.

**Prepositional relative clauses**

The results of this study for Prepositional relative clauses revealed that the [Practice NF] group did not experience any learning, while all groups with Feedback did. Additionally, the [Practice EC] group outperformed the [Practice NF] group on all three dependent measures at all test times, except for the delayed grammaticality judgment posttest, where significance was approached. Only one additional significant difference was found, with [Practice EP] outperforming [Practice NF] on the immediate written production posttest.

The obtained main effect for Time for each of the four feedback groups in this design is in line with recent meta-analyses that demonstrate a facilitative role of feedback in SLA (e.g., Russell & Spada, 2006; Li, 2010; Lyster & Saito, 2010) and provides counterevidence to claims that L2 development depends on exposure to positive evidence
alone and negative evidence is not necessary or even potentially harmful (e.g., Krashen, 1981; Schwartz, 1993; Truscott, 2007).

In terms of type of Feedback, the present study found no significant differences between groups. This is likely due to an insufficiently large sample size for the number of feedback levels under comparison (five). Consistently for all dependent measures, however, the plotted graphs suggested three levels of attainment, with the [Practice NF] group experiencing no learning, the [Practice EC] group being the highest achiever, and the other three Feedback groups clustering in the middle. In the discussion that follows, the role of type of Feedback is separately analyzed for each of the two feedback components investigated here, explicitness of negative evidence and prompting for error repair.

The fact that [Practice EC] was the only group that significantly outperformed the control group, [Practice NF], on all dependent measures and test times (except for grammaticality judgment at Time 3), supports Li (2010: 35), who in a meta-analysis of 33 published and unpublished studies for the 1991-2007 period found that explicit feedback worked better than implicit feedback on immediate and short-delayed posttests (7-29 days). In the realm of CALI specifically, the results for Prepositional relative clauses here align with research showing an edge of more explicit feedback conditions (Bowles, 2005; Lado, 2008; Nagata, 1993; Nagata & Swisher, 1995; Rosa & Leow, 2004b). On the other hand, it does not provide direct counterevidence for other studies that did not find a significant difference between more and less explicit feedback types (Camblor, 2006; Hsieh, 2007; Moreno, 2007; Sanz, 2004; Sanz & Morgan-Short, 2004), although as observed earlier future replication with a larger sample size may do so. As a matter of
fact, an experimental analysis performed on a duplication of the current data did yield the aforementioned three levels of achievement for all dependent measures, with the [Practice EC] group outperforming the Spot Feedback groups, which in turn outperformed the [Practice NF] group. Additionally, the present study also shows the importance of including a control, No Feedback group in the experimental design, since with the current number of participants the Spot Feedback groups never outperformed the control group, despite their significant gains.

In terms of prompting for error repair, the present study did not find an advantage of prompting over no prompting, as evidenced by the fact that there were no significant differences between the Prompt and Continue groups with the same type of negative evidence. In the two groups with Spot Feedback, the Prompt group experienced slightly higher gains across all three dependent measures. Interestingly, however, in the Explain Feedback groups it was the Continue group that experienced the highest gains. Furthermore, this was the only group that consistently outperformed the control group in all dependent measures, since [Practice EP] only outperformed [Practice NF] in the immediate written production posttest. Now, even if the difference between [Practice EC] and [Practice EP] never reached significance, the question that arises, given the inferior performance of the latter, is whether prompts can be detrimental.

One possible interpretation of these results is that intensive prompts can disrupt the communication flow and allocate too much weight on formal accuracy at the expense of meaning, thus hindering acquisition and in line with Krashen’s reservations about form-focused instruction. This is precisely why many authors have advocated for recasts over other feedback types, because they argue that recasts are time-saving, help maintain a
focus on meaning to a greater extent than other techniques, and are less threatening to student confidence (e.g., Han & Kim, 2008; Loewen & Philp, 2006; Nicholas et al., 2001).

Another possible interpretation of these results is that, in the absence of post-reformulation corrective feedback (either positive or negative), prompts may actually contribute to confuse the learner. In the present study, participants were prompted by the avatars if their initial production was ill-formed. After they modified their output, either successfully or not, the avatar moved on in the dialog and proceeded to ask the next question without providing an evaluation of the participant’s reformulation. This decision was made for experimental reasons, in order to control for the amount of negative evidence provided, similarly to McDonough (2005) (personal communication, May 2008) and differently from e.g., Muranoi (2000: 635). In this latter study, when the learners produced modified output, Muranoi “evaluate[d] the output by providing feedback (i.e., teacher repetitions), so that the learner [could] confirm that his or her hypothesis [was] valid.” It may be argued, then, that if no such post-reformulation evaluation is supplied, students may exit the communicative mini-episode deceitfully thinking that they got it right even if they did not. However, in the context of a battery of task-essential activities with intensive feedback (such as the one in this study) this is not likely, because participants are bound to encounter negative evidence again and again when they try out their new hypothesis in subsequent items, unless the specific reformulation occurs in the very last item in the interaction.

A third possible explanation to account for the non-significant advantage of [Practice EC] over [Practice EP] is that the potentially beneficial effects of prompts, if any, are not
intrinsic to prompts but rather, subject to whether or not learners are able to modify their ill-formed productions successfully, which is ultimately determined by the learners’ ability to use the negative evidence to which prompts are juxtaposed. Specifically, if the negative evidence is minimally explicit, the learner’s success will be determined by his or her opportunities to access positive evidence either retrospectively (in memory) or prospectively (in subsequent input), or to generate new hypotheses. In turn, if the negative evidence is more explicit (the grammatical rule is laid out), the learner’s ability to reformulate ill-formed output successfully will depend on his or her ability to decode the metalinguistic information provided. To better elucidate if this latter possibility could contribute to explain the performance gap between [Practice EC] and [Practice EP], the treatment transcripts of both groups were examined and compared. Participants were divided into “successful modifiers” and “unsuccessful modifiers,” depending on whether or not they were able to switch, at least once, from ungrammatical preposition stranding (PS) to grammatical preposition pied piping (PPP), and then, their performance during the treatment was linked to their subsequent performance on the written immediate posttests (these tests were chosen as illustrative of learning because they most closely resembled the treatment task). In this coding scheme, successful output modifications were considered any successful change in production after negative feedback. Thus, output modifications did not only occur in the [Practice EP] group but also in the group without prompts, [Practice EC], whenever a participant started out by using PS and then switched to PPP in subsequent items after receiving negative feedback.

Table 7.8 below reproduces the treatment transcript of an “unsuccessful modifier” in the [Practice EP] group. In this case, the student started the treatment by using PS in
items 22 and 26, repeating her ill-formed productions after receiving feedback. In item 28 she tried PS again but then, upon processing the feedback message, she tried a different strategy. The feedback message was: Ajá. Oye, la preposición ‘para’ no está en el lugar apropiado. Debes ponerla justo después del antecedente ‘compañía’ y antes del pronombre relativo ‘que’ [“Uhuh. Listen, the preposition ‘para’ is not in the right place. You should put it right after the antecedent ‘compañía’ and after the relative pronoun ‘que’”]. Following the feedback message, the student moved the preposition after —instead of before— the relative pronoun que (insufficient PPP). In subsequent items, she persisted on her use of insufficient PPP, clearly demonstrating her inability to fully decode the feedback messages that followed her productions. Finally, in the last item, she gave up trying, returning to her original PS.
Table 7.8 Treatment transcript of an “unsuccessful modifier” in [Explain EP]

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Participant’s output</th>
<th>Accuracy on targeted items</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>XXXX, 18, Indian</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>que yo adora</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>que estoy <strong>por</strong></td>
<td>X</td>
</tr>
<tr>
<td>22</td>
<td>que estoy <strong>por</strong></td>
<td>← PS</td>
</tr>
<tr>
<td>23</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>que proporciona ayudar a estudiantes Amer…</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>que you hable</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>que vivo <strong>con</strong></td>
<td>X</td>
</tr>
<tr>
<td>26</td>
<td>que vivo <strong>con</strong></td>
<td>← PS</td>
</tr>
<tr>
<td>27</td>
<td>que viene a nuestra casa</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>que ella trabaja <strong>para</strong></td>
<td>X</td>
</tr>
<tr>
<td>28</td>
<td>que <strong>para</strong> ella trabaja</td>
<td>← insufficient PPP</td>
</tr>
<tr>
<td>29</td>
<td>que esta al lado de</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>que guardo <strong>en</strong> mi dinero</td>
<td>X</td>
</tr>
<tr>
<td>30</td>
<td>que <strong>en</strong> guardo mi dinero</td>
<td>← insufficient PPP (x2)</td>
</tr>
<tr>
<td>31</td>
<td>que <strong>por</strong> tu puedes ver</td>
<td>X</td>
</tr>
<tr>
<td>31</td>
<td>que <strong>por</strong> puedes ver</td>
<td>← insufficient PPP (x3)</td>
</tr>
<tr>
<td>32</td>
<td>que robo mi dinero</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>que <strong>de</strong> estas hablando</td>
<td>X</td>
</tr>
<tr>
<td>33</td>
<td>que <strong>de</strong> estas hablando</td>
<td>← insufficient PPP (x4)</td>
</tr>
<tr>
<td>34</td>
<td>que <strong>en</strong> you puse.</td>
<td>X</td>
</tr>
<tr>
<td>34</td>
<td>que <strong>en</strong> puse</td>
<td>← insufficient PPP (x5)</td>
</tr>
<tr>
<td>35</td>
<td>que <strong>desde</strong> estan llamando</td>
<td>X</td>
</tr>
<tr>
<td>35</td>
<td>que <strong>desde</strong> estan llamando</td>
<td>← insufficient PPP (x6)</td>
</tr>
<tr>
<td>36</td>
<td>que <strong>con</strong> tengo mi contrato telefonico</td>
<td>X</td>
</tr>
<tr>
<td>36</td>
<td>que tengo mi contrato telefonico <strong>con</strong></td>
<td>← PS</td>
</tr>
</tbody>
</table>

*Note.* X = ill-formed output on first trial; ← = modified output; (xNumber) = Number of repetitions of a type of modified output; PS = preposition stranding; PP = pied-piping.
In contrast, Table 7.9 below reproduces the treatment transcript of a “successful modifier” in [Practice EP]. In this case, this participant went again through an initial stage of PS and, later on, insufficient PPP on item 26, like the former participant. However, on item 28 this participant was able to correctly reformulate his ill-formed production by using full PPP, and after 3 occasional errors on his first trial, he always reformulated successfully, answering the last items correctly on his first attempts.
Table 7.9 Treatment transcript of a successful modifier in [Explain EP]

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Participant’s output</th>
<th>Accuracy on targeted items</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>me llamo XXXXX, soy 19 años y soy de ingle...</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>que yo amo</td>
<td>X</td>
</tr>
<tr>
<td>22</td>
<td>que estoy aquí <strong>para</strong></td>
<td>← PS</td>
</tr>
<tr>
<td>22</td>
<td>que estoy aquí <strong>para</strong></td>
<td>← PS</td>
</tr>
<tr>
<td>23</td>
<td>dos cientos</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>que proporciona ayuda a estudiantes americanos</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>que hablo <strong>con</strong></td>
<td>X</td>
</tr>
<tr>
<td>25</td>
<td>que hablo <strong>con</strong></td>
<td>← PS</td>
</tr>
<tr>
<td>26</td>
<td>que vivo <strong>con</strong></td>
<td>X</td>
</tr>
<tr>
<td>26</td>
<td>que <strong>con</strong> vivo</td>
<td>← insufficient PPP</td>
</tr>
<tr>
<td>27</td>
<td>que viene a nuestra casa</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>que trabaja <strong>para</strong></td>
<td>X</td>
</tr>
<tr>
<td>28</td>
<td><strong>para</strong> que trabaja</td>
<td>← PPP</td>
</tr>
<tr>
<td>29</td>
<td>que estaba al lado de</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>en que guarda mi dinero</td>
<td>⊗</td>
</tr>
<tr>
<td>31</td>
<td>que puede ver <strong>por</strong></td>
<td>X</td>
</tr>
<tr>
<td>31</td>
<td><strong>por</strong> que puede ver</td>
<td>← PPP (x2)</td>
</tr>
<tr>
<td>32</td>
<td>que roba mi dinero</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>que habla <strong>de</strong></td>
<td>X</td>
</tr>
<tr>
<td>33</td>
<td><strong>de</strong> que habla</td>
<td>← PPP (x3)</td>
</tr>
<tr>
<td>34</td>
<td>en que poní</td>
<td>⊗</td>
</tr>
<tr>
<td>35</td>
<td>desde que estaba llamando</td>
<td>⊗</td>
</tr>
<tr>
<td>36</td>
<td><strong>con</strong> que tengo un contrato telefonico</td>
<td>⊗</td>
</tr>
</tbody>
</table>

*Note. X = ill-formed output on first trial; ⊗ = correct output on first trial; ← = modified output; (xNumber) = Number of repetitions of a type of modified output; PS = Preposition Stranding; PPP = Preposition Pied-Piping.*

As Table 7.10 below shows, both [Practice EC] and [Practice EP] comprised a pool of 17 participants each. In the [Practice EC] group all 17 participants were able to correctly
reformulate their output at least once, whereas in the [Practice EP] group there were 7 participants who never managed to provide a successful reformulation. In the [Practice EC] group no-one scored 0 on the immediate posttest; thirteen participants scored 3 or above, and the remaining 4 scored below. In contrast, in the [Practice EP] group all 7 participants who could not manage to reformulate their output successfully scored 0 on the posttest, and among the remaining 10 participants, 9 scored above 3 points and 1 scored below.

Table 7.10 Number of participants by successful modifications during the treatment and learning attainment in the immediate written posttest

<table>
<thead>
<tr>
<th></th>
<th>0 in treatment, 0 in posttest</th>
<th>+1 in treatment, &gt;3 in posttest</th>
<th>+1 in treatment, &lt;3 in posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Practice EC]</td>
<td>0</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>N= 17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Practice EP]</td>
<td>7</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>N= 17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clearly, then, L2 development was tightly linked to whether or not the participants were able to provide a successful reformulation during the treatment, which arguably was due to their ability to decode the metalinguistic information in the feedback. Interestingly, the results of the exit questionnaire revealed that even if [Practice EC] and [Practice EP] received the exact same metalinguistic rules in their feedback messages, the former group reported significantly higher levels of satisfaction than the latter in terms of how useful they thought the feedback was (Question 19) and how satisfied they were with its level of detail (Question 20). The next challenge, then, is to identify what determines the learners’
ability to decode explicit negative evidence (e.g., language aptitude, listening comprehension, or prior knowledge of the structure not evidenced on the treatment). To this regard, triangulation with the participants’ MLAT scores gathered during the data collection process may cast additional light into the picture. For now, however, to further refine these findings the unsuccessful modifiers in the [Practice EP] group were excluded from the pool. In the resulting pool, the successful modifiers got an average of 7.00 items right out of 10 in the immediate written posttest. Similarly, the participants in the [Practice EC] group scored an average of 7.35 out of 10. This suggests that in treatments with intensive explicit feedback including metalinguistic information, L2 development is solely determined by the students’ ability to decode the information in the negative evidence, and that prompts do not contribute an additional benefit.

It is hard to compare these findings with previous research, since this is possibly the first study to address the effects of prompts in an explicit treatment with metalinguistic information in isolation from positive evidence. Looking at the big picture, these findings may be interpreted as counterevidence to Lyster and Saito (2010), who in a meta-analysis of 15 classroom interaction studies found that prompts and explicit correction did not yield significantly different effects (p. 282-3). As discussed earlier, in the present study explicit negative evidence without prompts [Practice EC] yielded better results than prompts with minimally explicit negative evidence [Practice SP], since only the former group outperformed the control group. There is, however, a fundamental difference between the present study and the studies meta-analyzed by Lyster and Saito (2010). In the present study, access to positive evidence was minimized, both retrospectively (the structures were novel to learners) and prospectively (the interlocutor did not supply
positive evidence). In contrast, in the studies meta-analyzed by Lyster and Saito, participants had at least partial prior knowledge of the structures, and in some cases they had been taught the rules beforehand, so prompts in that case may have helped them to strengthen their previous knowledge representations (Nobuyoshi & Ellis, 1993) and to promote automatic retrieval of linguistic forms (de Bot, 1996).

As for the effects of prompts in less explicit treatments (no provision of rules), there is only one previous study, McDonough (2005), that has experimentally isolated the effects of prompts and negative evidence without positive evidence. Using logistic regression analysis, McDonough found that successfully modified output, in contrast to negative evidence by itself, was the only significant predictor of development for English question formation. Furthermore, she found that increasing the production of modified output increased the odds of development by a multiple of 10. She concluded (p. 91): “although I expected that both negative feedback and modified output would be significant predictors of ESL question development, the logistic regression model revealed that only modified output produced in response to negative feedback was significant.” In the present study, prompts did not have an impact on L2 development, since both [Practice SC] and [Practice SP] performed statistically similarly. However, as demonstrated earlier, there seemed to be no correlation between the provision of prompts and successful output modifications, since some participants in the [Practice SC] group did equally use the negative evidence received for one item to successfully modify their original production in the following items. In McDonough’s study, however, this was less likely because the negative evidence provided to her no-prompts group did not signal the corrective intent as evidently as in the present study. Specifically, in this study students
were told that they had made an error, as well as its locus and nature. Conversely, in McDonough the tutor repeated the ill-formed segment and moved on in the dialog, thus not only precluding opportunities for modified output but also perhaps obscuring the corrective intent. All in all, however, this study agrees with McDonough in that successfully modified output is the only predictor of learning, but this ultimately only reflects the ability of the learner to decode the negative evidence and access positive evidence either prospectively or retrospectively.

**Present subjunctive**

The results of this study for Present subjunctive revealed that the [Practice NF] and [Practice SC] groups experienced no significant learning, while all others did. However, the only observed difference was between the [Practice EC] and [Practice NF] groups, with the former outperforming the latter in oral production and approaching significance in written production.

Overall, these results confirm those obtained for Prepositional relative clauses, equally suggesting that, in the absence of positive evidence, more explicit negative evidence may have a beneficial role in L2 development, whereas prompts do not. However, the fact that a significant interaction Time x Feedback (with contrasts between [Practice EC] and [Practice NF]) was reached for only one of the three dependent measures and approached for another (as opposed to all measures across the board), suggests that the advantages or more explicit instruction may be moderated by the targeted form at play. Section 4 below elaborates on this issue, à propos of the results for the interaction Time x Form x Feedback.
3. Research question #3. Type of grammatical form and L2 development

The third research question addressed whether type of linguistic Form (i.e., more vs. less complex) has differential effects on L2 development as a result of a pedagogical intervention. Results showed that when all 8 experimental groups were taken together, students experienced significantly higher learning of Prepositional relative clauses than Present subjunctive on written production and grammaticality judgment measures, whereas for oral production the observed learning of the two structures was statistically similar.

Subsequent independent analyses for each of the 8 experimental groups revealed that the obtained interaction Time x Form originated in the Practice groups with Explain Feedback, both [Practice EC] and [Practice EP], on both written production and grammaticality judgment. A similar trend, however, was observed for the other Practice groups with Spot Feedback, where students also experienced higher learning gains for Prepositional relative clauses on written production and grammaticality judgment. In turn, the Exposure groups did not follow a clear pattern, and the No Feedback groups did not experience any significant learning of either structure.

The significantly better results obtained in written production and grammaticality judgment for Prepositional relative clauses in both explicit treatments, [Practice EC] and [Practice EP], corroborate that, as posed in Chapter 5, this structure was less challenging to students than Present subjunctive. Qualitative examination of the treatment and posttest transcripts revealed some of the problems that participants encountered in their attempts to process and implement the metalinguistic information in the feedback to
produce both structures (for illustrative purposes, see Tables 7.11 and 7.12 below, including a sample of a treatment item and the feedback delivered to the Explain groups for each targeted form). In the case of Prepositional relative clauses (Table 7.11), some participants demonstrated trouble understanding the exact location of the prepositions. Some of them consistently moved the preposition after (as opposed to before) the relative pronoun *que*, while others managed to understand the rule either immediately or after a few attempts. Some students, however, overgeneralized it to non-prepositional clauses, producing sentences such as *El grupo CERCA QUE esté un parada de autobús* ("The group NEXT THAT is a bus stop") instead of *El grupo QUE está CERCA DE una parada de autobús* ("The group THAT is NEXT TO a bus stop"). However, these students were a minority (only 5 out of 127 students overgeneralized this rule for one or more items in the immediate written production posttest).

Table 7.11 Sample item and example of Explain feedback for Prepositional RCs

<table>
<thead>
<tr>
<th>ITEM:</th>
<th><em>Sí, el nombre de la persona (THAT I SPOKE WITH) ________ es Pepa.</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>FEEDBACK:</td>
<td><em>Pepa, hablaré con ella. Oye, la posición de la preposición CON no es correcta. En inglés, dices la preposición al final. Por ejemplo: THAT bla, bla, bla... WITH, es Pepa. Pero en español, no. En español ponemos la preposición CON antes del pronombre relativo QUE.</em></td>
</tr>
</tbody>
</table>

In contrast, the rule for Present subjunctive was more complex to implement, since it invoked making a form-meaning connection, that is, participants had to decide whether or not the adjectival relative clause was describing a physical versus an imaginary world. As
a result, there were far more cases of overgeneralization of this structure (66/127 students overgeneralized the Present subjunctive in one or more items in the immediate written production posttest, as opposed to 5/127 for Prepositional relative clauses).

Table 7.12 Sample item and example of Explain feedback for Present subjunctive

<table>
<thead>
<tr>
<th>ITEM:</th>
<th>Quiero un apartamento (THAT IS) ______________ en (la playa / el centro / el campus).</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[I would like an apartment (THAT IS) ______________ (on the beach / downtown / on campus)].</td>
</tr>
<tr>
<td>FEEDBACK:</td>
<td>¡Buena elección! Oye, tienes un problema con el verbo ESTAR. Estamos hablando de un apartamento hipotético, en tu imaginación. Hay muuuchos apartamentos en el centro, en la playa y en el campus. Por eso, aquí tienes que usar el subjuntivo.</td>
</tr>
<tr>
<td></td>
<td>[“Good choice! Hey, you have a problem with the verb TO BE. We are talking about a hypothetical apartment here, one in your imagination. There are many apartments downtown, on the beach, and on campus. Consequently, you should use the subjunctive here.”]</td>
</tr>
</tbody>
</table>

In turn, the lack of a significant interaction Time x Form for oral production (i.e., participants demonstrated statistically similar learning for both forms in this test) was likely due to the nature of the task, that was more cognitively taxing for Prepositional relative clauses than for Present subjunctive. In the written production tests students had to translate a sentence from English into Spanish, under no time constraints. However, in the oral production tests participants had 7 seconds to join two sentences into a new one, using the relative pronoun que and all capitalized words. Table 7.13 below provides sample items for Present subjunctive and Prepositional relative clauses. As table 7.13 shows, to complete the item targeting Present subjunctive participants only had to decide whether to use indicative or subjunctive and conjugate the infinitive verb in (2), but the rest of the sentence stayed the same (que VIVA mucho fuera de casa [“who SPENDS-
SUBJ a lot of time outside the apartment”). In contrast, to complete the sentence targeting Prepositional relative clauses students had to perform two tasks, i.e., choosing the verb mood and rearranging the word order in (1). For that reason, under time constraints students were unable to complete these sentences more often than those merely targeting Present subjunctive.

