ABSTRACT

Title of Document: THE ROLE OF RULES, EXAMPLES AND INDIVIDUAL DIFFERENCES IN THE ACQUISITION OF DECLARATIVE AND PROCEDURAL SECOND LANGUAGE KNOWLEDGE

Ilina Kachinske, Doctor of Philosophy, 2016

Directed By: Dr. Robert M. DeKeyser, Second Language Acquisition

The relevance of explicit instruction has been well documented in SLA research. Despite numerous positive findings, however, the issue continues to engage scholars worldwide. One issue that was largely neglected in previous empirical studies - and one that may be crucial for the effectiveness of explicit instruction - is the timing and integration of rules and practice. The present study investigated the extent to which grammar explanation (GE) before practice, grammar explanation during practice, and individual differences impact the acquisition of L2 declarative and procedural knowledge of two grammatical structures in Spanish. In this experiment, 128 English-speaking learners of Spanish were randomly assigned to four experimental treatments and completed comprehension-based task-essential practice for interpreting object-verb (OV) and ser/estar (SER) sentences in Spanish.

Results confirmed the predicted importance of timing of GE: participants who received GE during practice were more likely to develop and retain their knowledge successfully. Results further revealed that the various combinations of rules and practice posed differential task demands on the learners and consequently drew on language aptitude and WM to a different extent. Since these correlations between individual differences and learning outcomes were the least observed in the conditions that received GE during practice, we argue that the suitable integration of rules and practice ameliorated task demands, reducing the burden on the learner, and accordingly mitigated the role of participants' individual differences. Finally, some evidence also showed that the comprehension practice that participants received for the two structures was not sufficient for the formation of solid productive knowledge, but was more effective for the OV than for the SER construction.

THE ROLE OF RULES, EXAMPLES AND INDIVIDUAL DIFFERENCES IN THE ACQUISITION OF DECLARATIVE AND PROCEDURAL SECOND LANGUAGE KNOWLEDGE

By

Ilina Kachinske

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Robert DeKeyser, Chair
Professor Michael Long
Professor Steven Ross
Professor Jeffrey Lidz
Professor Peter Afflerbach

Dedication

To my mother Aneta.

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Robert DeKeyser, for his invaluable guidance and investment in me over the past six years. I thank my wonderful professor for his inspiring mentorship, for his untiring dedication and motivation even when things looked impossible to me, for his invaluable support during all stages of my studies, and for the energy and zeal that he puts into advancing the field of second language acquisition. Without Professor DeKeyser's unfaltering mentorship, and encouragement, the completion of my PhD studies would not have been possible. Furthermore, I would like to thank the members of the committee: Dr. Michael Long, Dr. Steven Ross, Dr. Jeffrey Lidz, and Dr. Peter Afflerbach for their insightful comments and feedback.

I would also like to express my profound gratitude to Dr. Manel Lacorte and the Department of Spanish at the University of Maryland, without whom the data collection wouldn't have been possible. My thanks also go to the instructors, especially Carla Onate, who were extremely understanding and allowed me to take precious time from their classes.

My research assistants Jessica Nolasco and Gernam Zurita played a crucial role in one of the most strenuous stages of the dissertation, i.e., during data collection. I would also like to thank the students who were willing to participate in the experiment, as well as my own students who are very dear to me.

In addition, I would like to thank my colleagues and friends, especially Katya Solovyeva, Susan Benson, Payman Vafaee, Gisela Granena, Goretti Prieto Botana, Aashish Shenoy, Vera Mironova, for their endless support, encouragement, and fruitful conversations at different points during this journey. Additionally, I would like to immensely thank Dr. Goretti Prieto Botana for allowing me to use her experimental materials.

The process of writing the dissertation as well as my course of studies would not have been the same without the wonderful people I worked with in the School of Languages, Literatures, and Cultures. Pamela Deane, Dr. Lauretta Clough, Dr. Gabrielle Strauch, Michelle Dove, thank you for everything; for your love, trust, friendship, intellectual guidance and much more.

Writing the acknowledgements for this dissertation cannot pass without expressing gratitude to all of the professors in my life who contributed to my education, both as a scholar and as a person. I would, therefore, like to single out and thank my primary school teacher, Lenche Dimovska, my high school professors Svetlana Naumovska and Aleksandra Koshcicovska, my professors at the State University in Macedonia Professors Eleni Buzarovska, Ljupco Stefanovski, Marija Janeva, Violeta Svonson, Rajna Koska, Anastazija Kirkova-Naskova, my MA professors John Williams and Napoleon Katsos, as well as my statistics professors at the University of Maryland Dr.