**Table 7.13** Sample items by grammatical form in oral production test

<table>
<thead>
<tr>
<th>Present subjunctive</th>
<th>Prepositional RCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) <em>Busco un compañero de apartamento</em></td>
<td>(1) <em>Ayer FUI A HABLAR CON una consejera universitaria.</em></td>
</tr>
<tr>
<td>[“I’m looking for a roommate”]</td>
<td>[“Yesterday I went to talk to a university counselor”]</td>
</tr>
<tr>
<td>(2) <em>VIVIR mucho fuera de casa</em></td>
<td>(2) <em>Pepa</em></td>
</tr>
<tr>
<td>[“S/he is often out of the apartment”]</td>
<td>[“Her name is Pepa”]</td>
</tr>
<tr>
<td>(3) <em>Felipe busca un compañero de apartamento…</em></td>
<td>(3) <em>Pepa es la consejera universitaria…</em></td>
</tr>
<tr>
<td>[“Felipe is looking for a roommate…”]</td>
<td>[“Pepa is the university counselor…”]</td>
</tr>
</tbody>
</table>

In sum, the greater learning rates observed for Prepositional relative clauses in written production and grammaticality judgment tests supports previous literature evidencing the role of L2 grammatical form in mediating the effects of pedagogical interventions (see section 3 in Chapter 4). This role, however, may in turn be moderated by the type of assessment task, which may pose more challenges for one grammatical structure over another. In the next section, the role of type of form in explicit vs. implicit instruction is further discussed à propos of the interaction Time x Form x Feedback.

The fourth research question addressed whether the variables investigated in this dissertation —type of learner’s Agency, type of oral Feedback, and type of grammatical Form— worked in concert towards different L2 learning outcomes. This question was answered partially affirmatively. Results revealed a significant quadruple interaction Time x Form x Agency x Feedback in both written and oral production tasks, but not in the grammaticality judgment task. This quadruple interaction indicated that among the six experimental groups compared by the ANOVA, between-group differences arose for particular combinations of Form, Agency, and Feedback. Specifically, as discussed in section 1, learner’s Agency only played a role in L2 development (written and oral production) for the dyad with minimally explicit feedback ([Practice SP] vs. [Exposure SP]) and for the easier grammatical structure (Prepositional relative clauses).

Additionally, results yielded no significant interaction Time x Agency x Feedback and Time x Form x Agency. They did, however, reveal a significant interaction Time x Form x Feedback, but only in the grammaticality judgment task. These results were corroborated when Agency was excluded from the analysis and only the Practice groups were considered. The lack of a significant interaction Time x Form x Feedback for both written and oral production was due to the fact that for each of the two grammatical forms independently, Prepositional relative clauses and Present subjunctive, a significant interaction Time x Feedback was found (or approached, in the case of Present subjunctive in oral production) where the [Practice EC] group consistently outperformed the control group [Practice NF]. However, for grammaticality judgment the edge of [Practice EC]
over [Practice NF] only materialized for Prepositional relative clauses, not for Present subjunctive.

Overall, these results suggest that in the present design, more explicit feedback, as instantiated in the [Practice EC] group, proved more beneficial than less explicit feedback for both grammatical structures, and type of Form only played a role for grammaticality judgment. This is due to the fact that the [Practice EC] group did much better on this test for Prepositional relative clauses than for Present subjunctive, experiencing gains of 3 items out of 5 for the former and barely 1 item out of 5 for the latter. The question that arises, then, is why the edge of [Practice EC] over [Practice NF] for Present subjunctive only arose in written and oral production and not in grammaticality judgment. This is likely due to both the nature of the form and the test. Overall, across all three dependent measures, the results of explicit instruction in [Practice EC] were more modest for Present subjunctive than for Prepositional relative clauses. For example, in written production the [Practice EC] group experienced immediate gains of 64% for Prepositional relative clauses, while for Present subjunctive the gains only reached 40%. Similarly, in oral production the gains were of 35% versus 28%. Finally, in grammaticality judgment the gains were of 57% versus 23%. This suggests that participants had more difficulty in understanding and/or implementing the metalinguistic information provided for Present subjunctive. In both the written and oral tests, participants could produce some items by relying on memory, as tests included new items but also old items that appeared during the treatment phase. However, for grammaticality judgment all the targeted items were new in a way, because to preclude indirect exposure to positive evidence ungrammatical items rather than grammatical ones were used (see
Chapter 5, section 3.4 for more details). For example, one targeted item in the grammaticality judgment test was: *Carmen desea conocer a un hombre que la *QUIERE [“Carmen wants to *meet-IND a man who loves her”]. This item, which is grammatically incorrect, never occurred as such in the treatment. In order to judge whether or not this item was correct, participants had to conduct some analysis, which in turn was only possible if they had understood the rules correctly. Additionally, the fact that this test was administered at the very end of the session and its multiple choice nature, with three choices (“Correct / Incorrect / I don’t know”) may have enticed participants to rush it through, which would also explain their poorer results.

Consequently, then, with the observed exception of grammaticality judgment, more explicit instruction, as instantiated in the [Practice EC] group, had consistently greater effects than less explicit instruction for both targeted forms. With regards to the edge of explicit feedback found for the posited easier structure, Prepositional relative clauses, this study aligns with de Graaff (1997) and Ellis (2007), who found that explicit feedback worked better than implicit feedback for plural noun inflection in eXperanto and the English past tense morpheme, respectively. In contrast, these findings provide counterevidence to de Graaff (1997) and Nagata and Swisher (1995), who found that more explicit feedback and less explicit feedback worked equally fine with the position of negative words in eXperanto and Japanese verb inflection in passive sentences. One possible explanation for this divergence is that the structure in the present study was not as easy as those in de Graaff (1997) and Nagata and Swisher (1995), so participants in the less explicit conditions had more problems in finding the underlying rule. However, a more plausible explanation is that neither de Graaff nor Nagata and Swisher controlled
for prior grammatical explanation. In both of these studies, participants were exposed to
the grammatical rules before the treatment practice; hence, upon making errors and
receiving feedback during the treatment, participants could rely on remembering and
applying the rules at hand rather than having to infer them, as in the present study.
Arguably, this may explain why in the present study explicit feedback with metalinguistic
information proved more beneficial. Zooming out on the literature that explored the
effects of prior metalinguistic instruction (as opposed to feedback), the present study is in
line with the overwhelming majority of previous studies that found an edge of explicit
(2003), Robinson (1996), and Scott (1989). In contrast, only two studies did not find an
edge for metalinguistic explanation with simpler grammatical structures, namely, Ayoun
(2001) and Scott (1989). As discussed in Chapter 4, the results in Ayoun (2001) may be
due to the fact that the treatment practice was qualitatively different for the more and less
explicit conditions, favoring form-meaning connections only in the latter. In turn, the lack
of effects in Scott (1989) were found for oral production only, a test which elicited poor
results overall, thus minimizing the potential gap between the two conditions. In sum,
then, the present study, in context with the previous literature, corroborates the
increasingly clear picture that for easier grammatical forms more explicit instruction has
an edge over less explicit instruction.

With regards to Present subjunctive, the purported harder structure, the obtained edge
of explicit instruction as instantiated by the [Practice EC] group was less evident.
[Practice EC] outperformed [Practice NF] in oral production and approached significance
in written production, but failed to reach significance in grammaticality judgment.
Overall, these results support de Graaff (1997), Ellis (2007), and Nagata & Swisher (1995), who found that more explicit feedback worked better than less explicit feedback for harder grammatical structures, such as the position of objects in eXperanto, the comparative morpheme in English, and Japanese particles in passive sentences. In contrast, these findings partially contradict de Graaff (1997), who did not find a significant difference between the effects of more and less explicit feedback for imperative inflection in eXperanto. Zooming out on the literature on explicit instruction via prior metalinguistic explanation, the present study corroborates that the picture for harder structures is pretty complex. The positive effects found for written and oral production add up to the evidence gathered by Gass et al. (2003), Robinson (1996), and Scott (1989), who found an advantage of more explicit instruction for sentence-initial question words in Italian bi-clausal interrogatives, English locative pseudoclefts, and written French subjunctive. In contrast, the lack of a significant difference for grammaticality judgment is in line with Alanen (1995), Robinson (1996), and Scott (1989), for Finnish locative suffix –ssa and consonant voicing $t \rightarrow d$, English locative pseudoclefts, and oral French subjunctive. Finally, this study provides direct counterevidence to Ayoun (2001) and DeKeyser (1995), who found an advantage of implicit instruction over instruction with provision of metalinguistic rules in the development of French imparfait and prototypical rules for noun and verb suffixation in Implexan, respectively. As discussed in Chapter 4, the likely cause of the edge of implicit instruction in Alanen (2001) is that this group completed practice that enforced form-meaning connections to a greater extent than the explicit group. On the contrary, the results in DeKeyser (1995) cannot be attributed to a lack of control over external
variables but rather, the length of the treatment. In the present study, participants in the
less explicit conditions, [Practice SC] and [Practice SP], completed a total of 15 items (10
targeted, 5 non-targeted) in an average of 15 minutes. In contrast, the treatment in
DeKeyser’s study evolved along 20 sessions of 25 minutes each. Therefore, DeKeyser’s
study corroborates the claim that the effects of implicit instruction are only observable in
medium- to long-term treatments (e.g., Sanz & Morgan-Short, 2005).

Overall, then, the present study found that instruction with explicit feedback as
instantiated by the [Practice EC] group proved more beneficial than less explicit
instruction for both targeted forms, the posited easier one, Prepositional relative clauses,
and the harder one, Present subjunctive. However, the edge of explicit instruction was
more evident for the easier form. This partially confirms the thesis by Reber (1993), who
claimed that the benefits of explicit instruction are more likely to arise for simpler
grammatical structures. In contrast, these findings provide counterevidence to Hulstijn
and de Graaff (1994), who claimed that the advantages of explicit instruction would be
more evident for harder grammatical structures.

5. Conclusion and implications
The findings of this study make a contribution to the theory and practice of SLA. From a
theoretical perspective, this study contributes to fill a gap in the Interaction Hypothesis,
by illustrating that the facilitative effects of the main tenants of interaction (processing
input; producing output; forming, revising, and testing hypothesis in response to
corrective feedback; and modifying output) in L2 grammar development can be mediated
by the agency of the learner, who may assume the role of a performer or observer.
Specifically, this study found that the positive effects of active participation in practice only arise for more cognitively demanding learning conditions (in output-focused tasks with corrective feedback without provision of grammatical rules) and easier grammatical structures.

Additionally, this study makes a contribution to the growing SLA literature on the effects of feedback, by identifying the relative role of two of its components, negative evidence with different degrees of explicitness and prompting for error repair. Specifically, this study found that when positive evidence cannot be accessed either prospectively or retrospectively, and feedback is delivered intensively in the context of task-essential practice, prompting for error repair yields no significant effects, but successful output modifications, which are independent from prompting, are a clear predictor of L2 grammar development. In contrast, the explicitness of negative evidence did play a facilitative role, with more explicit feedback yielding better results than less explicit feedback (clear corrective intent without rule provision) for both easier and harder grammatical structures. However, these effects were more evident for easier forms (Prepositional relative clauses, as opposed to Present subjunctive).

From a pedagogical perspective, the findings of this dissertation have implications for the development of CALI technology. Specifically, this study proved that fully computerized instruction can promote L2 grammar development with great accuracy-time ratios. For example, after only 40 minutes of total practice (i.e., roughly 20 minutes per grammatical structure), the [Practice EC] group achieved learning gains of up to 65% in written production of Prepositional relative clauses and 40% in written production of Present subjunctive.
Also, this study showed that CALI tutors can successfully go beyond traditional input-based tasks and effectively use tasks that elicit production in simulated real-life situations, to embrace the spirit of task-based instruction.

Third, this study provided evidence that not only computerized written feedback but also computerized oral feedback can have beneficial effects for L2 development, even when it is conveyed in the target language.

Finally, and more interestingly, this study helped to elucidate when computerized instruction can have an edge over traditional technologies. Specifically, it was found that the edge of CALI practice only arises when combined with minimally explicit Feedback that signals the locus and nature of the error, without explaining the underlying rule. If the rule is provided, then non-interactive technologies such as language video series showcasing interactions can work just as well. However, as demonstrated by the results of the exit questionnaire, CALI introduces the fun factor against video-based technology, as evidenced by the fact that participants in the Exposure to practice conditions significantly reported that they would have preferred to practice, whereas performers declared to be significantly more satisfied with the role that they had been assigned.

6. Limitations and further research

The first obvious limitation in this study is related to sample size. The medium to high effect sizes obtained for the significant interaction Time x Feedback on most of the assessment measures suggest that replication of this study with a larger pool of participants may elicit additional between-group differences beyond the obtained contrast
between the [Practice EP] and [Practice NF] groups, along the three levels of achievement suggested by the plotted graphs.

The second limitation concerns the choice of the targeted forms. Clearly, future studies should target other grammatical structures to avoid, to the maximum extent possible, the potential incidence of generalization of prior knowledge on related structures, as in this study, where participants may have been able to recycle their existing knowledge of the subjunctive mood to the new context of adjectival relative clauses.

The third limitation of this study concerns the hybrid modality of the treatment materials. In Talking to Avatars, all avatar-to-student interactions occurred in the oral mode. For the avatar questions, additionally, written transcripts and translations were available, since it was crucial that participants understood the context cues to make the appropriate form-meaning connections when producing output. In turn, the interactions from student-to-avatars occurred in the written mode, to bypass the current limitations of speech recognition technology for non-native speakers. This hybrid modality can arguably increase the saliency of the corrective intent in the oral feedback messages, thus turning the less explicit condition into a more explicit one. However, this study did not operationalize explicitness as the presence of a clear indication of a corrective intent, but rather, as the provision of metalinguistic rules. On that note, future research addressing the effects of computerized (as opposed to computer mediated) implicit feedback is certainly warranted.

Fourth, in the spirit of creating more authentic CALI materials, this study was the first to investigate whether computerized oral feedback in the target language could lead to L2
grammar development. Although this proved to be the case, this study failed to investigate why participants were able to use this feedback with very different rates of success. To fill in this gap, future studies should incorporate measures of L2 listening comprehension and inquire students about the aspects of the feedback that they did not understand (e.g., grammar terminology, general language) via e.g., stimulated-recall protocols.

Fifth, this study did not control for individual differences in cognitive capacity or language aptitude. Participants, however, did complete a reduced version of the MLAT. Triangulating this information with the learning outcomes in future analyses will help to elucidate if language aptitude influenced the results of this study.

Sixth, the scoring protocol followed in this study was not nuanced enough to account for advancements in the developmental sequence of Spanish prepositional relative clauses. After listening to the feedback messages, a considerable number of students did move the prepositions closer to the relative pronoun (e.g., *La persona que vivo con* [“The person who I live with”] → *La persona que con vivo* [“The person who with I live”]). However, in the conservative scoring protocol used in this study these participants did not receive partial credit. Perhaps a more nuanced coding and scoring protocol may contribute a more revealing picture.

Seventh, although this study operationalized for attention by the observers to task completion via meaning-oriented check-up questions, with overall high rates, in lack of other measures of attention it is hard to determine why some observers did not experience any learning despite receiving repeated exposure to positive evidence (provided by the performers who learned) and feedback (provided by the avatars). In this sense, eye-
tracking technology could help determine whether these observers were exclusively focusing on the content part of the questions to be able to answer the follow-up questions correctly, rather than on both the content and the grammatical forms typed in by their peers. On this token, concurrent verbalizations were excluded from the design to avoid potential reactivity issues and to avoid narrowing the “experiential” gap between the participants in the Practice and Exposure to practice conditions. However, think-aloud protocols could have been used in the testing phase to tap into the participants’ levels of noticing and awareness of the targeted forms and their learning strategies.

Eighth, the present analysis addressed the role of learner’s uptake only qualitatively. On this note, future quantitative analyses using correlations may help to elucidate more precisely the impact of different types of output modifications in L2 grammar development.

Finally, although the present study did include new and old items in the assessment tests, no separate analyses were performed. Consequently, future analyses discriminating between old and new items may provide a more nuanced picture in terms of item and system learning under each of the experimental conditions in this study.
### APPENDIX 3.1

A classification of drills with descriptions (extracted from Paulston, 1970) and some examples

<table>
<thead>
<tr>
<th>TYPE OF DRILL</th>
<th>DESCRIPTION</th>
<th>EXAMPLE</th>
<th>INTENDED GOAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanical</td>
<td>- The response is totally controlled: there is only one correct answer in one correct form; - attention to or understanding of the L2 input is not required; - ideal for choral drilling.</td>
<td>T: “Happy. The happy baby. / Big. The big baby. / Spanish. The Spanish baby. / Healthy?” L: “The healthy baby.”</td>
<td>Adjective placement (also, gender and number agreement or lack thereof)</td>
</tr>
<tr>
<td>Meaningful</td>
<td>- The response is partially controlled: there is only one correct answer but the form may vary; - attention to and understanding of the L2 input is required. - choral drilling is possible, but not automatic.</td>
<td>Comprehension of written /aural texts: T: When did the man in this story get to work? L: He got to his office at 9 o’clock. Situational teaching: T: Where is Rob in relation to Susan? L: On her right.</td>
<td>Simple past, prepositions Locative expressions</td>
</tr>
<tr>
<td>Communicative</td>
<td>- The response is not controlled at all: the answer is unpredictable and the forms may vary; - attention to and understanding of the L2 input is required. - choral drilling is impossible.</td>
<td>T: What kind of movies do you like to watch in your free time? L: I like to watch action movies / I like watching action movies / I never watch movies.</td>
<td>Free variation in “to like + infinitive verb” vs. “to like + progressive verb”</td>
</tr>
</tbody>
</table>

T: Teacher; L: Learner.
APPENDIX 3.2

Examples of feedback moves based on their individual components (explicitness of negative evidence, positive evidence, and prompts for error repair).

These examples are inspired in a communicative situation from the materials used in this dissertation. In this particular situation students take part in a simulated adventure where they describe to a real-estate agent the kind of accommodation that they are looking for during a hypothetical semester abroad in Spain. In response to the real-estate agent, the imaginary student says: *Quiero una casa que TIENE dos habitaciones* (“I want a house that *HAS-IND two bedrooms”), in lieu of the target-like utterance *Quiero una casa que TENGA dos habitaciones* (“I want a house that *HAS-SUBJ two bedrooms*”).

<table>
<thead>
<tr>
<th>IMPLICIT NEGATIVE EVIDENCE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+ PE,</td>
<td>+ PE,</td>
</tr>
<tr>
<td>- Prompt</td>
<td>+ Prompt</td>
</tr>
<tr>
<td>Ah, vale, quieres una casa que tenga dos habitaciones. Tengo dos posibilidades. Una con piscina y otra sin piscina. ¿Cuál quieres?</td>
<td>Ah, vale, quieres una casa que tenga dos habitaciones. [El avatar recibe una llamada de teléfono y no la atiende.] ¿Perdona, qué dijiste?</td>
</tr>
<tr>
<td>[Oh, ok, you want a house that has-subjunctive two bedrooms. I have two possibilities. One with a swimming pool and one without. Which one do you want?]</td>
<td>[Oh, ok, you want a house that has-subjunctive two bedrooms. [The avatar receives a phone call and doesn’t answer.] Excuse me, you were saying?]</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>- PE,</td>
<td><em>Eso no es correcto en español, pero te entiendo. ¿Te gustaría tener piscina?</em></td>
</tr>
<tr>
<td>- PE,</td>
<td>(a) <em>Eso no es correcto en español. ¿Se ocurre otra forma de decirlo?</em></td>
</tr>
<tr>
<td></td>
<td>(b) <em>Perdona, ¿qué quieres decir?</em></td>
</tr>
<tr>
<td>+ PE,</td>
<td><em>Oh, ya veo. Quieres decir “quiero una casa que tenga dos habitaciones”. ¿Te gustaría tener piscina?</em></td>
</tr>
<tr>
<td>+ PE,</td>
<td><em>Oh, ya veo. Quieres decir “quiero una casa que tenga dos habitaciones”. Por favor reformula tu frase.</em></td>
</tr>
</tbody>
</table>
EXPLICIT NEGATIVE EVIDENCE (ERROR SPOTTING)

- PE,  "Tiene" no funciona aquí, pero bueno, te entiendo. ¿Te gustaría tener piscina?

- Prompt  ['Has-indicative" does not work here, but anyway, I got you. Would you like a swimming pool?]

- PE,  (a) ¿Tiene?

+ Prompt  ['Has-indicative"?]

(b) ¿Qué quieres decir con "tiene"?
[What do you mean by "has-indicative"?]

(c) "Tiene" no funciona aquí. ¿Se te ocurre otra forma de decirlo?
['Has-indicative" does not work here. Can you think of another way to put it?]

(d) No, "tiene" no. Quieres una casa que...
[No, not "has-indicative". You want a house that…]

(e) ¿Quieres una casa que tiene? Quieres una casa que...
[You want a house that has-indicative? You want a house that…]

(f) Quieres una casa que ¿qué?
[You want a house that what?]

+ PE,  Mira, “tiene” no funciona aquí. Debes decir “tenga”, pero bueno, te entiendo. ¿Te gustaría tener piscina?

- Prompt  [Look, “has-indicative" does not work in here. You should use “has-subjunctive”, but anyway, I got you. Would you like a swimming pool?]

+ PE,  Mira, “tiene” no funciona aquí. Debes decir “tenga”. ¿Entonces, cómo reformularías tu frase?

+ Prompt  [Look, “has-indicative” does not work here. You should use “has-subjunctive”. So, how would you rephrase your sentence?]
Mira, “tiene” no funciona aquí porque es indicativo, y estás describiendo una casa genérica, es decir, tu casa ideal, no una casa concreta que estás viendo en este momento. Bueno, ¿te gustaría tener piscina?

[Look, “has-indicative” does not work here because it is indicative, and you are describing a generic house, that is, your ideal house, not a specific house that you are seeing at this moment. Now, would you like a swimming pool?]

Mira, “tiene” no funciona aquí porque es indicativo, y estás describiendo una casa genérica, es decir, tu casa ideal, no una casa concreta que estás viendo en este momento. Entonces, ¿se te ocurre otra forma de decirlo?

[Look, “has-indicative” does not work here because it is indicative, and you are describing a generic house, that is, your ideal house, not a specific house that you are seeing at this moment. So, can you think of another way to say that?]

Mira, “tiene” no funciona aquí porque es indicativo, y estás describiendo una casa genérica, es decir, tu casa ideal, no una casa concreta que estás viendo en este momento. Por eso debes decir “tenga”, que es subjuntivo. Pero ahora dime, ¿te gustaría tener piscina?

[Look, “has-indicative” does not work here because it is indicative, and you are describing a generic house, that is, your ideal house, not a specific house that you are seeing at this moment. That is why you should use “tenga”, which is subjunctive. But now tell me, would you like a swimming pool?]

Mira, “tiene” no funciona aquí porque es indicativo, y estás describiendo una casa genérica, es decir, tu casa ideal, no una casa concreta que estás viendo en este momento. Por eso debes decir “tenga”, que es subjuntivo. Entonces, cómo reformularías tu frase?