Andre Rupp, Dr. Jeffrey Harring and Dr. Gregory Hancock, as well as my inspiring and wonderful professors at the department of second language acquisition Dr. Michael Long and Steven Ross and Dr. Colin Phillips from the Linguistics Department, to whom I am indebted for all the knowledge they passed on, for our thought-provoking conversations and motivation to grow. I was honored to be their student. I would like to thank all of you for the bits and pieces that you provided me with to pave the way to this journey.

I am especially grateful to my colleague and dearest friend Anastazija Kirkova-Naskova, without whom my career in linguistics would not have been possible, for her constant intellectual and emotional support, for being always there for me no matter what. I love you my sister.

I would like to thank my family, Amphilohije, my parents Aneta and Ljupco, my grandparents Ilija, ZIvka, Dusan, Anastasija, for their unconditional love and their ceaseless upbringing and nurturing of Truth in me, my uncle Riki, my cousins Filip and Luka. I am immensely thankful to my parents-in-law, Timothy and Judith Kachinske, without whose help and support the dissertation would not have been possible. They were extremely supportive and helpful in so many ways.

Finally, from the bottom of my heart I thank my husband, Adam Kachinske, *the superlative of my life*, and our wonderful children, Vasily and Jovana. Adam, I thank you for your bottomless love, inspiring conversations, unselfish sacrifices, the patience to listen to me talking about SLA, May you be eternally blessed!

Table of Contents

List of Tables

.

List of Figures

Chapter 1: Introduction

One fundamental observation in language learning is that first language (L1) acquisition is uniform and highly successful, whereas second language acquisition (SLA) by adults is variable and rather poor (Dörnyei, 2005). While L1 acquisition is characterized by implicit learning processes leading to largely implicit knowledge, the possibility of gaining implicit knowledge in a second language (L2) by means of implicit learning has been less clear. From a more applied perspective, the core issue has evolved around the possibility of obtaining, if not implicit knowledge, at least functionally equivalent automatized explicit knowledge. As highlighted in Lyster & Sato (2013), one important issue that arises in classroom settings is the extent to which explicit knowledge can become part of a learner's spontaneously used knowledge. Unlike children, adults tend to rely on problem-solving, hypothesis-testing strategies that are involved in explicit learning, especially instructed learners. This biases adults towards approaching any cognitive task by using explicit learning mechanisms, at least in instructed and experimental conditions. A key component in explicit learning is acquiring declarative knowledge, which corresponds to grammar rules and vocabulary in the area of language learning. Rules, therefore, can be considered as essential for adults' preferred way of learning. The field of instructed SLA, however, has been divided over the precise role that grammar rules play in the process of language acquisition. Questions such as whether rules should be provided preemptively or reactively, or whether there is any facilitative learning effect when rules are repeated, have been less researched and remain issues of debate. They will be the main motivation for the present study. Although there is increasing evidence that explicitly instructed learners are superior to learners who are not instructed in their knowledge (e.g., Goo et al., 2015; Norris & Ortega, 2000, Spada & Tomita, 2010), there is also some evidence that explicit learning mainly leads only to a short-term learning effect, with substantial knowledge losses over time (Lado et al., 2014; Stafford et al., 2011), that there is little to no benefit of explicit rules when practice is task-essential (e.g., Benati, 2004; Farley, 2004a; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996; Wong, 2004b; cf. Prieto Botana, 2013), that the effectiveness of providing explicit rules hinges upon the type of structure under investigation (e.g., Farley, 2004b; Fernández, 2008; Henry, Culman & VanPatten, 2009), and that whether learners can benefit from explicit instruction or not may depend on when it is delivered as well as on individual differences (ID) in cognitive aptitudes (e.g., Erlam, 2003; Robinson, 1996; Presson et al., 2014; Stafford et al., 2011). Studies that explored the effectiveness of explicit-deductive versus explicit-inductive instructional treatments, likewise, have yielded inconsistent and conflicting evidence (Abraham, 1985; Erlam, 2003; Haight et al., 2007; Herron & Tomasello, 1992; Robinson, 1996; Rosa & O'Neill, 1999; Shaffer, 1989).

In addition, studies that investigated the role of providing grammar rules have not been systematic in the way grammar rules have been delivered. While some studies provided rules prior to practice (e.g., Alanen, 1995), other studies had grammar rules being provided both before and during practice (e.g., DeKeyser, 1995; Robinson, 1996), only during practice (e.g., Lado et al., 2014; Presson et al., 2014; Stafford et al., 2011), or only after practice (e.g., Michas & Berry, 1994), making it very hard to draw any systematic conclusions regarding the beneficial role of rules. It is very likely that the timing and the manner in which explicit grammar rules are delivered makes a difference in the observed effectiveness of instruction types (Leow, 1998; Norris & Ortega, 2001; Watari, 2014).