[Look, “has-indicative” does not work here because it is indicative, and you are describing a generic house, that is, your ideal house, not a specific house that you are seeing at this moment. That’s why you should use “tenga”, which is subjunctive. So, how would you rephrase your sentence?]
# APPENDIX 4.1

Types of feedback moves researched in the CALI literature

<table>
<thead>
<tr>
<th>CALI STUDY</th>
<th>TARGET FORM</th>
<th>EXPERIMENTAL TASK AND EXAMPLE OF POSSIBLE STUDENT’S RESPONSE</th>
<th>LESS EXPLICIT FEEDBACK</th>
<th>MORE EXPLICIT FEEDBACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bowles (2005)</td>
<td>Spanish dative experiencer constructions with <em>gustar</em>, e.g.:</td>
<td>• Input-focused, translation task where students have to sequentially order the Spanish words that translate an English sentence. L1 sentence: “I like biology.” Maze options: <em>Yo</em>/<em>me</em>/<em>gusto</em>/<em>gusta</em>/<em>la biología</em>.</td>
<td>“Implicit feedback”: [Visual cue: a wall closing the maze.]</td>
<td>“‘Sorry! You might think this sentence should start with ‘yo’ in Spanish, but ‘to like’ is a special construction. In Spanish, instead of using ‘yo/el/él/ellas’ to start these sentences, use the object pronouns. Click or press any key to continue.”</td>
</tr>
<tr>
<td>Camblor (2006)</td>
<td>Spanish noun-adjective gender and number agreement, e.g.:</td>
<td>• Output-focused picture description task involving production of sentences containing the name and color of an object and its location in a room.</td>
<td>“Implicit feedback”: “Here is the sentence you just composed: Las plantas ANARANJADO está sobre la mesa. Rewrite your sentence in the space below.”</td>
<td>“Explicit feedback” “Here is the sentence you just composed: Las plantas ANARANJADO está sobre la mesa. Noun-adjective agreement in this sentence is not correct! Rewrite your sentence in the space below.”</td>
</tr>
<tr>
<td>Hsieh (2007)</td>
<td>Spanish dative experiencer constructions with <em>gustar</em>, e.g.:</td>
<td>• Input-focused, picture description task where students select words to gradually compose a sentence:</td>
<td>“Implicit feedback”: [Visual cue: a red X.]</td>
<td>“Explicit feedback” “Incorrect! The person here is the indirect object of <em>gustar</em>; therefore you need the preposition <em>a</em> ‘to’ in front of a personal name.”</td>
</tr>
<tr>
<td>Me <em>gusta la biología</em></td>
<td>[<em>&quot;I like biology.&quot;</em>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camblor (2006)</td>
<td>Spanish noun-adjective gender and number agreement, e.g.:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Las plantas anaranjadas</td>
<td>[<em>&quot;The orange plants.&quot;</em>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsieh (2007)</td>
<td>Spanish dative experiencer constructions with <em>gustar</em>, e.g.:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A <em>Carmen le gustan los insectos</em></td>
<td>[<em>&quot;Carmen likes insects.&quot;</em>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CALI STUDY</td>
<td>TARGET FORM</td>
<td>EXPERIMENTAL TASK AND EXAMPLE OF POSSIBLE STUDENT’S RESPONSE</td>
<td>LESS EXPLICIT FEEDBACK</td>
<td>MORE EXPLICIT FEEDBACK</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Moreno (2007)</td>
<td>Spanish pre-verbal direct object pronouns for third person singular and plural, feminine and masculine (lo, la, los, las), e.g.: La ama Juan.</td>
<td>“Juan loves her.”</td>
<td>“[-Explicit feedback]”: [Visual cue: a red traffic light.]</td>
<td>“[+Explicit feedback]”</td>
</tr>
</tbody>
</table>

In Task 1, “forking paths,” the student selects words to compose a sentence:

Start → Lo/la → ama Juan.

In Task 2, “sentence selection,” the student selects one of two sentences.

(A) La ama Juan / (B) La ama Pepe.

Student’s response:
In task 1, the student selects the incorrect pronoun lo.
<table>
<thead>
<tr>
<th>CALI STUDY</th>
<th>TARGET FORM</th>
<th>EXPERIMENTAL TASK AND EXAMPLE OF POSSIBLE STUDENT’S RESPONSE</th>
<th>LESS EXPLICIT FEEDBACK</th>
<th>MORE EXPLICIT FEEDBACK</th>
</tr>
</thead>
</table>
| Nagata (1993) and Nagata and Swisher (1995) | Japanese particles and passivization, e.g.:  
Ee, inu ni soko o arukaremosita.  
["I was affected by a dog’s walking through there.”] | • Output-focused, translation task based on context cues.  
"Your neighbor told you that the flowers in front of your house have been spoiled. Tell her that yes, a dog went and walked there (lit. walked through there) (i.e., you were affected by a dog’s walking through there).  
Neighbor: Otaku no mae no hana ga dame ni natte imasu yo.  
You: ______.”  
• Student’s response:  
The student types in *Ee, inu ni soko de arukaremosita. | “Traditional CALI feedback”:  
“Read the following messages:  
- DE is not expected to be used here.  
- O is missing.” | “Intelligent CALI feedback”  
“Read the following messages:  
<Particle error>  
- DE is not expected to be used here.  
- O is missing  
- In your sentence, SOKO is the ‘location’ (where the activity takes place), but it should be the ‘location’ (through which the action moves). Use the particle O to mark it.” |
| Rosa and Leow (2004)              | “If” clauses in Spanish contrary-to-fact conditional sentences in the past, e.g.:  
Si Hitler hubiera ganado la Guerra…  
["If Hitler had won the War…”] | • Input-focused, multiple choice fill-in-the-blank.  
Context: Hitler no ganó la Segunda Guerra Mundial.  
Puzzle: ______ el mundo actual sería muy diferente.  
Options: Si Hitler ganara la Guerra / Si Hitler hubiera ganado la Guerra / Si Hitler habría ganado la Guerra / Si Hitler ganaría la Guerra.  
• Student’s response:  
The student drags and drop the incorrect option Si Hitler ganara la Guerra. | “Implicit feedback”:  
“This is INCORRECT. Try again!” | “Explicit feedback”  
“INCORRECT. The condition expressed in this sentence refers to the PAST. In sentences like this you use  
SI + HUBIERA + PAST PARTICIPLE  
Now click OK, take the sentence fragment out of the puzzle piece, try another one and click FEEDBACK.” |
<table>
<thead>
<tr>
<th>CALI STUDY</th>
<th>TARGET FORM</th>
<th>EXPERIMENTAL TASK AND EXAMPLE OF POSSIBLE STUDENT’S RESPONSE</th>
<th>LESS EXPLICIT FEEDBACK</th>
<th>MORE EXPLICIT FEEDBACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanz (2004) and Sanz and Morgan-Short (2004)</td>
<td>Spanish pre-verbal direct object pronouns <em>(me, te, la, nos, los, las)</em>, e.g.: <em>Las invita Manuel al cine.</em> [“Manuel asks them to the movies.”]</td>
<td>- Two input-focused interpretation tasks. In Task 1, “picture interpretation,” the student selects, out of two pictures, the one that illustrates a sentence: <em>Las invita Manuel al cine.</em> In Task 2, “article interpretation,” the students read a Spanish text and select the meaning of its title out of two possible English translations. - Student’s response: The student selects the incorrect picture or title translation.</td>
<td>“[-Explicit feedback]”: “Sorry, try again.”</td>
<td>“[+Explicit feedback]” “Sorry! Look again and see the explanation below. Always remember: a direct object pronoun <em>(me/te/la/nos/los/las)</em> is never the subject.”</td>
</tr>
</tbody>
</table>
### APPENDIX 4.2
Empirical studies on the role of type of grammatical form in more versus less explicit instruction

<table>
<thead>
<tr>
<th>STUDY</th>
<th>FORMS</th>
<th>CONDITIONS</th>
<th>DURATION</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanen</td>
<td>Manipulated Finnish</td>
<td>Control</td>
<td>2 sessions</td>
<td>Written fill-in-the-blanks, word level</td>
</tr>
<tr>
<td>(1995)</td>
<td>- Locative suffixes (difficulty NR)</td>
<td>Enhance</td>
<td>(15 min. each)</td>
<td>- (R, R&amp;E) &gt; (C, E) only for some locative suffixes and phonological changes.</td>
</tr>
<tr>
<td>N= 36</td>
<td>- Phonological changes (difficulty NR)</td>
<td>Rule</td>
<td></td>
<td>- Lack of contrasts may be due to overgeneralization of alternative structures or low frequency in input.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rule &amp; Enhance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ayoun</td>
<td>French</td>
<td>Recasts</td>
<td>5 sessions</td>
<td>Written production, text level</td>
</tr>
<tr>
<td>(2001)</td>
<td>- Passé composé (easier)</td>
<td>Describing pictures</td>
<td>(time NR)</td>
<td>- Recasts = Grammar (for passé composé)</td>
</tr>
<tr>
<td>N= 145</td>
<td>- Imparfait (harder)</td>
<td>Reading recasts (positive or negative)</td>
<td></td>
<td>- Recasts &gt; Grammar (for imparfait)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Models</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grammar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STUDY</td>
<td>FORMS</td>
<td>CONDITIONS</td>
<td>DURATION</td>
<td>RESULTS</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>de Graaff</td>
<td>eXperanto</td>
<td>Explicit</td>
<td>10 sessions (1.5 h. each)</td>
<td>Written sentence judgment</td>
</tr>
<tr>
<td>(1997) N=54</td>
<td></td>
<td>- Inflection of plural nouns (simpler)</td>
<td></td>
<td>Written sentence + correction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Inflection of imperatives (harder)</td>
<td></td>
<td>Controlled written production, word level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Syntax</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Position of negative words (simpler)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Position of objects (harder)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implicit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Inflection of imperatives (harder)</td>
<td></td>
<td>E &gt; I for imperative inflection (harder morphology) and position of...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Position of negative words (simpler)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Position of objects (harder)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Explicit</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Inflection of plural nouns (simpler)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Inflection of imperatives (harder)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Syntax</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Position of negative words (simpler)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Position of objects (harder)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeKeyser</td>
<td>Implexan</td>
<td>Implicit-inductive</td>
<td>20 sessions (25 min. each)</td>
<td>Written production, sentence level</td>
</tr>
<tr>
<td>(1995) N=44</td>
<td></td>
<td>- Noun number + case (Categorical, ~ easier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Verb number + gender (Categorical, ~ easier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Explicit-deductive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Noun number + case (Categorical, ~ easier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Verb number + gender (Categorical, ~ easier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implicit-inductive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Noun number + case (Categorical, ~ easier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Verb number + gender (Categorical, ~ easier)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STUDY</td>
<td>FORMS</td>
<td>CONDITIONS</td>
<td>DURATION</td>
<td>RESULTS</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ellis (2007) N = 34</td>
<td>English</td>
<td>Recasts</td>
<td>2 sessions (30 min. each)</td>
<td>Timed oral imitation and correction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Past tense morpheme – ed (simpler)</td>
<td></td>
<td>Written grammaticality judgment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Comparative morpheme – er (harder)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metalinguistic feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Completing story-retelling and fill-in-the-blanks activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Listening to partial recasts</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- No instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gass et al. (2003) N= 34</td>
<td>Italian</td>
<td>+Focused attention</td>
<td>1 session (50 minutes)</td>
<td>Grammaticity judgment with error correction (syntax and morphology), or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Reading text with forms underlined</td>
<td></td>
<td>translation (lexicon)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Answering questions on structures</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Reading rules/vocabulary strategies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Focused attention</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Reading text</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Completing distracting practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lexicon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Vocabulary items (~ easier)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Morphosyntax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Pre-verbal direct and indirect object pronouns (~ medium)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Syntax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Sentence-initial question-words in bi-clausal interrogatives (~ harder)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Metalinguistic > Control; Recast = Control; Metalinguistic = Recast (for both forms)

Type of form only played a role for metalinguistic feedback in ungrammatical items. Students got better at identifying a missing morpheme if it was comparative (harder) rather than past tense (easier)

[+F] > [-F] in all 3 areas for 1st y. students
[+F] > [-F] in lexicon for 2nd year students
[+F] = [-F] in all areas for 3rd year students

[+F] was more beneficial for: syntax > morphology > lexicon
[-F] was more beneficial for: lexicon > morphology > syntax
<table>
<thead>
<tr>
<th>STUDY</th>
<th>FORMS</th>
<th>CONDITIONS</th>
<th>DURATION</th>
<th>RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nagata &amp; Swisher</td>
<td>Japanese</td>
<td>Traditional CALI feedback&lt;br&gt;Reading grammar lessons&lt;br&gt;Completing output practice&lt;br&gt;Reading feedback informing of missing or unexpected words</td>
<td>4 sessions (1 h. each)</td>
<td>Written translation task, sentence level&lt;br&gt;- I-CALI &gt; T-CALI for particles (harder)&lt;br&gt;- I-CALI = T-CALI for verb conjugations (simpler)</td>
</tr>
<tr>
<td>(1995)</td>
<td>- Verb inflection in passive constructions (simpler)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N= 32</td>
<td>- Particles in passive constructions (harder)</td>
<td>Intelligent CALI feedback&lt;br&gt;Reading grammar lessons&lt;br&gt;Completing output practice&lt;br&gt;Reading feedback informing of errors and the appropriate rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robinson</td>
<td>English</td>
<td>Implicit&lt;br&gt;Memorizing models&lt;br&gt;Incidental&lt;br&gt;Answering comprehension questions</td>
<td>2 sessions (time NR, 40 models)</td>
<td>Grammaticality judgment&lt;br&gt;- Greater gain for S-V inversion in all conditions&lt;br&gt;- Instructed &gt; Rule search, Implicit, trend for Incidental, for S-V inversion (easier).&lt;br&gt;- Instructed &gt; Rule search, for pseudoclefts (harder).</td>
</tr>
<tr>
<td>(1996)</td>
<td>- S-V inversion in object topicalization (easier)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N= 104</td>
<td>- Locative pseudocleft (harder)</td>
<td>Rule search&lt;br&gt;Reading models&lt;br&gt;Inferring rules&lt;br&gt;Inferring rules&lt;br&gt;Instructed&lt;br&gt;Reading rules&lt;br&gt;Reading models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STUDY</td>
<td>FORMS</td>
<td>CONDITIONS</td>
<td>DURATION</td>
<td>RESULTS</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------</td>
<td>-----------------------------------</td>
<td>----------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Scott (1989)</td>
<td>French</td>
<td>Implicit Listening to a text with exemplars</td>
<td>6 sessions (10 min. each)</td>
<td>Written fill-in-the-blanks, word level</td>
</tr>
<tr>
<td>N= 34</td>
<td>- Relative pronouns</td>
<td>Explicit Listening to rules</td>
<td></td>
<td>- E &gt; I, for both forms independently</td>
</tr>
<tr>
<td></td>
<td>(difficulty NR)</td>
<td>Listening to models</td>
<td></td>
<td>- Greater gains for subjunctive in both conditions (NR)</td>
</tr>
<tr>
<td></td>
<td>- Subjunctive</td>
<td></td>
<td></td>
<td>- According to Ellis (2007), edge of explicit instruction arose for more complex form (i.e., subjunctive).</td>
</tr>
<tr>
<td></td>
<td>(difficulty NR)</td>
<td></td>
<td></td>
<td>Oral fill-in-the-blanks, word-level</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- E = I</td>
</tr>
</tbody>
</table>
APPENDIX 5.1 A

Script of the treatment materials: Situation A (“Finding an apartment and a roommate”)

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SpotC</td>
</tr>
<tr>
<td>1</td>
<td>Avatar</td>
<td>[A-1 ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hola, ¡bienvenido a Málaga! Me llamo Pepa.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¿Cómo te llamas?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>[Text:]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(NOMBRE)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Avatar</td>
<td>[A-2 ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Encantada. ¿Es tu primera vez en España?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>[Text:]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Sí/ No)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Avatar</td>
<td>[A-3 ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>¡Muy bien! Esta es la oficina para estudiantes extranjeros. Yo voy a ayudarte a encontrar dos cosas: primero, un apartamento para tu año académico en España, y, después, un compañero de apartamento español. ¿Estás preparado?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>[Text:]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Sí, ¡vamos!)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### FEEDBACK TYPE

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Avatar</td>
<td>[A-4 ]</td>
<td>Vale, comencemos por el apartamento. ¿En qué área de la ciudad te gustaría vivir? ¿En la playa, en el centro de la ciudad o en el campus universitario?</td>
<td>ESTÉ (OK!)</td>
<td>[A-4-SC-OK] ¡Buena elección! Y muy bien, el verbo ESTAR es correcto.</td>
<td>[A-4-SC-KO] ¡Buena elección! Y muy bien, el verbo ESTAR es correcto. Estamos hablando de UN apartamento hipotético, en tu imaginación. Hay muuuchos apartamentos en el centro, en la playa y en el campus. Por eso, usa el subjuntivo.</td>
<td>[A-4-SP-KO] ¡Buena elección! Oye, tienes un problema con el verbo ESTAR. Estamos hablando de UN apartamento hipotético, en tu imaginación. Hay muuuchos apartamentos en el centro, en la playa y en el campus. Por eso, aquí tienes que usar el subjuntivo. Pero no te preocupes, continúa.</td>
</tr>
</tbody>
</table>

**Student**

Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Quiero un apartamento (THAT IS) ____________ en (la playa / el centro / el campus).
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Avatar</td>
<td>[A-5 ]</td>
<td>ESTÁ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mira, tengo varias opciones. En los apartamentos de la carpeta violeta hay cerca una parada de autobús; en los apartamentos de la carpeta verde no hay una parada de autobús cerca. ¿Qué grupo te interesa más?</td>
<td>(OK!)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• [A-5-XC-KO] Vale, tomo nota. Oye, tienes un problema con el verbo ESTAR. Aquí no buscas, aquí seleccionas un grupo real. Estás explicando cómo es un grupo. En todos los apartamentos del grupo 1 hay una parada de autobús cerca. Es algo factual. Cuando describes cómo es algo concreto, tienes que usar el indicativo. ¿Puedes reformular tu frase, por favor?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• [A-5-XC-KO] Vale, tomo nota. Oye, tienes un problema con el verbo ESTAR. Aquí no buscas, aquí seleccionas un grupo real. Estás explicando cómo es un grupo. En todos los apartamentos del grupo 1 hay una parada de autobús cerca. Es algo factual. Cuando describes cómo es algo concreto, tienes que usar el indicativo. ¿Puedes reformular tu frase, por favor?</td>
<td></td>
</tr>
</tbody>
</table>

Student: [Text:]
Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Me interesa más el grupo de la carpeta (violeta/verde); es decir, como dices, el grupo (THAT IS) _____________ (cerca/lejos) de la parada del autobús.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td># TENGA (*KO!)</td>
<td>[A-6-SP-KO] Vale, apunto el número de habitaciones. Pero ojo: el verbo TENER no es correcto. Ya sabes, cuando expresas ideas, cuando hablas de UN apartamento ideal en tu cabeza, UN apartamento genérico, un apartamento que no conoces todavía, tienes que usar el subjuntivo. Pero no pasa nada. Sigue adelante.</td>
<td>[A-6-SP-KO] Buen número de habitaciones. Pero ojo: el verbo TENER no es correcto. Ya sabes, cuando expresas ideas, cuando hablas de UN apartamento ideal en tu cabeza, UN apartamento genérico, un apartamento que no conoces todavía, tienes que usar el subjuntivo. ¿Puedes decir tu frase de otra manera?</td>
</tr>
<tr>
<td></td>
<td></td>
<td># TENGA (*KO!)</td>
<td>[A-6-XC-KO] Buen número de habitaciones. Pero ojo: el verbo TENER no es correcto. Ya sabes, cuando expresas ideas, cuando hablas de UN apartamento ideal en tu cabeza, UN apartamento genérico, un apartamento que no conoces todavía, tienes que usar el subjuntivo. ¿Puedes decir tu frase de otra manera?</td>
<td>[A-6-XC-KO] Buen número de habitaciones. Pero ojo: el verbo TENER no es correcto. Ya sabes, cuando expresas ideas, cuando hablas de UN apartamento ideal en tu cabeza, UN apartamento genérico, un apartamento que no conoces todavía, tienes que usar el subjuntivo. ¿Puedes decir tu frase de otra manera?</td>
</tr>
</tbody>
</table>

Student | [Text:] Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(Sí/No) deseó compartir mi habitación. Quiero un apartamento (THAT HAS) _____________ (una habitación/dos habitaciones).</td>
<td></td>
</tr>
<tr>
<td>Screen #</td>
<td>Speaker</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>7</td>
<td>Avatar</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Student: [Text:]
Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Busco un apartamento (THAT COSTS) _____________ (menos de 500/de 500 a 1000/más de 1000) euros al mes.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
</table>

**Student**

[Text:]

Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Vale, pues me interesan más los apartamentos de la serie (A/ B), es decir, como tú me dices, los apartamentos (THAT HAVE/ THAT DO NOT HAVE) ___________ piscina.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Avatar</td>
<td>Excelente. Ahora vamos a hablar de animales. ¿Tienes una mascota aquí en Málaga?</td>
<td>PERMITA PERMITAN (OK!)</td>
<td>SpotC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-9]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-9-SC-OK]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vale. Y oye, ¡qué buena gramática! El verbo PERMITIR es correcto.</td>
<td>SpotP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-9-SC-KO]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vale. Oye, tienes un problema con el verbo PERMITIR, pero no pasa nada, continúa.</td>
<td>ExplainC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-9-SP-KO]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vale. Oye, tienes un problema con el verbo PERMITIR. ¿Puedes reformular tu frase, por favor?</td>
<td>ExplainP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-9-XC-KO]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vale. Oye, tienes un problema con el verbo PERMITIR. Estás expresando preferencias personales. No conoces el edificio donde vas a vivir. Puede ser un edificio grande, pequeño, blanco, negro... Por eso, tienes que usar el subjuntivo. ¡Sigamos!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-9-XP-KO]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Vale. Y oye, ¡qué buena gramática! El verbo PERMITIR es correcto. Estás expresando preferencias personales. No conoces el edificio donde vas a vivir. Puede ser un edificio grande, pequeño, blanco, negro... Por eso, usa subjuntivo.</td>
<td></td>
</tr>
</tbody>
</table>

Student [Text:]
Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

(Sí/No) tengo una mascota aquí en Málaga. (Sí/No) necesito un edificio (THAT ALLOWS) ______________ tener animales.
**Student**

Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

En los Estados Unidos vivo en un apartamento (THAT ALLOWS/THAT DOES NOT ALLOW) ______________ tener animales.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
</table>

# POSEA (*KO!)

• [A-11-XC-KO] Tomo nota. El verbo POSEER no es correcto. Todavía no conoces tu apartamento final. ¡Hay muchas posibilidades! ¡Muchos apartamentos posibles! Por eso, el indicativo no funciona. ¿Sabes una cosa? ¡Yo no puedo vivir sin aire acondicionado!  
• [A-11-XP-KO] Tomo nota. El verbo POSEER no es correcto. Todavía no conoces tu apartamento final. ¡Hay muchas posibilidades! ¡Muchos apartamentos posibles! Por eso, el indicativo no funciona. ¿Sabes una cosa? ¡Yo no puedo vivir sin aire acondicionado! ¡Ahora, ¿puedes corregir tu frase, por favor? |