Theoretical accounts addressing questions such as when focus on form should occur in the classroom, whether grammar should be provided at the beginning, during, or only after learners have developed some initial communicative competence, or how frequently rules should be revisited in order to yield beneficial results, abound in the SLA literature (e.g., DeKeyser, 1997, 1998, 2015; DeKeyser & Criado-Sánchez, 2012a,b; Doughty & Williams, 1998; N. Ellis, 2002, 2005; R. Ellis, 2002, 2006; Hinkel, 2006; Larsen-Freeman, 2009; Lightbown, 1998, 2000; Long, 1991, 2007; Long & Robinson, 1998; Mitchell, 2000; Schmidt, 1990, 1993, 1994a, 1994b; 2001; Spada, 2007; VanPatten, 1996, 2002). However, there is surprisingly little empirical research in the field of SLA investigating the effect of some variables in explicit instruction, in particular the timing of explicit grammar explanation (GE) in relation to practice that leads to substantial immediate and long-term gains (Spada, 1997).

There is however, a large body of literature investigating the use of rules outside of the SLA field, both within Skill Acquisition Theory (SAT), and in other areas of Cognitive Psychology (CP), suggesting that rules are not always equally beneficial to learners, but rather that the specific timing of providing explicit rules may be essential for maximizing learning effects (e.g., Sallas et al., 2007). Furthermore, based on the current understanding of human cognitive architecture, in particular of the limits of working memory capacity (Baddeley, 2010; Just & Carpenter, 1992), it is reasonable to assume that any instructional treatment that makes heavy demands on working memory, such as an inductive one for instance, will not be very effective at best and detrimental to learning at worst (Kirschner et al., 2006), except perhaps for learners with very high WM. In a similar vein, we can assume that any instructional treatment that eliminates the need for learners to overburden their working memory, but instead focuses on proceduralizing their knowledge will prove to be a more effective one. Suggestions have been made in the field of educational psychology that novice learners should be provided with maximal guidance during the learning process, which can be relaxed only once learners have developed knowledge that is stored in long-term memory (Kirschner et al., 2006).

While type of instruction and practice L2 learners receive has an impact on second language (L2) grammar learning, learner-internal cognitive variables also play a role in any skill acquisition process, including language learning. As Larsen-Freeman (2009) notes, when teaching grammar, teachers should always be aware of who the learners are in terms of their preferences for learning, their cultural background, and their cognitive as well as affective individual characteristics. Matching L2 instruction to the way learners learn is believed to be the most effective way of instructing L2 learners and hence to be the crux of future SLA research (de Graaff & Housen, 2009). Evidence has accumulated that L2 ultimate attainment in adults is highly impacted by individual differences, in particular language aptitude (e.g., Carroll, 1981; DeKeyser, 2000; DeKeyser & Koeth, 2011; Dörnyei, 2005; Dörnyei & Skehan, 2003; Doughty et al., 2010; Linck et al., 2013; Skehan, 1990; Vatz et al., 2013). Evidence from experimental studies on Aptitude-by-Treatment Interaction (ATI) demonstrates that the benefits of a particular instructional treatment are constrained by individual differences (e.g. Erlam, 2005; Linck et al., 2013; Perrachione, Lee, Ha, & Wong, 2011; Robinson, 1997; Sheen, 2007; Suzuki, 2013; Vatz et al., 2013; Wesche, 1981). It may be important, therefore, to examine also to what extent individual differences mediate the effects of timing of different components of instruction. There is some evidence to suggest that different cognitive variables play a role at the early versus later stages of skill acquisition (Ackerman, 1988; Morgan-Short et al., 2014). Since studies so far have mainly explored the interaction between cognitive individual differences and instructional treatment by looking at the final outcome of learning, the secondary focus of the present dissertation is to look at this interaction in a more granular fashion.

In sum, although studies have demonstrated a beneficial effect of explicit instruction, studies have been very unsystematic in the timing and duration of providing explicit grammar rules, and no empirical study has directly contrasted and manipulated the provision of rules before and during practice. There are reasons to believe that the right conditions need to be in place so that explicit grammar rules become beneficial for the learner. Theoretical accounts from cognitive psychology predict that the best time to provide explicit rules would be just when learners need them. In that way the problem-solving process does not overburden our limited working memory capacity, and attentional resources are freed for the required proceduralization of declarative knowledge. Pedagogical observations to the same effect have also been made, suggesting that the best time to provide learners with grammar rules or error correction is precisely at the moment when learners are negotiating for meaning and need the linguistic information to get their meaning across appropriately and accurately (Long, 1991).

Situated in the framework of John Anderson's Adaptive Control of Thought (ACT-R) model of cognitive skill acquisition (Anderson, 1993; Anderson et al., 2004; Anderson & Lebiere, 1998), the goals of the present study are:

1. To explore the role of prepractice GE on the formation of declarative and proceduralized knowledge.
2. To explore the role of GE during practice on the formation of declarative and proceduralized knowledge.
3. To explore the extent to which and the stage at which cognitive individual differences modulate the effectiveness of the different sequences of rules and practice.

Specifically, this dissertation examined the relative effectiveness of four instructional treatments that differ in their timing of providing explicit rules. The first instructional group received both GE before and during practice (GEb+ GEd+), the second instructional group only received GE before practice (GEb+ GEd-), the third instructional group only received GE during practice (GEb- GEd+), and the fourth instructional group did not receive any GE (GEb- GEd). The dissertation also explored the degree to which different components of language learning aptitude and working memory are predictive of learning explicitly in the early versus later stages of the learning process.

The issue of the timing of explicit grammar rules is of high practical and theoretical importance. Current methodological trends found in Computer Assisted Instruction (CAI) or blended courses tend to have students go through the grammar rules themselves, prior to coming to the (virtual) classrooms, so that there is ample time for classroom communicative activities. The sequential and simultaneous ordering of rules and practice is of theoretical importance because it provides a platform to test the applicability of the Adaptive Control Theory (ACT-R) model of SAT to SLA. Such questions as when rules should be provided in relation to practice, whether the manipulation of the relationship between rules and practice, including the spacing between them or the sequential or simultaneous presentation differentially affect the proceduralization of declarative knowledge have not yet been addressed. Investigating the specific nuances that bring about the beneficial effects of explicit instruction is theoretically important for providing a better understanding of the learning process itself, but is also practically significant because it will inform language practitioners about the most optimal integration of rules and practice.

Chapter 2: Review of the Literature and Theoretical Framework

The following sections turn to the literature on explicit learning and practice. Specifically, the following sections examine (1) studies that have utilized explicit grammar rules and practice in SLA as part of at least one experimental condition, (2) studies that investigated the issue of deductive versus inductive presentation of rules, (3) research in CP on top-down and bottom-up learning processes, (4) theoretical cognitive models from CP relevant to instruction and practice, and (5) SLA and CP studies regarding individual differences in language and skill acquisition.

Explicit instruction and practice within SLA

The view of the facilitative effect of explicit teaching of grammar in the L2 language classroom has shifted dramatically over time, from grammar-based teaching to more communicative approaches. At the same time, strong empirical evidence from research on instructed second language acquisition (ISLA) provides support for the advantages of explicit learning, as demonstrated by studies conducted in classroom settings (e.g., Spada & Lightbown, 1993; White, Spada, Lightbown, & Ranta, 1991), and in laboratory settings (e.g., Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; N. Ellis, 1993; Robinson, 1996, 1997). This finding has been corroborated by several meta-analyses (Goo et al., in press; Norris & Ortega, 2000; Spada & Tomita, 2010). The general finding from the meta-analyses is that "on average, instruction that incorporates explicit (including deductive, inductive, as well as explicit information incorporated in feedback) techniques leads to more substantial effects than implicit instruction" (Norris & Ortega, 2000:500). Consistent with the results obtained by Norris & Ortega (2000), Goo et al. (2015) provide evidence for the effectiveness of explicit learning and its superiority to implicit learning including more recent studies (between 1999 and 2011), while Spada & Tomita (2010), demonstrate a facilitative effect of explicit information for both simple and complex rules and in both controlled and spontaneous use. It should be kept in mind, however, that several researchers have criticized the claim that evidence leans in favor of explicit learning on grounds of recurrent methodological biases that give an advantage to explicit learning (Doughty, 2003; de Graaff, 1997; Yang & Givón, 1997). Namely, in response to Norris and Ortega's (2000) meta-analysis, Doughty (2003) outlines several serious problems. First, the training participants receive is of very short duration, and as such, does not allow implicit learning to take place. The testing is immediately after the training, with very few experiments employing delayed post-tests. Moreover, there is also a bias towards explicit learning, in that the majority of the outcome measures are also explicit in nature, such as discrete-point or declarative knowledge-based measures. Finally, the target structures are rather simple, once again favoring explicit learning. Taken together, these characteristics do not allow implicit learning to develop, and many times place explicit learning at an advantage.