Student | [Text:] Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.  
(Si/No) me gusta el aire acondicionado, así que quiero un apartamento (THAT IS EQUIPPED WITH/ THAT IS NOT EQUIPPED WITH) ___________ aire acondicionado.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-12-SC-KO] Me parece bien. Y el verbo SER es correcto. ¡Buena gramática! Todavía no conoces a tu compañero. ¡Las posibilidades son infinitas! Por eso necesitas usar el subjuntivo.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-12-SP-KO] Me parece bien, pero ¡ojo!: tienes un problema con el verbo SER. No pasa nada, continúa.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-12-SP-KO] Me parece bien, pero ¡ojo!: tienes un problema con el verbo SER. ¿Puedes reformular tu frase, por favor?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-12-XC-KO] Me parece bien, pero ¡ojo!: tienes un problema con el verbo SER. Todavía no conoces a tu compañero. ¡Las posibilidades son infinitas! Por eso el indicativo no funciona aquí. Vamos, sigamos adelante.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-12-XP-KO] Me parece bien, pero ¡ojo!: tienes un problema con el verbo SER. Todavía no conoces a tu compañero. ¡Las posibilidades son infinitas! Por eso el indicativo no funciona aquí. ¿Puedes reformular tu frase, por favor?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Student**

[Text:]
Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Busco un compañero de apartamento (THAT IS) _____________ (chico/ chica).
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
</table>

Student

Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Busco a una persona (THAT “HAS”) _____________ (de 18 a 25/ de 25 a 30 / más de 30) años.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
</table>

Student  
[Text:]  
Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.  
Busco un compañero (THAT SMOKES/THAT DOES NOT SMOKE) ____________.
FEEDBACK TYPE

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Avatar</td>
<td>¿Y tú qué piensas del consumo de alcohol?</td>
<td>BEBEN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(OK!)</td>
<td>[A-15-SC-OK] Vale, el verbo BEBER es correcto. Por cierto, ten cuidado, que en España el alcohol es demasiado barato...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≠ BEBEN</td>
<td>[A-15-SC-KO] Vale. Oye, tu verbo BEBER no es correcto, pero no pasa nada. Por cierto, ten cuidado, que en España el alcohol es demasiado barato...</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*KO!)</td>
<td>[A-15-SP-KO] Vale. Oye, tu verbo BEBER no es correcto, ¿puedes corregirlo? Por cierto, ten cuidado, que en España el alcohol es demasiado barato...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-15-XC-KO] Vale. Oye, tu verbo BEBER no es correcto, Cuando dices lo que piensas sobre el mundo, cuando expresas tu opinión sobre algo, tu percepción de la realidad, tienes que usar el indicativo. Por cierto, ten cuidado, que en España el alcohol es demasiado barato...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[A-15-XP-KO] Vale. Oye, tu verbo BEBER no es correcto. Cuando dices lo que piensas sobre el mundo, cuando expresas tu opinión sobre algo, tu percepción de la realidad, tienes que usar el indicativo. ¿Puedes corregir tu frase? Por cierto, ten cuidado, que en España el alcohol es demasiado barato...</td>
</tr>
</tbody>
</table>

Student: [Text:]
Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

En mi opinión, las personas (THAT DRINK/ THAT DO NOT DRINK) ____________ alcohol son (más divertidas/más responsables).
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [A-16-XC-OK] Buena elección. Y muy bien, el verbo VIVIR es correcto. Ahora estás expresando tu preferencia sobre una persona que no conoces. Por eso, usa el subjuntivo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [A-16-SC-KO] Buena elección. Pero ¡ojo! Tienes un problema con el verbo VIVIR. No pasa nada, te entiendo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [A-16-SP-KO] Buena elección. Pero ¡ojo! Tienes un problema con el verbo VIVIR. ¿Puedes decir tu frase de otra manera?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [A-16-XC-KO] Buena elección. Pero ¡ojo! Tienes un problema con el verbo VIVIR. Ahora estás expresando tu preferencia sobre una persona que no conoces. Por eso, usa el subjuntivo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [A-16-XP-KO] Buena elección. Pero ¡ojo! Tienes un problema con el verbo VIVIR. Ahora estás expresando tu preferencia sobre una persona que no conoces. Por eso, usa el subjuntivo. ¿Puedes decir tu frase de otra manera?</td>
</tr>
</tbody>
</table>

**Student**

[Text:] Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

Busco un compañero (THAT LIVES) ____________ (en casa/ fuera de casa) la mayor parte del tiempo.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Student**

Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate.

(Sí/ No) me importa, porque me defino como una persona (THAT IS) ____________ (tradicional/ liberal/ moderada).
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-18 -SP-KO] Ok, tomo nota. Pero atención: tienes un problema con el verbo QUERER. Tú no sabes qué cosas te va a pedir tu compañero. Son peticiones futuras e indeterminadas. Por eso, debes usar el subjuntivo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-18-XP-KO] Ok, tomo nota. Pero atención: tienes un problema con el verbo QUERER. Tú no sabes qué cosas te va a pedir tu compañero. Son peticiones futuras e indeterminadas. Por eso, debes usar el subjuntivo.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-18-SP-KO] Ok, tomo nota. Pero atención: tienes un problema con el verbo QUERER. ¿Puedes reformular tu frase, por favor?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[A-18-XC-KO] Ok, tomo nota. Pero atención: tienes un problema con el verbo QUERER. Tú no sabes qué cosas te va a pedir tu compañero. Son peticiones futuras e indeterminadas. Por eso, debes usar el subjuntivo.</td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>[Text:] Please specify the kind of apartment you are looking for. Fill in the blanks and select the options as appropriate. (Sí/no) haré todas las cosas (THAT HE/SHE WANTS) _____________.</td>
<td></td>
<td>[A-18-XP-KO] Ok, tomo nota. Pero atención: tienes un problema con el verbo QUERER. ¿Puedes reformular tu frase, por favor?</td>
</tr>
<tr>
<td>Screen #</td>
<td>Speaker</td>
<td>Message</td>
<td>Answer</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Avatar</td>
<td>[A-19 ] Perfecto, pues ya tengo toda la información que necesito. Te llamaremos en un par de días. ¡Gracias y bienvenido a Málaga!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### APPENDIX 5.1 B

Script of the treatment materials: Situation B (“Reporting a theft”)

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.</td>
<td>Avatar</td>
<td>Hola, bienvenido a la Comisaría de la Policía de Málaga. Antes de continuar necesito tu nombre completo, tu edad y tu nacionalidad.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Student [Text:](NOMBRE, EDAD Y NACIONALIDAD)
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(OK!)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student</th>
<th>[Text:</th>
<th>Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Porque quiero practicar español y España es un país (THAT I LIKE) ________ (bastante/mucho).</td>
<td></td>
</tr>
<tr>
<td>Screen #</td>
<td>Speaker</td>
<td>Message</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Student [Text:] Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

La razón (THAT I AM HERE FOR) __________ es que me han robado dinero.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>Avatar</td>
<td>[B-23 ] Bueno, y ¿cuánto dinero te han robado?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td></td>
<td>[Text:] Me han robado _______ euros.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
[B-24] Lo siento. Bueno, mi experiencia como policía me dice que el dinero te lo ha robado alguien que tú conoces. Probablemente un amigo, tu compañero de apartamento, la persona que limpia tu casa, etcétera. Bien, vives con un compañero de apartamento, ¿verdad? ¿Cómo lo conociste?

*[Text:] En la Universidad, gracias a un servicio (THAT PROVIDES HELP TO AMERICAN STUDENTS) ___________ como yo…
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• [B-25-SP-KO] Pepa, hablaré con ella. Oye, la posición de la preposición CON no es correcta. ¿Puedes corregir tu frase?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• [B-6-XP-KO] Pepa, hablaré con ella. Oye, la posición de la preposición CON no es correcta. En inglés, dicen la preposición al final. Por ejemplo: THAT bla, bla... WITH, es Pepa. Pero en español, no. En español ponemos la preposición CON antes del pronombre relativo QUE. ¿Puedes corregir tu frase?</td>
</tr>
</tbody>
</table>

Student [Text:]
Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

Sí, el nombre de la persona (THAT I SPOKE WITH) es Pepa.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>Avatar</td>
<td>Bueno, volviendo a tu compañero de apartamento, ¿qué es lo que hace? ¿Estudia o trabaja?</td>
<td>CON_QUE (OK!)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[B-26-SP-KO] Vale, gracias por la información. OJO: Tienes un problema con la preposición CON. Su posición en la frase no es correcta. A ver, intenta corregir tu frase…</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[B-26-XC-KO] Vale, gracias por la información. OJO: Tienes un problema con la preposición CON. Su posición en la frase no es correcta. En inglés, puedes poner la preposición WITH al final de una frase, pero en español, no. En español tienes que poner la preposición CON justo detrás del antecedente, PERSONA. Pero comprendo lo que quieres decir. Continuemos.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[B-7-XP-KO] Vale, gracias por la información. OJO: Tienes un problema con la preposición CON. Su posición en la frase no es correcta. En inglés, puedes poner la preposición WITH al final de una frase, pero en español, no. En español tienes que poner la preposición CON justo detrás del antecedente, PERSONA. A ver, intenta corregir tu frase…</td>
</tr>
</tbody>
</table>

Student [Text:] Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

La persona (THAT I LIVE WITH) ___________ (es estudiante/trabaja).
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>FEEDBACK TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [B-27-SP-KO] Ya veo, la señora de la limpieza. Oye, tu cláusula relativa no es correcta. ¿Puedes corregir tu frase?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [B-27-XC-KO] Ya veo, la señora de la limpieza. Oye, tu cláusula relativa no es correcta. Tienes que usar el pronombre relativo QUE inmediatamente después de su antecedente, PERSONA. En esta oración, en inglés, hay una preposición, TO. Pero la preposición TO no está al final de la cláusula. Hay un objeto detrás: &quot;nuestra casa&quot;. Por eso, en inglés y en español el orden de las palabras es igual, no cambia.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• [B-8-XP-KO] Ya veo, la señora de la limpieza. Oye, tu cláusula relativa no es correcta. Tienes que usar el pronombre relativo QUE inmediatamente después de su antecedente, PERSONA. En esta oración, en inglés, hay una preposición, TO. Pero la preposición TO no está al final de la cláusula. Hay un objeto detrás: &quot;nuestra casa&quot;. Por eso, aquí el orden de las palabras es igual en inglés y en español, no cambia. ¿Puedes corregir tu frase?</td>
</tr>
</tbody>
</table>

Student: [Text:] Sí, hay otra persona (THAT COMES TO OUR HOUSE) _______ tres veces a la semana: la señora de la limpieza.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen #</td>
<td>Speaker</td>
<td>Message</td>
<td>Answer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Student | Text: | Estaba dentro de un libro (THAT IS NEXT TO MY BED) ________.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
</table>


|       |         | (*KO!)   | [B-30-SP-KO] Vale. Atención: tienes un error con la posición de la preposición EN. Ya sabes que en español nunca podemos terminar una frase con una preposición. En español es necesario poner la preposición EN antes del pronombre relativo QUE. ¿Puedes corregir tu error? | [B-30-SP-KO] Vale. Atención: tienes un error con la posición de la preposición EN. Ya sabes que en español nunca podemos terminar una frase con una preposición. En español es necesario poner la preposición EN antes del pronombre relativo QUE. ¿Puedes corregir tu error? |

**Student**

[Text:]
Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

No. No creo que nadie supiera el lugar (THAT I KEEP MY MONEY IN) ________.
<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>Avatar</td>
<td>¿Hay alguna manera de ver el interior de tu habitación desde fuera? ¿Hay algún agujero en la puerta?</td>
<td>POR_QUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(OK!)</td>
<td>• [B-31 -SC-OK]</td>
<td>Ya veo, muy interesante. Y bien, has colocado la preposición POR correctamente.</td>
<td>• [B-31 -XC-OK]</td>
<td>Ya veo, muy interesante. Y bien, has colocado la preposición POR correctamente. Como sabes, en español nunca podemos terminar una frase con una preposición. Aquí debes poner la preposición POR justo después de su antecedente, la PUERTA.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>≠ POR_QUE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(*KO!)</td>
<td>• [B-31-SC-KO]</td>
<td>Ya veo, muy interesante. Pero cuidado, porque no has colocado la preposición POR correctamente. De todos modos, continuemos.</td>
<td>• [B-31-SP-KO]</td>
<td>Ya veo, muy interesante. Pero cuidado, porque no has colocado la preposición POR correctamente. Como sabes, en español nunca podemos terminar una frase con una preposición. Aquí debes poner la preposición POR justo después de su antecedente, la PUERTA. De todos modos, continuemos.</td>
</tr>
</tbody>
</table>

Student | [Text:] | Please answer the police officer’s question. Fill in the blanks and select the options as appropriate. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sí, hay un agujero en la puerta (THAT YOU CAN SEE THROUGH) _________.</td>
</tr>
</tbody>
</table>
Entonces alguien pudo mirar por el agujero y descubrir dónde guardas tu dinero. ¿Qué te dice tu intuición? ¿Piensas que el dinero te lo ha robado tu compañero de apartamento o la señora de la limpieza?

Creo que la persona (THAT ROBBED MY MONEY) _______ puede ser (mi compañero de cuarto/la señora de la limpieza).
Yo personalmente pienso que el dinero lo robó la señora de la limpieza. Ella no vive en la casa, así que puede coger el dinero y esconderlo fuera. ¿Qué piensas de mi hipótesis?

DE_QUE

¡Qué buena gramática! Veo que estás aprendiendo… ☻ El orden de la preposición DE es correcto. Y bueno… ¡ya veremos quién es el ladrón!

# DE_QUE

Cuidado con tu gramática. El orden de la preposición DE no está bien. Pero bueno… ¡a ver si descubrimos quién es el ladrón!

(*KO!)

¡Qué buena gramática! Veo que estás aprendiendo… ☻ El orden de la preposición DE es correcto. En inglés puedes decir THAT bla bla bla ABOUT. En español, no. Debes poner la preposición DE antes del pronombre relativo QUE. Y bueno… ¡ya veremos quién es el ladrón!

Student

[Text:]

Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

Creo que la hipótesis (THAT YOU TALK ABOUT) _______ es (probable/improbable).
Ayer quería salir a cenar con amigos a un restaurante. Fui a mi habitación para coger dinero, pero el dinero no estaba en el libro (THAT I PUT IT IN) _______.

FEEDBACK TYPE

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
<th>SpotC</th>
<th>SpotP</th>
<th>ExplainC</th>
<th>ExplainP</th>
</tr>
</thead>
</table>

Student
### FEEDBACK TYPE

<table>
<thead>
<tr>
<th>Screen #</th>
<th>Speaker</th>
<th>Message</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td># DESDE QUE (*KO!) • [B-35-SC-KO] Vaya, eso es muy interesante. Pero OJO, la preposición DESDE no está en el lugar correcto. De todos modos, te entiendo, así que vamos a continuar.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Student**

[Text:] Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

Sí. Alguien llamó a mi casa muchas veces ayer, pero no pude reconocer el número de teléfono (THAT THEY WERE CALLING FROM) _____.

**Student**

[Text:] Please answer the police officer’s question. Fill in the blanks and select the options as appropriate.

Sí. Alguien llamó a mi casa muchas veces ayer, pero no pude reconocer el número de teléfono (THAT THEY WERE CALLING FROM) _____.
El nombre de la compañía (THAT I HAVE MY PHONE WITH) ________ es Cherrytel.
[B-37 ]
Déjame acceder a los registros de Cherrytel. 
Hmmm, sí, hay un número que se repite. [Coge el teléfono:] Hola, soy el inspector García. 
Estoy investigando un robo a un estudiante americano. Usted lo llamó por teléfono ayer varias veces. ¿Por qué? [Espera unos segundos]. Ah, ya veo… [Mirando a cámara] 
Es tu compañero de apartamento. Dice que encontró el dinero en tu libro por accidente y lo tomó prestado. Dice que te llamó varias veces para decírtelo, pero tú no contestabas el teléfono… Bueno, ¡problema resuelto!
APPENDIX 5.2 A

Untimed written production test: Pretest / Delayed posttest

Instructions: In the next pages you will be presented with a number of situations that you may encounter in real life and how you could react to them. The situations are in English, but your answers should be in Spanish. In order to help you, the beginning of each answer is already provided. Simply complete your answers by translating the CAPITALIZED English bits into their Spanish equivalents.

If you don't know a word, hover your mouse over it and you will get a Spanish gloss. Not all words are glossed. Just the most difficult ones. If you don't know a word, improvise. Please do not worry if you do not know how to express something in Spanish. I am more interested in your intuitions than anything else. Thanks!

1. At a shoe store, you want the employee to get you a size 10 of a particular model you like.

   YOU SAY: “Puede darme una talla 10 de los zapatos verdes THAT ARE BELOW THE ORANGE ONES?”

2. At the accommodation office, the agent is curious about the pet policy in your building in America.

   YOU SAY: “Vivo en un edificio THAT DOES NOT ALLOW ANIMALS IN THE COMMON AREAS.”

3. It's Spring break and when you get to your hotel they want to put you in a cheaper room.

   YOU SAY: “Quiero estar en la habitación THAT I'M PAYING FOR.”

4. The person sitting next to you on the plane has a heart attack.

   YOU YELL: “Is there anyone here WHO IS A DOCTOR?”
5. While in Spain, your host family asks you about the U.S. policy on guns.
   YOU SAY: “Hay muchas personas WHO HAVE GUNS IN THE USA.”

6. You are a Spanish teacher at a DC school and want to encourage your students to finish their tasks on time.
   YOU SAY: “Daré dos puntos extra a la persona WHO FINISHES FIRST.”

7. You are interviewing Obama and want to learn details about his personal life.
   YOU SAY: “Dígame tres cosas THAT YOU CANNOT LIVE WITHOUT.”

8. You get back to campus after Spring break and tell a friend about your vacation...
   YOU SAY: “Llegamos allí y el hotel THAT WE'RE STAYING AT IS TERRIBLE.”

9. You go to the police station to report a theft. The officer asks you why you're there.
   YOU SAY: “La razón THAT I’M HERE FOR IS...”

10. At the accommodation office, the agent asks you to choose between the apartments in the purple folder, that are close to a bus stop, and the ones in the green folder, that are not.
    YOU SAY: “Prefiero el grupo en la carpeta violeta, es decir, el grupo THAT IS CLOSE TO A BUS STOP.”

11. You go to the police station to report a theft. The officer asks you what your roommate does for a living.
    YOU SAY: “La persona THAT I LIVE WITH IS A STUDENT TOO.”

12. You go to the police station to report a theft. The officer asks you the name of the company of your cleaning lady.
    YOU SAY: “Lo siento, no recuerdo el nombre de la compañía THAT SHE WORKS FOR.”
13. You go to the police station to report a theft. The officer asks you if you noticed some suspicious activity lately. You received a number of calls from an unknown number.

YOU SAY: “No pude identificar el número THAT THEY WERE CALLING FROM.”

14. You go to the police station to report a theft. The officer asks you how you found your roommate.

YOU SAY: “En la Universidad, gracias a un servicio THAT PROVIDES HELP TO FOREIGN STUDENTS.”

15. You go to the police station to report a theft. The officer asks you if someone else other than your roommate has access to your apartment.

YOU SAY: “Sí, la señora de la limpieza THAT COMES TO OUR HOUSE 3 TIMES A WEEK.”

16. You go to the police station to report a theft. The officer asks you where you kept the money that got stolen.

YOU SAY: “Dentro de un libro THAT IS NEXT TO MY BED.”

17. You go to the police station to report a theft. The officer thinks the thief is your cleaning lady.

YOU SAY: “Pienso que la hipótesis THAT YOU’RE TALKING ABOUT IS UNLIKELY.”

18. You have a pet and need an apartment that allows pets in the building. You are at the University’s accommodations office.

YOU SAY: “Necesito encontrar un edificio THAT ALLOWS PETS.”

19. You just arrived in Spain and are looking for an apartment. You’re at the University’s accommodations office.

YOU SAY: “Me gustaría tener un apartamento THAT IS ON CAMPUS.”
20. You meet a friend you haven't seen in a while and are catching up with her.

YOU SAY: “Hey, te acuerdas de ese chico español THAT I ALWAYS TELL YOU ABOUT? HE’S MARRIED!”

21. Your mom asks you if you have her car keys.

YOU SAY: “Mamá, siempre las pongo en la caja THAT YOU PUT THEM IN“

22. You need to find a cheap apartment for your semester abroad in Spain. You are at the accommodations office.

YOU SAY: “Busco un apartamento THAT COSTS LESS THAN 500 EUROS A MONTH.”

23. You prefer a non-smoking roommate. You are at the University's accommodations office.

YOU SAY: “Prefiero un compañero de apartamento THAT DOESN’T SMOKE.”

24. You saw this really cool cooking program featuring Spanish cook José Andrés and tell your mom about it. She asks you who José Andrés is.

YOU SAY: “Es un hombre THAT COOKS SPANISH FOOD ON PBS.”

25. You work at the Office of International Programs at GU and need a French speaking assistant. You write a job ad.

YOU SAY: “La OIP necesita un ayudante WHO SPEAKS FRENCH.”

26. You’re a fan of Penélope Cruz but your friend has no clue who she is.

YOU SAY: “Es la actriz THAT PLAYS THE CRAZY WOMAN IN VICKY, CRISTINA, BARCELONA.”

27. You’re lost in Madrid and need to get back to your hotel. You ask a passerby for help.

YOU SAY: “Perdone, necesito encontrar un autobús THAT TAKES ME TO THE PLAZA MAYOR.”
28. You're writing a profile on Match.com and are filling in the “About your date” field.

YOU SAY: “Busco a alguien WHO GETS ME.”

29. Your semester abroad program requires you to share an apartment with a native speaker, but you want a private room. You are at the University's accommodations office.

YOU SAY: “Necesito un apartamento THAT HAS 2 ROOMS.”

30. You're sitting in class and you fancy one of the students. You're chatting with your friend.

YOU SAY: “Me gusta el chico/la chica THAT IS WEARING THE RED SHIRT BEHIND THE BIG GUY.”
APPENDIX 5.2 B

Untimed written production test: Immediate posttest

In the next pages you will be presented with a number of situations that you may encounter in real life and how you could react to them. The situations are in English, but your answers should be in Spanish. In order to help you, the beginning of each answer is already provided. Simply complete your answers by translating the CAPITALIZED English bits into their Spanish equivalents.

*If you don't know a word, hover your mouse over it and you will get a Spanish gloss* Not all words are glossed. Just the most difficult ones. If you don't know a word, improvise. Please do not worry if you do not know how to express something in Spanish. I am more interested in your intuitions than anything else. Thanks!

1. *At the accommodation office, the agent is curious about the pet policy in your building in America.*

   YOU SAY: “Vivo en un edificio THAT DOES NOT ALLOW ANIMALS IN THE COMMON AREAS.”

2. *While in Spain you learn about a terrorist group that supports the independence of the Basque country. Your host family asks you what you think of terrorism.*

   YOU SAY: “El terrorismo es un problema THAT EVERY COUNTRY SHOULD FIGHT AGAINST.”

3. *While in Spain, your host family asks you about the American lifestyle.*

   YOU SAY: “En los Estados Unidos hay demasiadas personas WHO LIVE TO WORK.”