In all three of the above-mentioned meta-analyses, instructions were coded as explicit in several instances: if rules were explicitly provided as part of the instruction (explicit deduction), if learners were instructed to find metalinguistic generalization about presented stimuli (explicit induction), or if metalinguistic feedback was provided during training. The final sample in the meta-analyses included both classroom-based and laboratory-based instruction studies. Since the present research explores the role of explicit grammar information in relation to practice, in a controlled laboratory setting, the literature review of the present paper will be limited to studies that provided explicit rules as part of a treatment in laboratory settings. The review will first start providing some necessary terminological definitions and will subsequently present empirical studies that compared the relative effectiveness of two or more instructional treatments by manipulating the degree of explicitness of the instruction.

Definitions of terms

One source of difficulty and confusion concerning the comparison of empirical studies in the field of SLA is the inconsistent use of terminology. The main purpose of this section is to define key terminology that will be used throughout the dissertation. One crucial distinction that should be made is between instruction, learning, knowledge and memory. Instruction can be implicit or explicit and incidental or intentional, depending on the learning process intended by the teacher. Learning itself can proceed implicitly or explicitly, incidentally or intentionally, not necessarily as intended by the teacher, but depending on how the learner approaches the learning task. The resulting knowledge can again be implicit or explicit, largely but not entirely because of the initial learning process. The defining criterion here is whether the learner is aware or not of the knowledge (as it is retrieved). If the target rules are not consciously accessible, knowledge is implicit, but when rules are part of a learner's conscious system then the knowledge is explicit.

The notions of explicit and implicit learning differ 'in the extent to which actions (or decisions) are driven by conscious beliefs (Berry, 1994: 147), or the extent to which learning occurs with (explicit) or without (implicit) concurrent awareness (DeKeyser, 1995). As DeKeyser (2003) highlights, it is also important to distinguish implicit and explicit learning from deductive and inductive learning. Deduction is defined as an instructional process that moves from the general to the specific, or from rules to examples. Deductive instruction involves explanation of a language rule at the beginning of a learning session, before students are exposed to more examples. On the other hand, induction is a process that moves from the specific to the general. Learners are exposed to instances illustrating a specific target rule and try to arrive at metalinguistic generalizations on their own (Decoo, 1996; DeKeyser, 1995, 2003; Erlam, 2003). Following an inductive treatment, learning can either proceed implicitly or explicitly; when the instruction is deductive, however, learning is always explicit.

To supplement these clarifications, when learners not only receive instructions to find rules when processing a specific input, but are also encouraged in their search by means of enhanced input, questions, or hints, then learning is not only explicit and inductive, but more specifically guided-inductive (Herron & Tomasello, 1992). An attempt will be made to classify all of the reviewed studies having grammar rule explanation as part of the treatment as being inductive, guided-inductive, or explicit-deductive.

Studies comparing implicit and explicit learning

A large number of studies in the SLA field sought to examine the relative effectiveness of explicit instruction by comparing it to a more implicit or incidental type of instruction. These studies overwhelmingly provide evidence for the beneficial effects of explicit instruction. However, when subsequent studies tried to isolate the reasons for the observed beneficial effects of explicit instruction, contradictory evidence emerged.

Although explicitness is not an issue in this study, it is important to review parts of these studies, as one of the main manipulations in our present study stems from a methodological inconsistency observed in them. Namely, when treatments involve provision of explicit grammar rules, no systematicity exists across studies for how and when the explicit rules are provided, how many times they are revisited and how they are practiced. As a result, it is hard to isolate the specific features of explicit grammar explanation that lead to beneficial learning effects. One general finding from these studies is that when explicit-deductive instruction is contrasted with implicit, incidental or guided-inductive instruction, learners benefit significantly more from the provision of explicit grammar explanation provided that it is combined with structured practice

(Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; Ellis, 1993; Kim, 2013; Michas & Berry, 1994; Robinson, 1996, 1997). What is not clear is what made explicit grammar rules beneficial for the acquisition of the target L2 rules, as explicit grammar explanation was operationalized very differently in these studies. Alanen (1995), Ellis (1993), Michas & Berry (1994) and Kim (2013) presented the explicit rules only before and de Graaff (1997) only during the practice session; DeKeyser (1995) and Robinson (1996) presented the explicit rules prior to the practice session, but also made them available for the participants during the practice. Since rules provide a fundamental role in explicit instruction, it is important to find out when and how often rules should be presented. From the existing studies it is not clear whether the advantage of EI was because rules were presented only at the beginning, before practice and then revisited during practice, or only during practice. In a follow-up reanalysis of Norris & Ortega's (2000) meta-analysis, Watari (2014) also reaches the same conclusion. Ninety-six of the 111 treatments analyzed by Norris & Ortega (2000) provided some metalinguistic explanation. The re-analysis of Watari (2014) demonstrates that of those, 35 treatments received metalinguistic intervention before the activities, 17 before and during the activities, 45 only during the activities, and none after the training session.

In addition, other important specifications such as how detailed the explicit information was, how it was delivered, aurally or visually, and whether there were comprehension questions to ensure that learners could understand the rules, have been left out. Following these studies, subsequent research wanted to gain a better understanding of what exactly makes explicit learning beneficial and what the exact role is of prepractice GE.

One of the studies that looked at the specifics of explicit instruction is Rosa & O'Neill (1999). Manipulating the provision of formal instruction (FI) and directions to search for target rules (RS), Rosa & O'Neill created five experimental groups: +FI, +RS; +FI, -RS; -FI, +RS; -FI, -RS; and

Treatment only (T-only) group which was equivalent to a control group; the latter completed the multiple-choice jigsaw puzzle without FI and RS. In line with previous research, and in particular with the Instructed condition in Robinson (1996), the two groups which received explicit instruction (+FI, +RS and +FI, -RS) outperformed the group that did not receive any formal instruction nor directions to look for rules (-FI, -RS). However, unlike the finding in Robinson's (1996) study, Rosa & O'Neill (1999) did not find any significant difference between the instructed, deductive (+FI, -RS) and the rule-search inductive group (-FI, +RS) in their improvement from pretest to posttest.

With almost identical experimental set-up, Stafford, Bowden & Sanz (2011) explored the effects of manipulating the provision of prepractice grammar explanation (+/-GE) and the degree of explicitness by providing more or less explicit feedback (+/-EF) in the initial acquisition of Latin case morphology. Four experimental conditions were created: +GE, +EF, +GE, -EF; -GE, +EF; -GE, -EF. Participants in all groups completed input-based, interactive practice, consisting of three written and three aural interpretation tasks. Feedback was provided after each response, regardless of its correctness. The more explicit feedback consisted of a correct/ incorrect statement followed by item-specific metalinguistic information reinforcing rules of Latin morphosyntax. The less explicit feedback was in the form of correct/ incorrect only. Initial acquisition was assessed by means of two interpretation tests, a grammaticality judgment test, and a production test.

First, all four groups significantly improved from pretest to posttest, and retained their acquired knowledge three weeks later, when measured with the aural and written interpretation tests. Second, results from the GJT, however, demonstrated that only the GE+ groups performed significantly better on the immediate posttest than on the pretest. However, the delayed posttest demonstrated that within three weeks these groups also had significant losses in accuracy (from posttest to delayed test). The group that received only explicit feedback did not change significantly, but the group that received less explicit feedback demonstrated significant changes from pretest to delayed test (mean pretest 16.7, mean delayed test 19.7).

Results from the written production test showed similar patterns: only the groups receiving GE significantly improved on the immediate posttest, but also showed significant losses in accuracy on the delayed test. However, contrary to the GJT data, the +GE, +EF group significantly sustained their improvement in written production scores from pretest to delayed posttest. The group receiving only explicit feedback approached significance from pretest to delayed posttest, while the group not receiving either explicit instruction or feedback did not change significantly. While the group with explicit feedback approached significance from pretest to delayed test, there was no significant post treatment change in the -GE-EF group.

One interesting finding is that the + GE groups showed clear benefits from the treatment on the written production test (more accurate marking of verb agreement and noun case), even after three weeks. The -GE, +EF group performed similarly to the +GE+EF in that they almost significantly improved from pretest to delayed test in accurate use of noun case morphology as assessed by the written production test, suggesting that explicit information provided at the moment when it is the most relevant may be equally good as prepractice GE.

As the authors point out, the GE in this study was highly structured and relatively brief, which may have impacted the results. What remains open to question is whether *repeated* exposure to GE facilitates language learning. Benati (2004) and Morgan-Short et al. (2010) claimed to have shown no clear benefit for repeated exposure to explicit instruction. However, in Benati (2004), the repeated exposure to GE was never followed by structured exposure to examples/ practice. In Morgan-Short et al. (2010) the explicit GE was provided aurally, for several structures, which must have been burdensome for participants to follow, understand and retain. It is also not clear whether the 33 meaningful examples of articles, nouns, verbs, and adverbs in the artificial language that the explicit group received in Morgan-Short et al. (2010), as opposed to 120 meaning examples of the implicit group, were delivered after the corresponding explicit GE or following the GE part. As we will see below, the CP literature suggests that the distribution and combination of rules and examples has an effect on learning.