4. *You lost your dog and decide to write an ad offering a reward.*

   YOUR AD SAYS: “Regalaré 200 euros a la persona WHO FINDS MY DOG.”
5. You go to the police station to report a theft. The officer asks you why you're there.

YOU SAY: “La razón THAT I'M HERE FOR IS...”

6. It's kind of late, you want to leave the club but the coat check guy can't find your coat.

YOU SAY: “Es el abrigo azul THAT IS BETWEEN THE BLACK AND THE WHITE COATS.”

7. At the accommodation office, the agent asks you to choose between the apartments in the purple folder, that are close to a bus stop, and the ones in the green folder, that are not.

YOU SAY: “Prefiero el grupo en la carpeta violeta, es decir, el grupo THAT IS CLOSE TO A BUS STOP.”

8. You go to the police station to report a theft. The officer asks you what your roommate does for a living.

YOU SAY: “La persona THAT LIVE WITH IS A STUDENT TOO.”

9. You go to the police station to report a theft. The officer asks you the name of the company of your cleaning lady.

YOU SAY: “Lo siento, no recuerdo el nombre de la compañía THAT SHE WORKS FOR.”

10. You go to the police station to report a theft. The officer asks you if you noticed some suspicious activity lately. You received a number of calls from an unknown number.

YOU SAY: “No pude identificar el número THAT THEY WERE CALLING FROM.”

11. Your next door neighbors had a big fight last night and the police came to your place to ask you some questions...

YOU SAY: “Los oí gritar, pero ignoro el motivo THAT THEY WERE ARGUING ABOUT.”
12. You go to the police station to report a theft. The officer asks you how you found your roommate.

YOU SAY: “En la Universidad, gracias a un servicio THAT PROVIDES HELP TO FOREIGN STUDENTS.”

13. You go to the police station to report a theft. The officer asks you where you kept the money that got stolen.

YOU SAY: “Dentro de un libro THAT IS NEXT TO MY BED.”

14. You go to the police station to report a theft. The officer asks you if someone else other than your roommate has access to your apartment.

YOU SAY: “Sí, la señora de la limpieza THAT COMES TO OUR HOUSE 3 TIMES A WEEK.”

15. You go to the police station to report a theft. The officer thinks the thief is your cleaning lady.

YOU SAY: “Pienso que la hipótesis THAT YOU'RE TALKING ABOUT IS UNLIKELY.”

16. You read an article on child exploitation in Latin America and tell a friend about it.

YOU SAY: “Es horrible ver las condiciones THAT THEY LIVE UNDER.”

17. You accidentally stained your apartment's carpet and need a stain remover, so you go to the nearest drug store.

YOU SAY: “Busco un producto THAT REMOVES STAINS FROM CARPETS.”

18. At a job interview, your potential employer asks you about your current job.

YOU SAY: “Trabajo en un banco. Soy la persona WHO GIVES LOANS TO STUDENTS.”
19. You have a pet and need an apartment that allows pets in the building. You are at the University's accommodations office.

YOU SAY: “Necesito encontrar un edificio THAT ALLOWS PETS.”

20. The Spanish class is over and someone forgot his/her cell phone on a chair. You ask your professor if he has a clue whose phone it is.

YOU SAY: “¿Por casualidad conoce a la persona THAT THIS PHONE BELONGS TO?”

21. You are watching the news with a friend and there's a report on people addicted to plastic surgery.

YOU SAY: “No me puedo creer el mundo THAT WE LIVE IN.”

22. You need to find a cheap apartment for your semester abroad in Spain. You are at the accommodations office.

YOU SAY: “Busco un apartamento THAT COSTS LESS THAN 500 EUROS A MONTH.”

23. You prefer a non-smoking roommate. You are at the University's accommodations office.

YOU SAY: “Prefiero un compañero de apartamento THAT DOESN'T SMOKE.”

24. While in Spain, your host family's dad asks you who The Jonas Brothers are.

YOU SAY: “Son unos hermanos WHO SING SONGS FOR TEENAGERS.”

25. At a restaurant, you ask the host about a friend you're supposed to be meeting there.

YOU SAY: “Estoy buscando a un amigo WHO HAS A PIERCING ON HIS NOSE.”

26. You are at a stationer's in Spain and want to buy a stapler, but you don't know the word.

YOU ASK: “¿Hay alguien aquí WHO UNDERSTANDS ENGLISH?”
27. You broke your arm and need to find someone who can type a paper for you. You post an ad in the University’s website.

YOU SAY: “Necesito una persona WHO CAN TYPE.”

28. You're going to a party, it's late, and forgot to buy a bottle of wine for the host. You ask a passerby for help.

YOU SAY: “Perdone, ¿conoce alguna tienda THAT SELLS ALCOHOL THIS LATE?”

29. Your semester abroad program requires you to share an apartment with a native speaker, but you want a private room. You are at the University's accommodations office.

YOU SAY: “Necesito un apartamento THAT HAS 2 ROOMS.”

30. You just arrived in Spain and are looking for an apartment. You're at the University's accommodations office.

YOU SAY: “Me gustaría tener un apartamento THAT IS ON CAMPUS.”
APPENDIX 5.3 A

Timed oral production test: Pretest / Delayed posttest (1/18)

Oral Production Task

- Hi!!!
- The following is an automated presentation.
- Please click your mouse only when you see a "Next" button (like the one which will appear in 3 seconds).

- In this task you will see a picture of a person or a thing and 2 sentences with information about it.
- Your goal is to link these 2 sentences into 1 by using the relative pronoun "QUE" and including all the words in RED color.
- State your answers orally by talking clearly into the microphone.
- You'll have 7 seconds to provide an answer. If you are unsure, please don't remain quiet, just give it your best shot.

- Let's start with a couple of examples! This will give you a clear idea of what to do.

PEDRO

- Soy director de cine (I'm a filmmaker).
- HACER películas en España (To make films in Spain).
- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:
- Pedro es un director de cine...

- Got it?
- Let's see one more example.
- This time, practice as if it were the real test. Don't forget to speak up into the microphone.
Timed oral production test: Pretest / Delayed posttest (2/18)

- ARE YOU READY?
  
  - Great! Now please call Luis or the research assistant so they can start the audio recording.

- From now on, this presentation will proceed automatically.
  - When the dialog bubble pops up, speak clearly into the microphone. You'll have 7 seconds. If you are unsure what to say, just trust your gut and give it a shot!!
  - Click NEXT now to get started and please DON'T click your mouse or hit any key for the remainder of the slideshow.

- Remember to speak clearly into the microphone when the dialog bubble pops up!!

- Ok, this will start in 3, 2, 1...
Timed oral production test: Pretest / Delayed posttest (3/18)

CAROLINA

Tengo un hijo de 20 años.
(I have a 20 year old son)
ESTUDIAN EN Georgetown.
(He studies at Georgetown)

- Use the relative pronoun QUE
- Use all the RED words.
- Start with the following sequence:

  Carolina tiene un hijo de 20 años...

CARMEN

Me gustan los hombres moderados.
(I like mellow men)
Sólo BEBER alcohol CON las comidas.
(They only drink alcohol with their meals)

- Use the relative pronoun QUE
- Use all the RED words.
- Start with the following sequence:

  A Carmen le gustan los hombres moderados...
Timed oral production test: Pretest / Delayed posttest (4/18)

**FELIPE**

*Busco un compañero de apartamento.*
(I'm looking for a roommate)

*VIVIR mucho fuera de casa.*
(S/he is often out of the apartment)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:

  - *Felipe busca un compañero de apartamento...*

*Speak up now! (7 seconds)*

**GUILLERMO**

*Me defino como un político conservador.*
(I define myself as a conservative)

*SER liberal EN materia social.*
(I am liberal with social matters)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:

  - *Guillermo se define como un político conservador...*

*Speak up now! (7 seconds)*
Timed oral production test: Pretest / Delayed posttest (5/18)

**RAFAEL**

**VIVIR sólo PARA una cosa.**
(I live for just one thing)

*El tenis.*
(Tennis)

- Use the relative pronoun **QUE**
- Use all the RED words.
- Start with the following sequence:
  - *El tenis es la única cosa...*

**ROSA**

*Ayer FUI A HABLAR CON una consejera universitaria.*
(Yesterday I went to talk to a university counselor)

*Pepa.*
(Her name is Pepa)

- Use the relative pronoun **QUE**
- Use all the RED words.
- Start with the following sequence:
  - *Pepa es la consejera universitaria...*
Timed oral production test: Pretest / Delayed posttest (6/18)

**JAVIER**

*Soy piloto.*
(I'm a pilot)

*VIVIR ENTRE Washington y Madrid.*
(I live between Washington and Madrid)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Javier es un piloto...

**Sergi**

*Necesito nuevos amigos.*
(I need new friends)

*PRACTICAR yoga conmigo.*
(They practice yoga with me)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Sergi necesita nuevos amigos...
Timed oral production test: Pretest / Delayed posttest (7/18)

KASEY

_Necesito encontrar un trabajo._
(I need to find a job)

_PAGAR bien._
(It pays well)

- Use the relative pronoun _QUE_.
- Use all the RED words.
- Start with the following sequence:

  - Kasey necesita encontrar un trabajo...

Speak up now! (7 seconds)

ANTONIO

_Qüiero empezar un negocio._
(I want to start a business)

_HACERME millonario._
(It makes me a millionaire)

- Use the relative pronoun _QUE_.
- Use all the RED words.
- Start with the following sequence:

  - Antonio quiere empezar un negocio...

Speak up now! (7 seconds)
Timed oral production test: Pretest / Delayed posttest (8/18)

KEVIN

*Soy un estudiante de intercambio.*
(I'm an exchange student)
*ADORAR estudiar EN España.*
(I love studying in Spain)

- Use the relative pronoun **QUE**.
- Use all the RED words.
- Start with the following sequence:

  - **Kevin es un estudiante de intercambio...**

Speak up now! (7 seconds)

CONSUELO

*Soy empleada doméstica.*
(I'm a housekeeper)
*A veces ROBAR dinero DE las casas.*
(Sometimes I steal money from the houses)

- Use the relative pronoun **QUE**.
- Use all the RED words.
- Start with the following sequence:

  - **Consuelo es una empleada doméstica...**

Speak up now! (7 seconds)
Timed oral production test: Pretest / Delayed posttest (9/18)

CRISTINA

Deseo vivir en un país multicultural.
(I want to live in a multicultural country)
HABLAR varias lenguas.
(It has several languages)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Cristina desea vivir en un país multicultural...

Speak up now!
(7 seconds)

Una PUERTA

En la puerta hay un agujero.
(There is a peephole in the door)
Tú PODER VER el otro lado POR el agujero.
(You can see the other side through the peephole)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - En la puerta hay un agujero...

Speak up now!
(7 seconds)
Timed oral production test: Pretest / Delayed posttest (10/18)

LAURA

QUIERO ENCONTRAR UN COMPAÑERO DE APARTAMENTO.
(I want to find a roommate)

QUE IMPARAR LA CASA FRECUENTEMENTE.
(She wants to clean the house often)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  Laura quiere encontrar un compañero...

LUCÍA

BUSCO UN COMPAÑERO DE APARTAMENTO.
(I'm looking for a roommate)

TENER ÚLTIMO DE 25 AÑOS.
(S/he is older than 25)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  Laura busca un compañero de apartamento...
Timed oral production test: Pretest / Delayed posttest (11/18)

**MAOLE**

*ESCRIBIR reportajes.*
(I write reports)

*Generalmente SOBRE política.*
(They’re generally about politics)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:
  
  - **La política es el tema...**

**MARIANO**

*Soy político.*
(I’m a politician)

*ESTAR CONTRA el matrimonio gay.*
(I’m against gay marriage)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:
  
  - **Mariano es un político...**
Timed oral production test: Pretest / Delayed posttest (12/18)

**ÓSCAR**

*Quiero encontrar un apartamento.*
(I want to find an apartment)

*ESTAR equipado con aire acondicionado.*
(It is equipped with air conditioning)

- Use the relative pronoun **QUE**
- Use all the RED words.
- Start with the following sequence:

  - Óscar quiere encontrar un apartamento...

**PILAR**

*Vivo en un apartamento barato.*
(I live in a cheap apartment)

*No TENER piscina EN el edificio.*
(It doesn’t have a pool in the building)

- Use the relative pronoun **QUE**
- Use all the RED words.
- Start with the following sequence:

  - Pilar vive en un apartamento barato...
Timed oral production test: Pretest / Delayed posttest (13/18)

RON

Estoy pensando en cambiar mi coche por algún otro.
(I'm thinking about trading in my car for some other one)
GASTAR menos gasolina.
(It uses less gas)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  Ron piensa en cambiar su coche por algún otro...

ALFONSO

TRABAJAR BAJO mucha presión.
(I work under a lot of stress)
No puedo soportarlo más.
(I can't take it anymore)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  Alfonso ya no puede soportar la presión...
CARLOS

Mi teléfono OPERAR CON una compañía telefónica.
(My phone plan is with a phone company)
La compañía es Verizon.
(The company is Verizon)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  • Verizon es la compañía...

Speak up now! (7 seconds)

JUAN

Busco un compañero de apartamento.
(I'm looking for a roommate)
SER chico.
(He is a guy)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  • Juan busca un compañero de apartamento...
Timed oral production test: Pretest / Delayed posttest (15/18)

**ELENA**

_Soy psicóloga._
(I'm a therapist)

_TRATAR a pacientes CON problemas de identidad._
(I treat patients with identity issues)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Elena es una psicóloga...

**MAITE**

_VIVIR EN una ciudad española._
(I live in a Spanish city)

_Málaga._
(Málaga)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Málaga es la ciudad española...
Timed oral production test: Pretest / Delayed posttest (16/18)

### JASON

Tengo un primo muy inteligente.
(I have a smart cousin)

TRABAJAR PARA el gobierno.
(He works for the government)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Jason tiene un primo muy inteligente...

### PABLO

GUARDAR mi dinero **EN** un lugar secreto.
(I hide my money in a secret place)

**Nadie conoce ese lugar.**
(Nobody knows this place)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Nadie conoce el lugar secreto...
Timed oral production test: Pretest / Delayed posttest (17/18)

**LUCAS**

Ayer **PONER** mi dinero **EN** un libro.
(Yesterday I put my money in a book)

*Hoy el dinero no está ahí.*
(Today the money is gone)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  
  - Hoy mi dinero no está en el libro...

**PEPE**

**ESTAR** orgulloso **DE** una cosa.
(I'm proud of something)

*Mi hijo Juan.*
(My son Juan)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  
  - Juan es la cosa...
• You’re done with the oral test!
• Please call Luis or a researcher assistant to move on to the written test.
• Remind them to stop and label your audio-recording! 😊
APPENDIX 5.3 B

Timed oral production test: Immediate posttest (1/15)

**ÍKER**

*Soy portero.*
(I'm a goalkeeper)

*JUGAR EN el Real Madrid.*
(I play for Real Madrid)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - **Íker es un portero...**

**POTOMAC**

*PASAR POR muchas ciudades.*
(It passes through many cities)

*Una es Washington.*
(One of them is Washington)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - **Washington es una de las ciudades...**

Speak up now! (7 seconds)
Timed oral production test: Immediate posttest (2/15)

**FELIPE**

*Busco un compañero de apartamento.*  
(I'm looking for a roommate)  
**VIVIR mucho fuera de casa.**  
(S/he is often out of the apartment)

- Use the relative pronoun **QUE**.  
- Use all the **RED** words.  
- Start with the following sequence:

  - *Felipe busca un compañero de apartamento...*

**ROSA**

*Ayer FUI A HABLAR CON una consejera universitaria.*  
(Yesterday I went to talk to a university counselor)  
**Pepa.**  
[Her name is Pepa]

- Use the relative pronoun **QUE**.  
- Use all the **RED** words.  
- Start with the following sequence:

  - *Pepa es la consejera universitaria...*
Timed oral production test: Immediate posttest (3/15)

**GUILLERMO**

*I define myself as a conservative.*

*SER liberal EN materia social.*

(I am liberal with social matters)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Guillermo se define como un político conservador...

**BALTASAR**

*I am a judge.*

*SOLUCIONAR conflictos ENTRE las personas.*

(I solve conflicts between people)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Baltasar es un juez...
Timed oral production test: Immediate posttest (4/15)

**MARCELO**

*Necesito nuevos amigos.*  
(I need new friends)  
*BAILAR salsa conmigo.*  
(They dance salsa with me)

- Use the relative pronoun **QUE**  
- Use all the **RED** words.  
- Start with the following sequence:
  
  - Marcelo quiere nuevos amigos...

**KEVIN**

*Soy un estudiante de intercambio.*  
(I'm an exchange student)  
*ADORAR estudiar EN España.*  
(I love studying in Spain)

- Use the relative pronoun **QUE**  
- Use all the **RED** words.  
- Start with the following sequence:
  
  - Kevin es un estudiante de intercambio...
Timed oral production test: Immediate posttest (5/15)

**ANA**

_Necesito encontrar una persona._  
(I need to find a person)  
COMPRAR mi casa.  
(She buys my house)

- Use the relative pronoun **QUE**  
- Use all the **RED** words.  
- Start with the following sequence:

  - Ana necesita encontrar una persona...

**CONSUELO**

_Soy empleada doméstica._  
(I'm a housekeeper)  
_A veces ROBAR dinero DE las casas._  
(Sometimes I steal money from the houses)

- Use the relative pronoun **QUE**  
- Use all the **RED** words.  
- Start with the following sequence:

  - Consuelo es una empleada doméstica...
Timed oral production test: Immediate posttest (6/15)

**LOLA**

*Quiero conocer a un hombre.*
(I want to meet a man)

*ENAMORARSE de mi.*
(He falls in love with me)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Lola quiere conocer a un hombre...

**EDUARD**

*Quiero una sociedad más educada.*
(I want a better-educated society)

*LEER más libros.*
(It reads more books)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Eduard quiere una sociedad más educada...
Timed oral production test: Immediate posttest (7/15)

**Una PUERTA**

*En la puerta hay un agujero.*  
(There is a peephole in the door)

*Tú PODER VER el otro lado POR el agujero.*  
(You can see the other side through the peephole)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - *En la puerta hay un agujero...*

**JUAN**

*CANTAR sólo PARA una persona.*  
(I sing only for one person)

*Mi mujer.*  
(My wife)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - *Su mujer es la única persona...*
Timed oral production test: Immediate posttest (8/15)

LAURA

Quiero encontrar un compañero de apartamento.
(I want to find a roommate)

QUERER LIMPIAR la casa frecuentemente.
(S/he wants to clean the house often)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  Laura quiere encontrar un compañero...

LUCÍA

Busco un compañero de apartamento.
(I'm looking for a roommate)

TENER más de 25 años.
(S/he is older than 25)

- Use the relative pronoun QUE.
- Use all the RED words.
- Start with the following sequence:

  Laura busca un compañero de apartamento...
Timed oral production test: Immediate posttest (9/15)

CARMEN

Me gustan los hombres moderados.
(I like mellow men)
Sólo BEBER alcohol CON las comidas.
(They only drink alcohol with their meals)

- Use the relative pronoun QUE
- Use all the RED words.
- Start with the following sequence:

  A Carmen le gustan los hombres moderados...

FANGORIA

Somos un grupo.
(We're a band)
HACER música PARA bailar.
(We make dance music)

- Use the relative pronoun QUE
- Use all the RED words.
- Start with the following sequence:

  Fangoria es un grupo...
Timed oral production test: Immediate posttest (10/15)

**Óscar**

*Quiero encontrar un apartamento.*  
(I want to find an apartment)  
*ESTAR equipado con aire acondicionado.*  
(It is equipped with air conditioning)

- Use the relative pronoun **QUE**.  
- Use all the **RED** words.  
- Start with the following sequence:

  * Óscar quiere encontrar un apartamento...

**Pilar**

*Vivo en un apartamento barato.*  
(I live in a cheap apartment)  
*No TENER piscina EN el edificio.*  
(It doesn’t have a pool in the building)

- Use the relative pronoun **QUE**.  
- Use all the **RED** words.  
- Start with the following sequence:

  * Pilar vive en un apartamento barato...
Timed oral production test: Immediate posttest (11/15)

**Juan**

*Busco un compañero de apartamento.*
(I'm looking for a roommate)

*SÉR chico.*
(He is a guy)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Juan busca un compañero de apartamento...

**Bebe**

*Quiero componer una canción.*
(I want to write a song)

*CambiAR el mundo.*
(It changes the world)

- Use the relative pronoun **QUE**.
- Use all the **RED** words.
- Start with the following sequence:

  - Bebe quiere componer una canción...
Timed oral production test: Immediate posttest (12/15)

Fermin

No CORRER nunca SIN una cosa.
(I never run without one thing)

Mis zapatillas de la suerte.
(My lucky shoes)

- Use the relative pronoun QUE
- Use all the RED words.
- Start with the following sequence:

  • Sus zapatillas de la suerte son la cosa...

CARLOS

Mi teléfono OPERAR CON una compañía telefónica.
(My phone plan is with a phone company)

La compañía es Verizon.
(The company is Verizon)

- Use the relative pronoun QUE
- Use all the RED words.
- Start with the following sequence:

  • Verizon es la compañía...
Timed oral production test: Immediate posttest (13/15)

**ISABEL**

_Soy escritora._
(I'm a writer)

_ESCRIBIR libros SOBRE las mujeres._
(I write books about women)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Isabel es una escritora...

**HÉCTOR**

_ENSEÑAR EN una universidad._
(I teach at a University)

_Georgetown._
(Georgetown)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Georgetown es la universidad...
Timed oral production test: Immediate posttest (14/15)

**LUCAS**

Ayer **PONER** mi dinero **EN** un libro.
(Yesterday I put my money in a book)

*Hoy el dinero no está ahí.*
(Today the money is gone)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:

  * Hoy mi dinero no está en el libro...

**ANTONIO**

**Soy actor.**
(I'm an actor)

**ESTAR casado CON** Melanie Griffith.
(I'm married to Melanie Griffith)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:

  * Antonio es un actor...
Timed oral production test: Immediate posttest (15/15)

PABLO

**GUARDAR mi dinero EN un lugar secreto.**
(I hide my money in a secret place)

* Nadie conoce ese lugar.
(Nobody knows this place)*

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - Nadie conoce el lugar secreto...

Speak up now! (7 seconds)

---

**iPHONE**

**Soy un aparato.**
(I'm a gadget)

*Todos HABLAR DE mi.*
(Everybody talks about me)

- Use the relative pronoun **QUE**
- Use all the **RED** words.
- Start with the following sequence:
  - El iPhone es el aparato...

Speak up now! (7 seconds)
APPENDIX 5.4

Untimed grammaticality judgment test

Instructions: Next you will see 20 English sentences and their translations into Spanish. Decide whether the Spanish translations are grammatically correct or incorrect. If you don't know the answer, select I don't know. Focus specifically on the CAPITALIZED words.

PLEASE NOTE: You don't have to judge whether the word choice is correct (e.g., "ser" vs. "estar"). Rather, what we ask you to do here is to determine whether the sentence is grammatical or ungrammatical. For example, in English "I don't do nothing" is ungrammatical, because you shouldn't use double negation.

1. ENGLISH: Alex works for an organization that helps political refugees.
   SPANISH: Alex trabaja para una organización que ayuda a refugiados políticos.

2. ENGLISH: Antonio bought a car that runs on electricity.
   SPANISH: Antonio ha comprado un coche que funciona con electricidad.

3. ENGLISH: Carmen wants to meet a man that loves her.
   SPANISH: Carmen desea conocer a un hombre que la quiere.

4. ENGLISH: Coffee is something that I cannot live without.
   SPANISH: El café es una cosa que no puedo vivir sin.

5. ENGLISH: The professor will award the student who gets the highest grade at the end of the course.
   SPANISH: El profesor premiará al estudiante que al final del curso obtiene la nota más alta.

6. ENGLISH: Kasey lives in an apartment that has outside views.
   SPANISH: Kasey vive en un apartamento que tiene vistas al exterior.

7. ENGLISH: The house I live in doesn’t have heat.
   SPANISH: La casa que vivo en no tiene calefacción.
8. **ENGLISH:** The business my dad works for is Coca-Cola.
**SPANISH:** La empresa que mi padre trabaja PARA es Coca-Cola.

9. **ENGLISH:** Laura has a car that doesn’t use much gas on long trips.
**SPANISH:** Laura tiene un coche que consume poca gasolina EN los viajes largos.

10. **ENGLISH:** Lola has a brother who is a doctor.
**SPANISH:** Lola tiene un hermano que ES médico.

11. **ENGLISH:** Lucía lives in a house that I can see from mine.
**SPANISH:** Lucía vive en una casa que puedo ver DESDE la mía.

12. **ENGLISH:** María is the mother of a girl who is under the influence of drugs.
**SPANISH:** María es madre de una hija que vive BAJO los efectos de las drogas.

13. **ENGLISH:** María needs to buy a phone that has access to the internet.
**SPANISH:** María necesita comprar un teléfono que TIENE acceso a Internet.

14. **ENGLISH:** My dad is the person who I’m most proud of.
**SPANISH:** Mi padre es la persona que estoy más orgulloso DE.

15. **ENGLISH:** Mónica wants to find a roommate who speaks Spanish.
**SPANISH:** Mónica quiere encontrar un compañero de apartamento que HABLA español.

16. **ENGLISH:** I don’t like people who smoke in the workplace.
**SPANISH:** No me gustan las personas que fuman EN las oficinas.

17. **ENGLISH:** I can’t imagine the rough times that you are going through.
**SPANISH:** No puedo imaginar los momentos de dolor que tú estás pasando POR.

18. **ENGLISH:** Obama has two daughters who study a lot.
**SPANISH:** Obama tiene dos hijas que ESTUDIAN mucho.

19. **ENGLISH:** I want to see the last movie that Penélope Cruz starred in.
**SPANISH:** Quiero ver la última película que INTERPRETA Penélope Cruz.
20.  ENGLISH: I want to see a movie that is fun. Any suggestions?
SPANISH: Quiero ver una película que ES divertida. ¿Sugerencias?
APPENDIX 5.5

Preliminary morphology review session (1/9)

- In the following slides we will review how to conjugate the different tenses that you know.
- Write your answers on the sheet provided and then check them with the solutions in this powerpoint presentation.
- If you're wrong, write the correct answer next to your answer. Your sheet will be collected at the end of this presentation.

- Review of Spanish tenses

- Preterit

  - I'm assuming you already know the preterit, so just write down the requested forms for the following verbs.
  - Yo comer → yo comí.
  - Yo bailar → yo...
  - Él salir → él...
  - Yo cantar → yo...

- Ok, now click to check your answers:

  - Yo bailar → yo bailé
  - Él salir → él salió
  - Yo cantar → yo cantó
Preliminary morphology review session (2/9)

- How did it go? Feel free to go back by using the Page Up keys to review this information.

Imperfect

- Ok, now let's review the imperfect. Write the forms for the following verbs:
  - Yo cantar:
  - Leer:
  - Yo reír:
  - Yo caminar:

- Now click to check your answers!

- Yo cantaba
- Leía
- Yo reía
- Yo caminaba

- How did it go? Feel free to backtrack by using the Page Up keys to review this information.
Present subjunctive

- There are 3 types of verbs, according to their ending:
  - -AR
  - -ER
  - -IR

- When conjugating the present subjunctive, these verbs change the vowel of their ending in the following way:
  - -AR \rightarrow e
  - -ER, -IR \rightarrow a

Let’s see some examples:

- AR \rightarrow e
  - acabar \rightarrow que yo acabe
  - adorar \rightarrow que yo adores
  - cocinar \rightarrow que yo cocine

- Ok, now it's your turn. Write down the "Yo" form of the following verbs in the subjunctive. Then click to proceed.
  - Estar
  - Estudiar
  - Fumar
Preliminary morphology review session (4/9)

- Ok, click again to check your answers.

- AR → e
  - Estar → que yo esté
  - Estudiar → que yo estudie
  - Fumar → que yo fume

- Excellent, now let's see some examples of verbs ending in -ER, -IR. Do you remember the vowel they change to?

- That's right, -ER, -IR verbs change into a.
  For example:
  - Beber → que yo beba
  - Vivir → que yo viva

- So now it's your turn. Write down your answers and then click your mouse.

  - Comprender
  - Permitir
  - Leer
  - Partir

- Ok, click to check if you are right!
Preliminary morphology review session (5/9)

- Comprender  → que yo comprenda.
- Permitir    → que yo permita.
- Leer       → que yo lea.
- Partir      → que yo parte.

- Got them right?
  - If no, please use the Page Up key to review this information again.
  - If you got your answers right, click your mouse to proceed.

- Ok, you know that this is not the whole story. Spanish can get tricky because it has some exceptions!
- However, if you know the present indicative well, it will be very easy for you to learn the forms of the present subjunctive.

- For example, remember that some verbs in the present indicative are stem changing. That is, they change the vowel of their stem (not their ending). E.g.:
  - COSTAR  → Este es un poco.
  - QUERER  → Yo no quiero mucho.

- Did you remember these forms in the present INDICATIVE?

- Alright. Well, the same applies to the subjunctive. So think of the “él” forms of the following verbs, write down your answers and then proceed.
  - COSTAR  → que esta casa...
  - QUERER  → que él...
Preliminary morphology review session (6/9)

- Click to check your answers!

- COSTAR  → que esta casa CUESTE
- QUERER  → que él QUIERA

- Got them right?
  - If you have any questions, please ask Luis or the research assistant.

- Ok, another exception is with verbs ending in -CAR, -GAR. When forming the subjunctive with these verbs, you have to make some changes to preserve the SOUND of the verb.
  - Think of how you would form the "Yo" forms of the following verbs and then check your answers:
    - Practicar  → que yo...
    - Pagar     → que yo...

- Click to check your answers!

- Practicar  → que yo practique
- Pagar      → que yo pague
Preliminary morphology review session (7/9)

- Finally, all verbs that are irregular in the present indicative are also irregular in the present subjunctive.
- Let's go down memory lane and think of the "Yo" forms of the following verbs in the present INDICATIVE:
  - Hacer → HOY yo...
  - Ser → HOY yo...
  - Tener → HOY yo...
  - Venir → HOY yo...

- Now click to check your answers!

- So now think of the "Yo" forms in the present SUBJUNCTIVE!
  - Hacer → que yo...
  - Ser → que yo...
  - Tener → que yo...
  - Venir → que yo...

- Now click to check your answers!

- Hacer → que yo haga
  - Ser → que yo sea
  - Tener → que yo tenga
  - Venir → que yo venga
Preliminary morphology review session (8/9)

- Got it?
- Ok, to finish, let's do a mini-test. Write down the requested forms of the verbs in the following slide. Then check your answers.

1. Yo acabar (presente indicativo)
2. Yo fumar (pretérito)
3. Él beber (presente subjuntivo)
4. Yo vivir (presente indicativo)
5. Él hablar (imperfecto)
6. Él gastar (presente subjuntivo)
7. Yo llevar (pretérito)
8. Él proporcionar (presente subjuntivo)
9. Yo querer (presente subjuntivo)
10. Yo hacer (presente indicativo)
11. Él ser (presente subjuntivo)
12. Yo pagar (pretérito subjuntivo)
13. Él estar (presente indicativo)
14. Yo comprender (presente subjuntivo)
15. Él practicar (presente subjuntivo)
Preliminary morphology review session (9/9)

- Ok, now click to check your answers! 

- 1. Yo acabar (presente indicativo) ACABO
- 2. Yo fumar (pretérito) FUMÉ
- 3. Él beber (presente subjuntivo) BEBA
- 4. Yo vivir (presente indicativo) VIVO
- 5. Él hablar (imperfecto) HABLABA
- 6. Él gastar (presente subjuntivo) CASTE
- 7. Yo llevar (pretérito) LLEVÉ
- 8. Él proporcionar (presente subjuntivo) PROPORCIONE
- 9. Yo querer (presente subjuntivo) QUIERA
- 10. Yo hacer (presente indicativo) HAGO
- 11. Él ser (presente subjuntivo) SEA
- 12. Yo pagar (presente subjuntivo) PAGUE
- 13. Él estar (presente indicativo) ESTÁ
- 14. Yo comprender (presente subjuntivo) COMPRENDA
- 15. Él practicar (presente subjuntivo) PRACTIQUE

- You're done!
- Call Luis or the research assistant to complete the second task.
APPENDIX 5.6

Exit questionnaire (1/8)

**Question 1**

Your input is crucial to interpret the results of this study. The following is a brief survey on your perceptions about “Talking to Avatars.” Please fill in the following fields as accurately as possible. All information is anonymous and confidential, so please speak your mind. If you agree with this, select “I Agree” below.

- Agree
- Disagree

Moving to the next question prevents changes to this answer.

**Question 2**

Now that you have completed the experiment, can you identify its purpose? Yes! No? Please explain.

**Question 3**

The author of this experiment was trying to teach you some Spanish. Can you identify which grammatical structures or vocabulary items he was trying to teach you? Please be as specific as possible.

Moving to the next question prevents changes to this answer.
Exit questionnaire (2/8)

Question 4

Did you review/go over any of these structures between the first and last day of this experiment? Please be honest, no matter what you did. It is important to interpret whether any potential learning was exclusively due to the software used in this experiment or was aided by other factors. So again, the question is:

Did you review/go over any of these structures outside this laboratory setting?

- Yes
- No

Moving to the next question prevents changes to this answer.

Question 5

If you checked “yes” in the previous question, please select all the options that apply. Otherwise select N/A:

- I consulted the class or other grammar textbook on these structures.
- I asked one of my Spanish professors about these structures.
- I asked a classmate about these structures.
- I asked a Spanish speaker about these structures.
- My teacher discussed these structures in class.
- Other
- N/A

Moving to the next question prevents changes to this answer.

Question 6

If you selected “Other” earlier, please specify how you learned about these structures. Otherwise please type N/A.

Moving to the next question prevents changes to this answer.

Question 7

If you did review/go over any of the structures taught in this experiment, when did you do that?

- After the first session
- After the second session

Moving to the next question prevents changes to this answer.
Exit questionnaire (3/8)

**Question 8**

What is your overall evaluation of "Talking to Avatars," the piece of software used in this study?

- 1. I hated it
- 2. I disliked it
- 3. Not sure
- 4. I liked it
- 5. I loved it.

⚠️ Moving to the next question prevents changes to this answer.

**Question 9**

What in particular did you like and didn’t like about this piece of software?

⚠️ Moving to the next question prevents changes to this answer.

**Question 10**

If you had used any type of commercial software for language learning purposes before this experiment, please state the name of that piece of software and whether you liked it better or worse than Talking to Avatars, explaining why.

⚠️ Moving to the next question prevents changes to this answer.
Exit questionnaire (4/8)

**Question 11**

Please rate the following statement. The program is easy to use. I understood the instructions and knew what to do at every point.
1. (1) It was extremely difficult to use
2. (2) It was difficult to use
3. (3) It was neither difficult nor easy to use
4. (4) It was easy to use
5. (5) It was extremely easy to use.

⚠ Moving to the next question prevents changes to this answer.

**Question 12**

I didn't have any technical problems while interacting with the program (e.g., the videos froze, took too much time to download, the sound was bad, etc.).
1. (1) The software had lots of problems
2. (2) The software had some problems
3. (3) I don't have an opinion
4. (4) Everything ran smoothly
5. (5) Everything ran superibly

⚠ Moving to the next question prevents changes to this answer.

**Question 13**

If you encountered technical problems, please specify them below. Otherwise, please write N/A:

⚠ Moving to the next question prevents changes to this answer.
Exit questionnaire (5/8)

**Question 14**

Do you think you learned some Spanish by taking part in this experiment? Please provide your answer on a scale of 1 to 5, where 1 is "I didn't learn anything at all" and 5 is "I learned a lot".

- 1
- 2
- 3
- 4
- 5

⚠️ Moving to the next question prevents changes to this answer.

**Question 15**

Exclude the activities in the tests and focus solely on the activities in the practice phase (i.e., finding accommodation, reporting a theft). Now, answer the following question: I found the number of practice sentences to be.

- 1. Too few for learning purposes
- 2. Slightly fewer than I needed
- 3. Far more than I needed
- 4. Slightly more than I needed
- 5. The amount was perfect.

⚠️ Moving to the next question prevents changes to this answer.

**Question 16**

I found the content of the activities (e.g., finding accommodation, reporting a theft) to be interesting and engaging.

- 1. They were extremely boring
- 2. They were boring
- 3. They were ok
- 4. They were pretty engaging
- 5. They were extremely engaging

⚠️ Moving to the next question prevents changes to this answer.

**Question 17**

I found the content of the activities (e.g., finding accommodation, reporting a theft) to be useful for real life in a Spanish-speaking country.

- 1. They were not useful at all
- 2. They were not very useful
- 3. They were ok
- 4. They were pretty useful
- 5. They were super-useful

⚠️ Moving to the next question prevents changes to this answer.
Exit questionnaire (6/8)

**Question 18**

While interacting with Talking to Avatars, some students received feedback (the avatars told them if their Spanish was ok or not) while others did not. If you received feedback, was it easy to understand? If you did not receive any feedback, please select N/A.

- (1) it was super-difficult to understand
- (2) it was difficult to understand
- (3) it was neither difficult nor easy
- (4) it was easy to understand
- (5) it was super-easy to understand
- (6) Not applicable

⚠ Moving to the next question prevents changes to this answer.

**Question 19**

If you received feedback, do you think it was useful for learning? If you did not receive feedback, select Not Applicable.

- (1) it wasn’t useful at all
- (2) it wasn’t very useful
- (3) it was ok
- (4) it was pretty useful
- (5) it was extremely useful.
- (6) Not applicable

⚠ Moving to the next question prevents changes to this answer.

**Question 20**

I received oral feedback, but I’d rather the avatar had given me a different type of content (more or less explanatory).

- (1) I was very unsatisfied with the level of detail
- (2) I was unsatisfied with the level of detail
- (3) it was ok
- (4) I liked the level of detail
- (5) I loved the level of detail.
- (6) Not applicable

⚠ Moving to the next question prevents changes to this answer.
Exit questionnaire (7/8)

Question 21
The instructor gave me feedback, but I'd rather s/he had provided me with written instead of oral feedback.
- (1) Absolutely, I hated the oral feedback here.
- (2) Yes, I think written feedback would have been better
- (3) I don't care.
- (4) No, I think oral feedback was better than having written feedback
- (5) No, I hated the oral feedback here.
- (6) Not applicable (I didn't get feedback)

⚠ Moving to the next question prevents changes to this answer.

Question 22
This software is good as a supplement to learn Spanish outside the classroom.
- (1) No, it wouldn't help me at all
- (2) No, I wouldn't help me much
- (3) Not sure
- (4) Yes, it would help me somewhat
- (5) Yes, it would help me a great deal

⚠ Moving to the next question prevents changes to this answer.

Question 23
I'd like to practice Spanish by using this type of software on a regular basis.
- (1) No, I'm totally against it.
- (2) No, I wouldn't like that much
- (3) Not sure
- (4) Yes, I'd like to
- (5) Totally, I'd love to.

⚠ Moving to the next question prevents changes to this answer.

Question 24
I feel more comfortable working with this software than in a regular classroom with a professor.
- (1) No, I'm way more comfortable in a regular classroom.
- (2) No, I'm slightly more comfortable in a regular classroom.
- (3) Not sure
- (4) Yes, I feel slightly more comfortable learning with the software.
- (5) Yes, I'm way more comfortable learning with the software.

⚠ Moving to the next question prevents changes to this answer.
Exit questionnaire (8/8)

**Question 25**
The "Transcribe" and "Translate" buttons were helpful.
- (1) They didn’t help me at all
- (2) They didn’t help me much
- (3) Not sure
- (4) They helped me somewhat
- (5) They helped me a lot

⚠️ Moving to the next question prevents changes to this answer.

**Question 26**
In this experiment some groups of students got to interact with the avatars, while the other groups of students got to watch these interactions without taking part in them. Think of the group you fell in, and rate the following statement: “I wish I had fallen in the opposite group.”
- (1) Totally, I hated my group
- (2) Somewhat, I didn’t like my group much
- (3) Not sure
- (4) No, I prefer the group I fell in
- (5) No, I absolutely loved my group.

⚠️ Moving to the next question prevents changes to this answer.

**Question 27**
In this experiment some students were asked to correct their mistakes in an extra trial, while others were simply asked to move on. If you received feedback, do you wish you had fallen in the opposite group (extra trial vs. no extra trial)? If you didn’t receive feedback at all, select “Not applicable.”
- (1) Totally, I hated my group
- (2) Somewhat, I didn’t like my group much
- (3) Not sure
- (4) No, I prefer the group I fell in
- (5) No, I loved my group.
- (6) Not applicable.

⚠️ Moving to the next question prevents changes to this answer.

**Question 28**
Thank you very much for completing the questionnaire! Please select “I agree” to finish and call Luis or the research assistant.
- Agree
- Disagree

⚠️ Click Submit to complete this assessment.

Save and Submit
APPENDIX 6.1

Written production accuracy for Prepositional relative clauses: Descriptive statistics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Feedback</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Spot &amp; Continue</td>
<td>16</td>
<td>.13</td>
<td>.34</td>
<td>3.19</td>
<td>4.07</td>
<td>2.81</td>
<td>4.13</td>
</tr>
<tr>
<td></td>
<td>Spot &amp; Prompt</td>
<td>16</td>
<td>.69</td>
<td>a,b</td>
<td>3.94</td>
<td>3.75</td>
<td>3.63</td>
<td>4.03</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Continue</td>
<td>17</td>
<td>.35</td>
<td>a,b</td>
<td>6.76</td>
<td>3.78</td>
<td>6.29</td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>17</td>
<td>.06</td>
<td>a,b</td>
<td>4.29</td>
<td>4.30</td>
<td>4.29</td>
<td>4.10</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>15</td>
<td>.13</td>
<td>.35</td>
<td>.27</td>
<td>1.03</td>
<td>.53</td>
<td>2.07</td>
</tr>
<tr>
<td>Exposure</td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>.00</td>
<td>.00</td>
<td>.64</td>
<td>1.45</td>
<td>1.14</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>13</td>
<td>.77</td>
<td>a,b</td>
<td>4.31</td>
<td>4.33</td>
<td>4.31</td>
<td>4.15</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>15</td>
<td>.27</td>
<td>.70</td>
<td>.40</td>
<td>.91</td>
<td>1.00</td>
<td>2.36</td>
</tr>
</tbody>
</table>

Superscripts indicate significance of change in accuracy mean scores for each individual group across time. Paired \( t \)-test values for intervals Time 1-2, Time 1-3, and Time 2-3, respectively, are:

- Practice SC, \( t(15) = 2.995; 2.595; -.686; p = .009**; .020*; .503. 
- Practice SP, \( t(15) = 3.828; 3.226; -.608; p = .002**; .006**; .552. 
- Practice EC, \( t(16) = 7.160; 5.997; -1.367; p = .000**; .000**; .191. 
- Practice EP, \( t(16) = 4.123; 4.307; .000; p = .001**; .001**; 1.00. 
- Practice NF, \( t(14) = .459; .727; 1.00; p = .653; .479; .334. 
- Exposure SP, \( t(13) = 1.662; 1.544; 1.102; p = .120; .147; .291. 
- Exposure EP, \( t(12) = 2.777; 2.822; .000; p = .017*; .015*; 1.00. 
- Exposure NF, \( t(14) = .619; 1.319; 1.598; p = .546; 208; 132. 

575
## APPENDIX 6.2

Oral production accuracy for Prepositional relative clauses: Descriptive statistics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Feedback</th>
<th>N</th>
<th>Pretest</th>
<th>Posttest (a)</th>
<th>Delayed (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Practice</td>
<td>Spot &amp; Continue</td>
<td>16</td>
<td>.00&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.00</td>
<td>1.25&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>.07&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>.27</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Continue</td>
<td>17</td>
<td>.00&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>.00</td>
<td>3.47</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>17</td>
<td>.12&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>.33</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>15</td>
<td>.13</td>
<td>.35</td>
<td>.00</td>
</tr>
<tr>
<td>Exposure</td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>.14</td>
<td>.36</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>13</td>
<td>.38&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.87</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>15</td>
<td>.07</td>
<td>.26</td>
<td>.20</td>
</tr>
</tbody>
</table>

Superscripts indicate significance of change in accuracy mean scores for each individual group across time. Paired t-test values for intervals Time 1-2, Time 1-3, and Time 2-3, respectively, are:

- Practice SC, \( t(15) = 1.927; 2.309; 2.300; p = .073; .036*; .036*. \)
- Practice SP, \( t(13) = 3.069; 2.844, 1.421; p = .009**; .014*; .179. \)
- Practice EC, \( t(16) = 4.150, 4.654; 1.912; p = .001**; .000**; .074 \)
- Practice EP, \( t(16) = 3.077; 3.050; .719; p = .007**; .008**; p = .483. \)
- Practice NF, \( t(14) = -1.468; -1.00; 1.00, p = .164; .334; .334. \)
- Exposure SP, \( t(13) = 1.421; 1.00; -1.099, p = .179; .336; .292. \)
- Exposure EP, \( t(12) = 2.091; 3.120; 1.915, p = .058; .009**; .080. \)
- Exposure NF, \( t(14) = .619; 1.451; 1.234; p = .546; .169; .238 \)

576
APPENDIX 6.3

Grammaticality judgment accuracy for Prepositional relative clauses: Descriptive statistics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Feedback</th>
<th>N</th>
<th>Pretest</th>
<th>Posttest (a)</th>
<th>Delayed (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Practice</td>
<td>Spot &amp; Continue</td>
<td>16</td>
<td>-.19\textsuperscript{a}</td>
<td>1.83</td>
<td>1.81</td>
</tr>
<tr>
<td></td>
<td>Spot &amp; Prompt</td>
<td>16</td>
<td>.13\textsuperscript{a}</td>
<td>1.67</td>
<td>1.88</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Continue</td>
<td>17</td>
<td>.18\textsuperscript{a,b}</td>
<td>1.47</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>17</td>
<td>-.29\textsuperscript{a,b}</td>
<td>1.57</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>15</td>
<td>-.27</td>
<td>2.02</td>
<td>-.87</td>
</tr>
<tr>
<td>Exposure</td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>-.21</td>
<td>1.37</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>13</td>
<td>.46\textsuperscript{a,b}</td>
<td>1.76</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>15</td>
<td>.27</td>
<td>1.53</td>
<td>-.13</td>
</tr>
</tbody>
</table>

Superscripts indicate significance of change in accuracy mean scores for each individual group across time. Paired \( t \)-test values for intervals Time 1-2, Time 1-3, and Time 2-3, respectively, are:

- Practice SC, \( t(15) = 2.530; 1.580; -1.346; p = .023*; .135; .198.  
- Practice SP, \( t(15) = 3.530; 1.587; -1.090; p = .003**; .133; .293.  
- Practice EC, \( t(16) = 4.110, 4.015; -.457; p = .001**; .001**; .653.  
- Practice EP, \( t(16) = 2.782; 3.636; .814; p = .013*; .002**; .428.  
- Practice NF, \( t(14) = -1.188; -.292; 1.388, p = .255; .774; .187.  
- Exposure SP, \( t(13) = .673; .123; -.718; p = .513; .904; .486.  
- Exposure EP, \( t(12) = 4.266; 4.538; -.341; p = .001**; .001**; .739.  
- Exposure NF, \( t(14) = -.751; -1.177; -.552; p = .465; .259; .589.  