In summary, Stafford et al.'s results demonstrated that when learners received metalinguistic feedback alone, they learned and retained their knowledge over a period of three weeks. However, only the group that received prepractice GE plus metalinguistic feedback showed significant improvements in the transfer from input-based practice to output assessment. This group was also the only group that sustained their knowledge over a period of three weeks when measured by a written production test.

In a follow-up study, Lado, Bowden, Stafford & Sanz (2014) investigated the role of incorporating metalinguistic information together with "correct/incorrect" type of feedback. Results demonstrated that while there was a clear advantage of providing concurrent explicit information during practice activities on the immediate posttest, these gains almost disappeared by the time of the delayed tests. An advantage of providing immediate metalinguistic feedback was also observed for performance on untrained items as well as for transfer of skill from input-based practice to productive use of target morphosyntactic structures. Results also demonstrate greater maintenance of gains in the group not receiving metalinguistic explanation.

Finally, Presson, MacWhinney, & Tokowicz (2014) also zoomed in on the beneficial role of explicit teaching. Similarly to Stafford et al. (2011), in their first experiment, they compared the effects of explicit rule presentation plus corrective feedback with those of feedback only. The study also investigated the effects of providing frequent as opposed to diverse exemplars and their interaction with plus or minus rule presentation. The learners' task was to categorize nouns in French by grammatical gender. Employing a within-subjects repeated-measures design, 38 participants were exposed to all four conditions: frequent exemplars with rule instruction, frequent exemplars with feedback only, diverse exemplars with rule instruction, and diverse exemplars with feedback only. During three 20-35 minutes-training sessions, participants were shown a noun in French with its English translation. Their task was to indicate whether the noun was masculine or feminine. For the feedback-only block, after each participant's response, feedback in the form of "correct" and "incorrect" was provided. For the rule instruction block, everything was identical except that after only incorrect trials a short cue explanation as well as a prompt were also displayed that guided participants' attention to the orthographic cues at the end of the words. Pretest to posttest results demonstrated that all instructional conditions produced learning. More relevant to the present literature review is the finding that instruction with rule presentation was more effective than learning without rule presentation, consistent with the reviewed studies and the general trend observed in SLA. In addition, the rule instruction with feedback condition outperformed the feedback-only condition, suggesting that feedback with metalinguistic explanation is better than "correct/incorrect" feedback only. In their second experiment, the effects of the rule presentation condition of Experiment 1 (fromage, -AGE -> masculine) were compared to a cue-highlighting treatment (fromAGE). Correctness feedback was also provided. In accord with previous research (Alanen, 1995), rule instruction produced better performance than mere feature focusing. Results from both experiments together provide sound evidence that the advantage of rule instruction was due to the explicit instruction learners received, and not because their attention was directed to relevant features in the input, suggesting that "the explicitness could increase the transparency of cue mappings and strengthen the resulting representations" (p.728).

Numerous studies within the Processing Instruction (PI) literature have also looked at the role of preemptively providing explicit instruction prior to meaningful input practice. Contrary to what has been observed in the studies reviewed above, earlier PI studies concluded that the role of explicit instruction is redundant when meaningful structured-input practice follows (e.g. Benati, 2004; Farley, 2004a; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996). More recent PI studies have demonstrated that the effectiveness of GE depends on the type of structure, in particular, that prepractice GE is beneficial for more complex linguistic structures (e.g., Farley, 2004b; Fernandez, 2008). However, the conclusions that can be drawn from the PI literature should not be taken for granted due to several methodological shortcomings. Namely, the structured input within the PI studies needs to be task-essential in order to yield beneficial results; when task-essentialness in not present in the tasks at hand, explicit information becomes crucial for learning (Prieto Botana, 2013). Prieto Botana (2013) furthermore demonstrates that even when practice is task-essential, explicit information improves learning in terms of more consistent and durable gains than the ones observed from task-essential practice only. In addition, a recent study has pointed to another methodological gap in the PI studies with regards to the lack of learning when explicit information is provided to participants: Vafaee & DeKeyser (2014) demonstrated that before PI studies claim that EI does not have any additional benefit over Structured Input, special attention should be paid whether participants actually process and understand the EI that is presented to them.