577
APPENDIX 6.4

Written production accuracy for Present subjunctive: Descriptive statistics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Feedback</th>
<th>N</th>
<th>Pretest M</th>
<th>Pretest SD</th>
<th>Posttest M (a)</th>
<th>Posttest SD (a)</th>
<th>Delayed M (b)</th>
<th>Delayed SD (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Spot &amp; Continue</td>
<td>18</td>
<td>.39</td>
<td>1.09</td>
<td>1.56</td>
<td>2.53</td>
<td>1.28</td>
<td>2.95</td>
</tr>
<tr>
<td></td>
<td>Spot &amp; Prompt</td>
<td>17</td>
<td>.65&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>1.11</td>
<td>2.41</td>
<td>2.50</td>
<td>2.12</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Continue</td>
<td>17</td>
<td>.29&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>.69</td>
<td>4.24</td>
<td>3.15</td>
<td>3.65</td>
<td>3.53</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>17</td>
<td>.88&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.05</td>
<td>2.94</td>
<td>2.79</td>
<td>2.53</td>
<td>3.68</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>16</td>
<td>.00</td>
<td>.89</td>
<td>1.25</td>
<td>2.11</td>
<td>1.31</td>
<td>2.63</td>
</tr>
<tr>
<td>Exposure</td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>- .14&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.77</td>
<td>2.14</td>
<td>2.35</td>
<td>1.79</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>13</td>
<td>.23&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>1.01</td>
<td>3.23</td>
<td>3.49</td>
<td>3.69</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>14</td>
<td>.43</td>
<td>1.28</td>
<td>.86</td>
<td>1.91</td>
<td>.36</td>
<td>.84</td>
</tr>
</tbody>
</table>

Superscripts indicate significance of change in accuracy mean scores for each individual group across time. Paired t-test values for intervals Time 1-2, Time 1-3, and Time 2-3, respectively, are:

- Practice SC, t(17) = 1.760; 1.262; -.519; p = .096; .224; .610.
- Practice SP, t(16) = 3.156; 3.234; -.629; p = .006**; .005**; .538.
- Practice EC, t(16) = 5.156; 3.781; -.979; p = .000**; .002**; .342.
- Practice EP, t(16) = 3.038; 1.902; -.835; p = .008**; .075; .416.
- Practice NF, t(15) = 2.030; 1.801; .141; p = .060; .092; .889.
- Exposure SP, t(13) = 3.133; 2.038; -.489; p = .008**; .062; .633.
- Exposure EP, t(12) = 3.144; 3.143; 1.032; p = .008**; .008**; .323.
- Exposure NF, t(13) = 1.147; -.179; -.846; p = .272; .861; .413.
APPENDIX 6.5

Oral production accuracy for Present subjunctive: Descriptive statistics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Feedback</th>
<th>N</th>
<th>Pretest M</th>
<th>Pretest SD</th>
<th>Posttest M</th>
<th>Posttest SD</th>
<th>Delayed M</th>
<th>Delayed SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>Spot &amp; Continue</td>
<td>17</td>
<td>.59</td>
<td>1.42</td>
<td>1.35</td>
<td>2.89</td>
<td>1.06</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>Spot &amp; Prompt</td>
<td>15</td>
<td>.93(^a)</td>
<td>1.91</td>
<td>2.33</td>
<td>2.72</td>
<td>2.00</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Continue</td>
<td>17</td>
<td>.29(^a,b)</td>
<td>1.21</td>
<td>3.12</td>
<td>3.04</td>
<td>2.88</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>17</td>
<td>.53(^a,b)</td>
<td>1.12</td>
<td>2.00</td>
<td>2.37</td>
<td>1.88</td>
<td>2.39</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>16</td>
<td>.50</td>
<td>1.21</td>
<td>.44</td>
<td>1.90</td>
<td>1.25</td>
<td>2.32</td>
</tr>
<tr>
<td>Exposure</td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>.14(^a,b)</td>
<td>.53</td>
<td>1.79</td>
<td>2.78</td>
<td>1.86</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>13</td>
<td>.69(^b)</td>
<td>1.80</td>
<td>2.23(^b)</td>
<td>3.61</td>
<td>3.31</td>
<td>3.59</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>14</td>
<td>.00</td>
<td>1.66</td>
<td>.57</td>
<td>1.34</td>
<td>-.14</td>
<td>.66</td>
</tr>
</tbody>
</table>

Superscripts indicate significance of change in accuracy mean scores for each individual group across time. Paired t-test values for intervals Time 1-2, Time 1-3, and Time 2-3, respectively, are:

- Practice SC, \(t(16) = .912; 1.017; -.551; p = .375; .324; .589\).
- Practice SP, \(t(14) = 2.432; 2.049; -.892; p = .029*; .060; .388\).
- Practice EC, \(t(16) = 4.431; 3.895; -.368; p = .000**; .001**; .718\).
- Practice EP, \(t(16) = 2.742; 2.255; -.287; p = .014*; .038*; .778\).
- Practice NF, \(t(15) = -.180; 1.772; 1.500; p = .860; .097; .154\).
- Exposure SP, \(t(13) = 2.182; 2.326; .249; p = .048*; .037*; .807\).
- Exposure EP, \(t(12) = 1.606; 2.640; 2.344; p = .134; .022*; .037*\).
- Exposure NF, \(t(13) = 1.096; -.263; -1.587; p = .293; .797; .136\).

579
APPENDIX 6.6

Grammaticality judgment accuracy for Present subjunctive: Descriptive statistics

<table>
<thead>
<tr>
<th>Agency</th>
<th>Feedback</th>
<th>N</th>
<th>Pretest</th>
<th></th>
<th>Posttest</th>
<th></th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Practice</td>
<td>Spot &amp; Continue</td>
<td>18</td>
<td>.00</td>
<td>1.37</td>
<td>.39</td>
<td>.98</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>Spot &amp; Prompt</td>
<td>17</td>
<td>.06\text{a,b}</td>
<td>.97</td>
<td>.88</td>
<td>1.45</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Continue</td>
<td>17</td>
<td>.47\text{a}</td>
<td>.94</td>
<td>1.65</td>
<td>1.80</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>17</td>
<td>.59</td>
<td>1.06</td>
<td>1.12</td>
<td>1.69</td>
<td>.94</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>16</td>
<td>-.19</td>
<td>1.22</td>
<td>.19</td>
<td>1.11</td>
<td>.13</td>
</tr>
<tr>
<td>Exposure</td>
<td>Spot &amp; Prompt</td>
<td>14</td>
<td>-.14</td>
<td>.66</td>
<td>.64</td>
<td>1.34</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>Explain &amp; Prompt</td>
<td>13</td>
<td>.77</td>
<td>1.83</td>
<td>2.08</td>
<td>2.60</td>
<td>2.15</td>
</tr>
<tr>
<td></td>
<td>No feedback</td>
<td>14</td>
<td>-.36</td>
<td>.93</td>
<td>.00</td>
<td>1.11</td>
<td>.14</td>
</tr>
</tbody>
</table>

Superscripts indicate significance of change in accuracy mean scores for each individual group across time. Paired *t*-test values for intervals Time 1-2, Time 1-3, and Time 2-3, respectively, are:

- Practice SC, *t*(17) = 1.046; .531; -.416; *p* = .310; .602; .682.
- Practice SP, *t*(16) = 3.160; 2.252; .152; *p* = .006**; .039*; .881.
- Practice EC, *t*(16) = 2.787; 1.844; -.410; *p* = .013*; .084; .687.
- Practice EP, *t*(16) = 1.232; .643; -.469; *p* = .236; .529; .645.
- Practice NF, *t*(15) = 1.307; .639; -.151; *p* = .211; .533; .882.
- Exposure SP, *t*(13) = 1.924; 1.035; -.392; *p* = .077; .319; .701.
- Exposure EP, *t*(12) = 1.596; 1.656; .201; *p* = .137; .124; .844.
- Exposure NF, *t*(13) = .960; 1.047; .342; *p* = .355; .314; .738.
APPENDIX 6.7

RQ1 - Summary of ANOVA results for accuracy in Prepositional relative clauses for Spot & Prompt treatment: Time by Agency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>79.243</td>
<td>39.622</td>
<td>11.098</td>
<td>.000**</td>
<td>.284</td>
<td>.989</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>104.003</td>
<td>104.003</td>
<td>7.333</td>
<td>.011*</td>
<td>.208</td>
<td>.743</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>26.577</td>
<td>13.288</td>
<td>3.722</td>
<td>.030*</td>
<td>.117</td>
<td>.659</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>20.881</td>
<td>10.440</td>
<td>7.545</td>
<td>.001**</td>
<td>.225</td>
<td>.932</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>11.167</td>
<td>5.583</td>
<td>4.035</td>
<td>.023*</td>
<td>.134</td>
<td>.695</td>
</tr>
<tr>
<td><strong>Grammaticality judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>16.786</td>
<td>8.393</td>
<td>3.539</td>
<td>.036*</td>
<td>.112</td>
<td>.636</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>30.957</td>
<td>30.957</td>
<td>5.781</td>
<td>.023*</td>
<td>.171</td>
<td>.641</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>8.119</td>
<td>4.060</td>
<td>1.712</td>
<td>.190</td>
<td>.058</td>
<td>.345</td>
</tr>
</tbody>
</table>
APPENDIX 6.8

RQ1 - Summary of ANOVA results for accuracy in Prepositional relative clauses for Explain & Prompt treatment: Time by Agency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>296.785</td>
<td>148.392</td>
<td>21.225</td>
<td>.000**</td>
<td>.431</td>
<td>1.000</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>1.336</td>
<td>1.336</td>
<td>.060</td>
<td>.808</td>
<td>.002</td>
<td>.056</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>2.385</td>
<td>1.192</td>
<td>.171</td>
<td>.844</td>
<td>.006</td>
<td>.075</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>100.200</td>
<td>50.100</td>
<td>14.301</td>
<td>.000**</td>
<td>.338</td>
<td>.998</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>4.027</td>
<td>4.027</td>
<td>.353</td>
<td>.557</td>
<td>.012</td>
<td>.089</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>4.066</td>
<td>2.033</td>
<td>.580</td>
<td>.563</td>
<td>.020</td>
<td>.142</td>
</tr>
<tr>
<td><strong>Grammatical judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>107.060</td>
<td>53.530</td>
<td>20.994</td>
<td>.000**</td>
<td>.429</td>
<td>1.000</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>18.954</td>
<td>18.954</td>
<td>3.970</td>
<td>.056</td>
<td>.124</td>
<td>.486</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>1.060</td>
<td>.530</td>
<td>.208</td>
<td>.813</td>
<td>.007</td>
<td>.081</td>
</tr>
</tbody>
</table>
APPENDIX 6.9

RQ1 - Summary of ANOVA results for accuracy in Prepositional relative clauses for No Feedback treatment: Time by Agency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>5.267</td>
<td>2.633</td>
<td>2.206</td>
<td>.120</td>
<td>.073</td>
<td>.432</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>1.344</td>
<td>1.344</td>
<td>.355</td>
<td>.556</td>
<td>.013</td>
<td>.089</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>.556</td>
<td>.278</td>
<td>.233</td>
<td>.793</td>
<td>.008</td>
<td>.085</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>.800</td>
<td>.400</td>
<td>1.333</td>
<td>.272</td>
<td>.045</td>
<td>.276</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>.900</td>
<td>.900</td>
<td>1.099</td>
<td>.303</td>
<td>.038</td>
<td>.173</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>1.067</td>
<td>.533</td>
<td>1.778</td>
<td>.178</td>
<td>.060</td>
<td>.357</td>
</tr>
<tr>
<td><strong>Grammaticality judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>4.200</td>
<td>2.100</td>
<td>1.188</td>
<td>.312</td>
<td>.041</td>
<td>.250</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>4.011</td>
<td>4.011</td>
<td>.569</td>
<td>.457</td>
<td>.020</td>
<td>.113</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>2.156</td>
<td>1.078</td>
<td>.610</td>
<td>.547</td>
<td>.021</td>
<td>.147</td>
</tr>
</tbody>
</table>
APPENDIX 6.10

RQ1 - Summary of ANOVA results for accuracy in Present subjunctive for Spot & Prompt treatment: Time by Agency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>72.639</td>
<td>36.320</td>
<td>11.329</td>
<td>.000**</td>
<td>.281</td>
<td>.990</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>4.950</td>
<td>4.950</td>
<td>.642</td>
<td>.430</td>
<td>.022</td>
<td>.121</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>1.242</td>
<td>.621</td>
<td>.194</td>
<td>.824</td>
<td>.007</td>
<td>.079</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>41.182</td>
<td>20.591</td>
<td>8.946</td>
<td>.000**</td>
<td>.249</td>
<td>.966</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>5.294</td>
<td>5.294</td>
<td>.404</td>
<td>.530</td>
<td>.015</td>
<td>.094</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>1.550</td>
<td>.775</td>
<td>.337</td>
<td>.716</td>
<td>.012</td>
<td>.101</td>
</tr>
<tr>
<td><strong>Grammaticality judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>12.098</td>
<td>6.049</td>
<td>4.361</td>
<td>.017*</td>
<td>.131</td>
<td>.733</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>2.328</td>
<td>2.328</td>
<td>.665</td>
<td>.422</td>
<td>.022</td>
<td>.124</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>.442</td>
<td>.221</td>
<td>.159</td>
<td>.853</td>
<td>.005</td>
<td>.073</td>
</tr>
</tbody>
</table>
APPENDIX 6.11

RQ1 - Summary of ANOVA results for accuracy in Present subjunctive for Explain & Prompt treatment: Time by Agency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>126.932</td>
<td>63.466</td>
<td>14.077</td>
<td>.000**</td>
<td>.335</td>
<td>.998</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>1.575</td>
<td>1.575</td>
<td>.094</td>
<td>.761</td>
<td>.003</td>
<td>.060</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>12.132</td>
<td>6.066</td>
<td>1.346</td>
<td>.269</td>
<td>.046</td>
<td>.278</td>
</tr>
<tr>
<td>Oral production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>63.162</td>
<td>31.581</td>
<td>9.627</td>
<td>.000**</td>
<td>.256</td>
<td>.976</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>8.125</td>
<td>8.125</td>
<td>.618</td>
<td>.438</td>
<td>.022</td>
<td>.118</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>7.429</td>
<td>3.714</td>
<td>1.132</td>
<td>.330</td>
<td>.039</td>
<td>.240</td>
</tr>
<tr>
<td>Grammaticality judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>15.725</td>
<td>7.863</td>
<td>3.243</td>
<td>.047*</td>
<td>.104</td>
<td>.595</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>13.595</td>
<td>13.595</td>
<td>1.885</td>
<td>.181</td>
<td>.063</td>
<td>.264</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>4.259</td>
<td>2.129</td>
<td>.878</td>
<td>.421</td>
<td>.030</td>
<td>.194</td>
</tr>
</tbody>
</table>
APPENDIX 6.12

RQ1 - Summary of ANOVA results for accuracy in Present subjunctive for No Feedback treatment: Time by Agency

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>11.323</td>
<td>5.661</td>
<td>2.482</td>
<td>.093</td>
<td>.081</td>
<td>.478</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>2.105</td>
<td>2.105</td>
<td>.443</td>
<td>.511</td>
<td>.016</td>
<td>.099</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>7.234</td>
<td>3.617</td>
<td>1.586</td>
<td>.214</td>
<td>.054</td>
<td>.322</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>1.586</td>
<td>.793</td>
<td>.472</td>
<td>.626</td>
<td>.017</td>
<td>.123</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>7.700</td>
<td>7.700</td>
<td>1.685</td>
<td>.205</td>
<td>.057</td>
<td>.241</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>8.786</td>
<td>4.393</td>
<td>2.614</td>
<td>.082</td>
<td>.085</td>
<td>.500</td>
</tr>
<tr>
<td><strong>Grammaticality judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>2.993</td>
<td>1.497</td>
<td>1.161</td>
<td>.321</td>
<td>.040</td>
<td>.245</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>.287</td>
<td>.287</td>
<td>.127</td>
<td>.725</td>
<td>.005</td>
<td>.064</td>
</tr>
<tr>
<td>Time * Agency</td>
<td>2</td>
<td>.193</td>
<td>.097</td>
<td>.075</td>
<td>.928</td>
<td>.003</td>
<td>.061</td>
</tr>
</tbody>
</table>
APPENDIX 6.13

RQ2 - Summary of ANOVA results for accuracy in both Forms combined for the 5 Practice groups: Time by Feedback

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>746.059</td>
<td>373.030</td>
<td>70.607</td>
<td>.000</td>
<td>.482</td>
<td>1.000</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>462.324</td>
<td>115.581</td>
<td>6.010</td>
<td>.000</td>
<td>.240</td>
<td>.981</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>204.561</td>
<td>25.570</td>
<td>4.840</td>
<td>.000</td>
<td>.203</td>
<td>.998</td>
</tr>
<tr>
<td>Oral production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>246.797</td>
<td>123.399</td>
<td>37.757</td>
<td>.000</td>
<td>.338</td>
<td>1.000</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>206.525</td>
<td>51.631</td>
<td>3.775</td>
<td>.008</td>
<td>.169</td>
<td>.871</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>122.471</td>
<td>15.309</td>
<td>4.684</td>
<td>.000</td>
<td>.202</td>
<td>.997</td>
</tr>
<tr>
<td>Grammaticality judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>131.982</td>
<td>65.991</td>
<td>23.275</td>
<td>.000</td>
<td>.234</td>
<td>1.000</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>187.277</td>
<td>46.819</td>
<td>7.883</td>
<td>.000</td>
<td>.293</td>
<td>.997</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>42.759</td>
<td>5.345</td>
<td>1.885</td>
<td>.066</td>
<td>.090</td>
<td>.777</td>
</tr>
</tbody>
</table>
APPENDIX 6.14

RQ2 - Summary of ANOVA results for accuracy in Prepositional relative clauses for the 5 Practice groups: Time by Feedback

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>598.582</td>
<td>299.291</td>
<td>60.733</td>
<td>.000**</td>
<td>.444</td>
<td>1.00</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>433.645</td>
<td>108.411</td>
<td>5.891</td>
<td>.000**</td>
<td>.237</td>
<td>.978</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>200.315</td>
<td>25.039</td>
<td>5.081</td>
<td>.000**</td>
<td>.211</td>
<td>.999</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>183.608</td>
<td>91.804</td>
<td>33.636</td>
<td>.000**</td>
<td>.312</td>
<td>1.00</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>157.523</td>
<td>39.381</td>
<td>4.291</td>
<td>.004**</td>
<td>.188</td>
<td>.914</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>89.001</td>
<td>11.125</td>
<td>4.076</td>
<td>.000**</td>
<td>.181</td>
<td>.991</td>
</tr>
<tr>
<td><strong>Grammaticality judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>131.367</td>
<td>65.684</td>
<td>22.055</td>
<td>.000**</td>
<td>.225</td>
<td>1.00</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>160.065</td>
<td>40.016</td>
<td>6.999</td>
<td>.000**</td>
<td>.269</td>
<td>.992</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>66.935</td>
<td>8.367</td>
<td>2.809</td>
<td>.006**</td>
<td>.129</td>
<td>.935</td>
</tr>
</tbody>
</table>
APPENDIX 6.15

RQ2 - Summary of ANOVA results for accuracy in Present subjunctive for the 5 Practice groups: Time by Feedback

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>205.009</td>
<td>102.505</td>
<td>28.841</td>
<td>.000**</td>
<td>.265</td>
<td>1.000</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>118.058</td>
<td>29.515</td>
<td>2.882</td>
<td>.028*</td>
<td>.126</td>
<td>.756</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>50.583</td>
<td>6.323</td>
<td>1.779</td>
<td>.085</td>
<td>.082</td>
<td>.749</td>
</tr>
<tr>
<td>Oral production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>86.960</td>
<td>43.480</td>
<td>16.900</td>
<td>.000**</td>
<td>.180</td>
<td>1.000</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>60.527</td>
<td>15.132</td>
<td>1.437</td>
<td>.230</td>
<td>.069</td>
<td>.427</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>44.964</td>
<td>5.621</td>
<td>2.185</td>
<td>.031*</td>
<td>.102</td>
<td>.846</td>
</tr>
<tr>
<td>Grammaticality judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>21.558</td>
<td>10.779</td>
<td>7.221</td>
<td>.001**</td>
<td>.083</td>
<td>.931</td>
</tr>
<tr>
<td>Feedback</td>
<td>4</td>
<td>44.891</td>
<td>11.223</td>
<td>2.855</td>
<td>.029*</td>
<td>.125</td>
<td>.752</td>
</tr>
<tr>
<td>Time * Feedback</td>
<td>8</td>
<td>5.583</td>
<td>.698</td>
<td>.468</td>
<td>.878</td>
<td>.023</td>
<td>.213</td>
</tr>
</tbody>
</table>
APPENDIX 6.16

RQ3 - Summary of results of 3x2 repeated-measures ANOVA for all 8 experimental groups: Time by Form

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>892.754</td>
<td>446.377</td>
<td>96.260</td>
<td>.000**</td>
<td>.280</td>
<td>1.000</td>
</tr>
<tr>
<td>Form</td>
<td>1</td>
<td>56.154</td>
<td>56.154</td>
<td>3.557</td>
<td>.060</td>
<td>.014</td>
<td>.468</td>
</tr>
<tr>
<td>Time * Form</td>
<td>2</td>
<td>37.203</td>
<td>18.602</td>
<td>4.011</td>
<td>.019*</td>
<td>.016</td>
<td>.716</td>
</tr>
<tr>
<td>Oral production</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>334.580</td>
<td>167.290</td>
<td>58.887</td>
<td>.000**</td>
<td>.196</td>
<td>1.000</td>
</tr>
<tr>
<td>Form</td>
<td>1</td>
<td>5.703</td>
<td>5.703</td>
<td>.559</td>
<td>.455</td>
<td>.002</td>
<td>.116</td>
</tr>
<tr>
<td>Time * Form</td>
<td>2</td>
<td>9.028</td>
<td>4.514</td>
<td>1.589</td>
<td>.205</td>
<td>.007</td>
<td>.337</td>
</tr>
<tr>
<td>Grammaticality judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>2</td>
<td>159.538</td>
<td>79.769</td>
<td>34.618</td>
<td>.000**</td>
<td>.123</td>
<td>1.000</td>
</tr>
<tr>
<td>Form</td>
<td>1</td>
<td>11.709</td>
<td>11.709</td>
<td>1.881</td>
<td>.171</td>
<td>.008</td>
<td>.277</td>
</tr>
<tr>
<td>Time * Form</td>
<td>2</td>
<td>15.334</td>
<td>7.667</td>
<td>3.327</td>
<td>.037*</td>
<td>.013</td>
<td>.629</td>
</tr>
</tbody>
</table>
**APPENDIX 6.17**

RQ3 - Summary of results of separate 3x2 repeated-measures ANOVAs for each experimental group: Time by Form

<table>
<thead>
<tr>
<th>Time x Form</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Written production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>2</td>
<td>19.309</td>
<td>9.655</td>
<td>1.943</td>
<td>.152</td>
<td>.057</td>
<td>.388</td>
</tr>
<tr>
<td>Practice SP</td>
<td>2</td>
<td>11.974</td>
<td>5.987</td>
<td>1.750</td>
<td>.182</td>
<td>.053</td>
<td>.353</td>
</tr>
<tr>
<td>Practice EC</td>
<td>2</td>
<td>36.314</td>
<td>18.157</td>
<td>3.528</td>
<td>.035*</td>
<td>.099</td>
<td>.637</td>
</tr>
<tr>
<td>Practice EP</td>
<td>2</td>
<td>32.882</td>
<td>16.441</td>
<td>3.116</td>
<td>.051*</td>
<td>.089</td>
<td>.580</td>
</tr>
<tr>
<td>Practice NF</td>
<td>2</td>
<td>5.474</td>
<td>2.737</td>
<td>1.317</td>
<td>.276</td>
<td>.043</td>
<td>.274</td>
</tr>
<tr>
<td>Exposure SP</td>
<td>2</td>
<td>9.452</td>
<td>4.726</td>
<td>1.414</td>
<td>.252</td>
<td>.052</td>
<td>.290</td>
</tr>
<tr>
<td>Exposure EP</td>
<td>2</td>
<td>1.103</td>
<td>.551</td>
<td>.086</td>
<td>.917</td>
<td>.004</td>
<td>.062</td>
</tr>
<tr>
<td>Exposure NF</td>
<td>2</td>
<td>4.694</td>
<td>2.347</td>
<td>1.712</td>
<td>.190</td>
<td>.060</td>
<td>.344</td>
</tr>
<tr>
<td><strong>Oral production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>2</td>
<td>9.207</td>
<td>4.603</td>
<td>1.385</td>
<td>.258</td>
<td>.043</td>
<td>.287</td>
</tr>
<tr>
<td>Practice SP</td>
<td>2</td>
<td>4.081</td>
<td>2.040</td>
<td>1.050</td>
<td>.357</td>
<td>.037</td>
<td>.224</td>
</tr>
<tr>
<td>Practice EC</td>
<td>2</td>
<td>12.608</td>
<td>6.304</td>
<td>1.487</td>
<td>.234</td>
<td>.044</td>
<td>.306</td>
</tr>
<tr>
<td>Practice EP</td>
<td>2</td>
<td>2.882</td>
<td>1.441</td>
<td>.545</td>
<td>.582</td>
<td>.017</td>
<td>.136</td>
</tr>
<tr>
<td>Practice NF</td>
<td>2</td>
<td>3.170</td>
<td>1.585</td>
<td>1.889</td>
<td>.160</td>
<td>.061</td>
<td>.377</td>
</tr>
<tr>
<td>Exposure SP</td>
<td>2</td>
<td>8.357</td>
<td>4.179</td>
<td>2.379</td>
<td>.103</td>
<td>.084</td>
<td>.460</td>
</tr>
<tr>
<td>Exposure EP</td>
<td>2</td>
<td>.333</td>
<td>.167</td>
<td>.038</td>
<td>.963</td>
<td>.002</td>
<td>.055</td>
</tr>
<tr>
<td>Exposure NF</td>
<td>2</td>
<td>4.009</td>
<td>2.005</td>
<td>1.738</td>
<td>.185</td>
<td>.060</td>
<td>.349</td>
</tr>
<tr>
<td><strong>Grammaticality judgment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice SC</td>
<td>2</td>
<td>11.474</td>
<td>5.737</td>
<td>2.074</td>
<td>.134</td>
<td>.061</td>
<td>.412</td>
</tr>
<tr>
<td>Practice SP</td>
<td>2</td>
<td>3.675</td>
<td>1.837</td>
<td>.953</td>
<td>.391</td>
<td>.030</td>
<td>.208</td>
</tr>
<tr>
<td>Practice EC</td>
<td>2</td>
<td>16.490</td>
<td>8.245</td>
<td>3.321</td>
<td>.042*</td>
<td>.094</td>
<td>.609</td>
</tr>
<tr>
<td>Practice EP</td>
<td>2</td>
<td>18.588</td>
<td>9.294</td>
<td>3.805</td>
<td>.027*</td>
<td>.106</td>
<td>.673</td>
</tr>
<tr>
<td>Practice NF</td>
<td>2</td>
<td>3.689</td>
<td>1.844</td>
<td>1.342</td>
<td>.269</td>
<td>.044</td>
<td>.278</td>
</tr>
<tr>
<td>Exposure SP</td>
<td>2</td>
<td>1.024</td>
<td>.512</td>
<td>.284</td>
<td>.754</td>
<td>.011</td>
<td>.093</td>
</tr>
<tr>
<td>Exposure EP</td>
<td>2</td>
<td>5.564</td>
<td>2.782</td>
<td>1.092</td>
<td>.344</td>
<td>.044</td>
<td>.231</td>
</tr>
<tr>
<td>Exposure NF</td>
<td>2</td>
<td>5.074</td>
<td>2.537</td>
<td>1.498</td>
<td>.233</td>
<td>.053</td>
<td>.306</td>
</tr>
</tbody>
</table>
APPENDIX 6.18

RQ1/RQ4 - Summary of 3x2x2x3 ANOVA on Written production accuracy for counterparted treatments (Spot & Prompt, Explain & Prompt, No Feedback): Time by Form by Agency by Feedback

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2</td>
<td>430.336</td>
<td>215.168</td>
<td>45.554</td>
<td>.000**</td>
<td>.354</td>
<td>1.000</td>
</tr>
<tr>
<td>Form</td>
<td>1</td>
<td>6.489</td>
<td>6.489</td>
<td>.830</td>
<td>.365</td>
<td>.010</td>
<td>.147</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>17.330</td>
<td>17.330</td>
<td>1.117</td>
<td>.294</td>
<td>.013</td>
<td>.181</td>
</tr>
<tr>
<td>Feedback</td>
<td>2</td>
<td>360.045</td>
<td>180.022</td>
<td>11.602</td>
<td>.000**</td>
<td>.218</td>
<td>.992</td>
</tr>
<tr>
<td>Form x Agency</td>
<td>1</td>
<td>4.283</td>
<td>4.283</td>
<td>.548</td>
<td>.461</td>
<td>.007</td>
<td>.113</td>
</tr>
<tr>
<td>Form x Feedback</td>
<td>2</td>
<td>24.291</td>
<td>12.146</td>
<td>1.554</td>
<td>.218</td>
<td>.036</td>
<td>.321</td>
</tr>
<tr>
<td>Agency x Feedback</td>
<td>2</td>
<td>58.547</td>
<td>29.273</td>
<td>1.887</td>
<td>.158</td>
<td>.043</td>
<td>.382</td>
</tr>
<tr>
<td>Form x Agency x Feedback</td>
<td>2</td>
<td>34.926</td>
<td>17.463</td>
<td>2.234</td>
<td>.114</td>
<td>.051</td>
<td>.443</td>
</tr>
<tr>
<td>Time x Form</td>
<td>2</td>
<td>6.559</td>
<td>3.280</td>
<td>1.266</td>
<td>.285</td>
<td>.015</td>
<td>.272</td>
</tr>
<tr>
<td>Time x Agency</td>
<td>2</td>
<td>5.188</td>
<td>2.594</td>
<td>.549</td>
<td>.578</td>
<td>.007</td>
<td>.140</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>4</td>
<td>138.618</td>
<td>34.655</td>
<td>7.337</td>
<td>.000**</td>
<td>.150</td>
<td>.996</td>
</tr>
<tr>
<td>Time x Form x Agency</td>
<td>2</td>
<td>9.194</td>
<td>4.597</td>
<td>1.774</td>
<td>.173</td>
<td>.021</td>
<td>.368</td>
</tr>
<tr>
<td>Time x Form x Feedback</td>
<td>4</td>
<td>17.668</td>
<td>4.417</td>
<td>1.705</td>
<td>.151</td>
<td>.039</td>
<td>.515</td>
</tr>
<tr>
<td>Time x Agency x Feedback</td>
<td>4</td>
<td>9.774</td>
<td>2.444</td>
<td>.517</td>
<td>.723</td>
<td>.012</td>
<td>.172</td>
</tr>
<tr>
<td>Time x Form x Agency x Feedback</td>
<td>4</td>
<td>25.835</td>
<td>6.459</td>
<td>2.493</td>
<td>.045*</td>
<td>.057</td>
<td>.700</td>
</tr>
</tbody>
</table>
**APPENDIX 6.19**

RQ1/RQ4 - Summary of 3x2x2x3 ANOVA on Oral production accuracy for counterparted treatments (Spot & Prompt, Explain & Prompt, No Feedback): Time by Form by Agency by Feedback

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2</td>
<td>147.303</td>
<td>73.652</td>
<td>29.136</td>
<td>.000**</td>
<td>.265</td>
<td>1.000</td>
</tr>
<tr>
<td>Form</td>
<td>1</td>
<td>10.020</td>
<td>10.020</td>
<td>1.898</td>
<td>.172</td>
<td>.023</td>
<td>.275</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>.156</td>
<td>.156</td>
<td>.016</td>
<td>.901</td>
<td>.000</td>
<td>.052</td>
</tr>
<tr>
<td>Feedback</td>
<td>2</td>
<td>189.389</td>
<td>94.695</td>
<td>9.420</td>
<td>.000**</td>
<td>.189</td>
<td>.975</td>
</tr>
<tr>
<td>Form x Agency</td>
<td>1</td>
<td>.051</td>
<td>.051</td>
<td>.010</td>
<td>.922</td>
<td>.000</td>
<td>.051</td>
</tr>
<tr>
<td>Form x Feedback</td>
<td>2</td>
<td>5.659</td>
<td>2.830</td>
<td>.536</td>
<td>.587</td>
<td>.013</td>
<td>.136</td>
</tr>
<tr>
<td>Agency x Feedback</td>
<td>2</td>
<td>23.639</td>
<td>11.819</td>
<td>1.176</td>
<td>.314</td>
<td>.028</td>
<td>.251</td>
</tr>
<tr>
<td>Form x Agency x Feedback</td>
<td>2</td>
<td>8.081</td>
<td>4.040</td>
<td>.765</td>
<td>.468</td>
<td>.019</td>
<td>.176</td>
</tr>
<tr>
<td>Time x Form</td>
<td>2</td>
<td>1.336</td>
<td>.668</td>
<td>.408</td>
<td>.666</td>
<td>.005</td>
<td>.115</td>
</tr>
<tr>
<td>Time x Agency</td>
<td>2</td>
<td>.421</td>
<td>.210</td>
<td>.083</td>
<td>.920</td>
<td>.001</td>
<td>.062</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>4</td>
<td>71.620</td>
<td>17.905</td>
<td>7.083</td>
<td>.000**</td>
<td>.149</td>
<td>.994</td>
</tr>
<tr>
<td>Time x Form x Agency</td>
<td>2</td>
<td>2.947</td>
<td>1.473</td>
<td>.900</td>
<td>.409</td>
<td>.011</td>
<td>.203</td>
</tr>
<tr>
<td>Time x Form x Feedback</td>
<td>4</td>
<td>3.192</td>
<td>.798</td>
<td>.487</td>
<td>.745</td>
<td>.012</td>
<td>.164</td>
</tr>
<tr>
<td>Time x Agency x Feedback</td>
<td>4</td>
<td>16.233</td>
<td>4.058</td>
<td>1.605</td>
<td>.175</td>
<td>.038</td>
<td>.487</td>
</tr>
<tr>
<td>Time x Form x Agency x Feedback</td>
<td>4</td>
<td>16.430</td>
<td>4.107</td>
<td>2.509</td>
<td>.044*</td>
<td>.058</td>
<td>.703</td>
</tr>
</tbody>
</table>
### APPENDIX 6.20

RQ1/RQ4 - Summary of 3x2x2x3 ANOVA on Grammaticality judgment accuracy for counterparted treatments (Spot & Prompt, Explain & Prompt, No Feedback): Time by Form by Agency by Feedback

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>PES</th>
<th>OP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2</td>
<td>77.106</td>
<td>38.553</td>
<td>16.462</td>
<td>.000**</td>
<td>.166</td>
<td>1.000</td>
</tr>
<tr>
<td>Form</td>
<td>1</td>
<td>.892</td>
<td>.892</td>
<td>.294</td>
<td>.589</td>
<td>.004</td>
<td>.084</td>
</tr>
<tr>
<td>Agency</td>
<td>1</td>
<td>2.140</td>
<td>2.140</td>
<td>.306</td>
<td>.581</td>
<td>.004</td>
<td>.085</td>
</tr>
<tr>
<td>Feedback</td>
<td>2</td>
<td>238.201</td>
<td>119.100</td>
<td>17.045</td>
<td>.000**</td>
<td>.291</td>
<td>1.000</td>
</tr>
<tr>
<td>Form x Agency</td>
<td>1</td>
<td>.412</td>
<td>.412</td>
<td>.136</td>
<td>.713</td>
<td>.002</td>
<td>.065</td>
</tr>
<tr>
<td>Form x Feedback</td>
<td>2</td>
<td>6.976</td>
<td>3.488</td>
<td>1.152</td>
<td>.321</td>
<td>.027</td>
<td>.247</td>
</tr>
<tr>
<td>Agency x Feedback</td>
<td>2</td>
<td>55.703</td>
<td>27.852</td>
<td>3.986</td>
<td>.022*</td>
<td>.088</td>
<td>.699</td>
</tr>
<tr>
<td>Form x Agency x Feedback</td>
<td>2</td>
<td>11.452</td>
<td>5.726</td>
<td>1.891</td>
<td>.157</td>
<td>.044</td>
<td>.383</td>
</tr>
<tr>
<td>Time x Form</td>
<td>2</td>
<td>1.530</td>
<td>.765</td>
<td>.473</td>
<td>.624</td>
<td>.006</td>
<td>.126</td>
</tr>
<tr>
<td>Time x Agency</td>
<td>2</td>
<td>.701</td>
<td>.351</td>
<td>.150</td>
<td>.861</td>
<td>.002</td>
<td>.073</td>
</tr>
<tr>
<td>Time x Feedback</td>
<td>4</td>
<td>51.814</td>
<td>12.954</td>
<td>5.531</td>
<td>.000**</td>
<td>.118</td>
<td>.975</td>
</tr>
<tr>
<td>Time x Form x Agency</td>
<td>2</td>
<td>3.611</td>
<td>1.806</td>
<td>1.117</td>
<td>.330</td>
<td>.013</td>
<td>.244</td>
</tr>
<tr>
<td>Time x Form x Feedback</td>
<td>4</td>
<td>27.350</td>
<td>6.837</td>
<td>4.229</td>
<td>.003**</td>
<td>.092</td>
<td>.920</td>
</tr>
<tr>
<td>Time x Agency x Feedback</td>
<td>4</td>
<td>8.789</td>
<td>2.197</td>
<td>.938</td>
<td>.443</td>
<td>.022</td>
<td>.293</td>
</tr>
<tr>
<td>Time x Form x Agency x Feedback</td>
<td>4</td>
<td>3.816</td>
<td>.954</td>
<td>.590</td>
<td>.670</td>
<td>.014</td>
<td>.192</td>
</tr>
</tbody>
</table>
**APPENDIX 6.21**

Summary of ANOVA results for exit questionnaire (1/2)

<table>
<thead>
<tr>
<th>Item</th>
<th>Practice SC</th>
<th>Practice SP</th>
<th>Practice EC</th>
<th>Practice EP</th>
<th>Practice NF</th>
<th>Exposure SP</th>
<th>Exposure EP</th>
<th>Exposure NF</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Overall evaluation?</td>
<td>3.88 .5</td>
<td>3.80 .86</td>
<td>4.06 .25</td>
<td>3.76 .66</td>
<td>3.79 .802</td>
<td>3.73 .59</td>
<td>3.67 .49</td>
<td>3.54 .52</td>
<td>3.79 .509</td>
</tr>
<tr>
<td>12. Did the tutor run smoothly?</td>
<td>3.94 1.24</td>
<td>4.00 1.13</td>
<td>3.44 1.36</td>
<td>4.00 .93</td>
<td>4.21 1.05</td>
<td>4.00 1.13</td>
<td>4.17 1.11</td>
<td>4.15 .69</td>
<td>3.97 .639</td>
</tr>
<tr>
<td>14. How much did you learn?</td>
<td>2.94 .77</td>
<td>3.20 .86</td>
<td>3.75 .58</td>
<td>2.88 .70</td>
<td>2.86 .77</td>
<td>3.07 .70</td>
<td>3.50 1.09</td>
<td>2.92 .862</td>
<td>3.14 .018*</td>
</tr>
<tr>
<td>15. How was the amount of items?</td>
<td>4.19 1.05</td>
<td>4.20 .68</td>
<td>3.88 1.02</td>
<td>3.88 1.11</td>
<td>3.93 .99</td>
<td>4.07 .96</td>
<td>4.00 1.04</td>
<td>3.69 1.32</td>
<td>3.98 .900</td>
</tr>
<tr>
<td>16. Was the content engaging?</td>
<td>3.88 .50</td>
<td>3.60 .74</td>
<td>3.63 .72</td>
<td>3.82 .73</td>
<td>3.43 .76</td>
<td>3.33 .62</td>
<td>3.50 .52</td>
<td>3.69 .85</td>
<td>3.62 .351</td>
</tr>
<tr>
<td>17. Was it useful for real life?</td>
<td>4.06 .44</td>
<td>3.73 .70</td>
<td>3.75 .68</td>
<td>3.76 .56</td>
<td>4.07 .62</td>
<td>3.80 .68</td>
<td>3.67 .49</td>
<td>3.69 .75</td>
<td>3.82 .475</td>
</tr>
<tr>
<td>18. Was feedback comprehensible?</td>
<td>3.69 1.01</td>
<td>3.77 1.23</td>
<td>3.13 1.12</td>
<td>3.50 1.15</td>
<td>N/A N/A</td>
<td>N/A N/A</td>
<td>N/A N/A</td>
<td>N/A N/A</td>
<td>3.52 .439</td>
</tr>
<tr>
<td>19. Was feedback useful?</td>
<td>3.73 .88</td>
<td>3.77 1.23</td>
<td>4.60 .51</td>
<td>3.50 1.32</td>
<td>N/A N/A</td>
<td>N/A N/A</td>
<td>N/A N/A</td>
<td>N/A N/A</td>
<td>3.90 .027*</td>
</tr>
</tbody>
</table>
Summary of ANOVA results for exit questionnaire (2/2)

<table>
<thead>
<tr>
<th>Item</th>
<th>Practice SC</th>
<th>Practice SP</th>
<th>Practice EC</th>
<th>Practice EP</th>
<th>Practice NF</th>
<th>Exposure SP</th>
<th>Exposure EP</th>
<th>Exposure NF</th>
<th>TOTAL</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. Was feedback detailed enough?</td>
<td>2.87 .83</td>
<td>3.25 .96</td>
<td>3.73 .88</td>
<td>2.87 .91</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3.18</td>
<td>.033*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Did you like the oral mode?</td>
<td>3.07 1.03</td>
<td>3.14 1.01</td>
<td>2.73 1.16</td>
<td>2.75 1.06</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>2.92</td>
<td>.643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. Is this tutor a good learning aid?</td>
<td>4.00 .52</td>
<td>4.07 1.03</td>
<td>4.00 .52</td>
<td>4.12 .60</td>
<td>3.79 .80</td>
<td>4.13 .64</td>
<td>4.00 .43</td>
<td>3.69</td>
<td>.63</td>
<td>3.98</td>
<td>.619</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. Would you use this regularly?</td>
<td>3.38 .72</td>
<td>3.40 1.06</td>
<td>3.56 .89</td>
<td>3.59 .79</td>
<td>3.43 1.02</td>
<td>3.71 .82</td>
<td>3.50 .80</td>
<td>3.46</td>
<td>.78</td>
<td>3.50</td>
<td>.971</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Do you prefer this to face to face?</td>
<td>2.63 .81</td>
<td>2.93 1.16</td>
<td>3.06 1.44</td>
<td>3.35 .99</td>
<td>3.00 1.47</td>
<td>2.80 1.08</td>
<td>2.25 1.05</td>
<td>3.23</td>
<td>1.23</td>
<td>2.92</td>
<td>.288</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. Did transcript buttons help?</td>
<td>4.13 1.15</td>
<td>4.40 .83</td>
<td>4.13 .96</td>
<td>4.06 1.14</td>
<td>4.50 .65</td>
<td>4.67 .62</td>
<td>4.08 .90</td>
<td>4.31</td>
<td>1.18</td>
<td>4.28</td>
<td>.577</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. Did you like your Agency?</td>
<td>4.00 .52</td>
<td>3.93 .70</td>
<td>4.13 .50</td>
<td>3.65 .86</td>
<td>3.43 1.16</td>
<td>3.00 .84</td>
<td>2.50 1.09</td>
<td>2.69</td>
<td>1.18</td>
<td>3.47</td>
<td>.000**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. Did you like Prompt/Continue?</td>
<td>3.33 1.23</td>
<td>3.93 1.03</td>
<td>3.47 .91</td>
<td>3.88 .70</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>3.66</td>
<td>.244</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


*Language Learning & Technology, 4*, 60-81.


Han, Z. (2000). Persistence of the implicit influence of NL: The case of the pseudo-


*TESOL Quarterly, 36*, 543–572.


*Language Learning Journal, 36*(1), 35-44.


Hardison, D. M. (2004). Generalization of computer-assisted prosody training:

Quantitative and qualitative findings. *Language Learning & Technology, 8*, 34-

52.


*Applied Linguistics, 10*, 331-359.


Kowal, M., & Swain, M. (1997). From semantic to syntactic processing: How can we promote it in the immersion classroom? In R. K. Johnson & M. Swain (Eds.),


Unpublished doctoral dissertation, Georgetown University, Washington, D. C.


Leaver, B. L., Rifkin, B., Shekhtman, B., Banner, D., Chaput, P., Evans-Romaine, K., et al. (2004). Apples and oranges are both fruit, but they don’t taste the same: A response to Wynne Wong and Bill VanPatten. *Foreign Language Annals, 37*(1), 125-132.


Tasks and language learning: Integrating theory and practice. (pp. 9-34).
Clevedon, UK; Philadelphia: Multilingual Matters.


Wong, W., & VanPatten, B. (2003). The evidence is IN: Drills are OUT. *Foreign Language Annals, 36*(3), 403-423.

Wong, W., & VanPatten, B. (2004). Beyond experience and belief (or, waiting for the evidence): A reply to Leaver et al.'s 'Apples and oranges.' *Foreign Language Annals, 37*(1), 133-142.