Deductive and inductive instruction in SLA

The relative effectiveness of providing explicit instruction has also been investigated by comparing explicit-deductive versus explicit-inductive approaches when learning L2 grammar (e.g., El-Banna & Ibrahim, 1985; Erlam, 2003; Haight et al., 2007; Herron & Tomasello, 1992; Nagata, 1997; Rosa & O'Neill, 1999; Rose & Ng, 2001; Seliger, 1975; Shaffer, 1989; Takimoto, 2005; Wang; 2002; Xia, 2005). The results reported by these studies are contradictory and inconclusive. While some studies report no overall significant difference between the two approaches (Rosa & O'Neill, 1999; Shaffer, 1989), studies such as Erlam (2003) and Robinson (1996) demonstrated an advantage for deductive instruction. Although the instructional treatments in Robinson (1996) are not formally called deductive and inductive, the more implicit ones (the memory and comprehension-focused groups), as well as the rule-search group can be regarded as receiving inductive treatment, and the instructed one a deductive treatment. Rules in the deductive treatment were presented at the beginning of the training session but were also available for reference during the practice activities. Erlam (2003) also found that deductive instruction worked better than inductive instruction in the learning of direct object pronouns in French, at least in classroom settings. Sixty-nine participants were randomly assigned into deductive, inductive or a control group. The deductive group received explicit instruction on the direct object pronoun before proceeding to practice activities. During all practice sessions, a complete chart with direct object pronouns was available for their reference. Practice involved both comprehension and production activities. The inductive group never received any metalinguistic explanation of the rules, but during the practice activities participants were encouraged to come up with an explanation regarding why something was correct or incorrect after each response. This approach resembles the guided inductive approach taken in studies reviewed below. Results from the immediate and delayed posttests indicated that the deductive group significantly outperformed both the inductive and the control group.

Other studies attempted to investigate a more cognitively motivated inductive approach, consisting of guiding learners to come up with a generalization about the presented stimuli, usually accomplished through questions posed by the instructor. Haight, Herron, & Cole (2007), as well as Herron & Tomasello (1992), compared the effectiveness of a deductive versus guided inductive instructional approach and demonstrated an advantage for the guided inductive approach. Both studies explain this advantage being due to the active role of the learner in the construction of meaning and form.

Interim concluding remarks

From the SLA studies reviewed here we can draw the following conclusions and highlight remaining questions:

1. Explicit instruction produces demonstrable learning advantages, at least in laboratory studies when explicit learning is contrasted with implicit learning.
2. Providing rules in isolation does not positively impact acquisition (Michas & Berry, 1994).
3. It is only when the provision of rules is accompanied either by exposure to exemplars or by structured practice that explicit instruction becomes beneficial (Ellis, 1993; Michas & Berry, 1994).
4. Inconsistent results are obtained across studies which investigate the provision and timing of explicit information:
    a. Explicit-deductive > explicit-inductive (Alanen, 1995; Ellis, 1993; Erlam, 2003; Kim, 2013; Presson et al., 2014; Robinson, 1996)
    b. Explicit-deductive < explicit-inductive (Kim, 2013)

c. Explicit-deductive = explicit-inductive (Abraham, 1985; Rosa & O'Neill, 1999; Shaffer, 1989).
d. Explicit guided-inductive > explicit-inductive (Robinson, 1997)
e. Explicit guided-inductive > explicit-deductive (Haight et al., 2007; Herron & Tomasselo, 1992)

5. The studies that looked into the role of metalinguistic information before practice or feedback during practice have observed immediate gains but also significant knowledge losses (Lado et al., 2014; Stafford et al., 2011).
6. There seems to be evidence that providing concurrent explicit information either in the form of metalinguistic rules or in the form of negative evidence (feedback correct-incorrect) eliminates the need for pre-practice GE (Presson et al., 2014; Stafford et al., 2011).
7. When explicit information is delivered as part of feedback, it is not clear how detailed the negative evidence should be. Conflicting results have been obtained from two studies. While negative evidence alone was better than negative evidence plus metalinguistic feedback in Lado et al., (2014), Presson et al., (2014) demonstrated that providing concurrent explicit information yielded superior results than negative evidence alone.
8. Explicit instruction varies greatly from study to study. While rules were presented only at the beginning of the training session in Ellis (1993), Michas & Berry (1994) and Kim (2013), rules were presented both prior to the practice session and during the practice session in DeKeyser (1995) and Robinson (1996). In de Graaff (1996) rules were only presented while practicing. In other studies, more or less explicit information was incorporated as part of feedback (Stafford et al., 2011; Lado et al., 2014; Presson et al.,

2014). It is quite reasonable to assume, then, that the inconsistencies in results obtained from empirical studies so far may be due at least in part to the unsystematic way of delivering explicit-deductive instruction.

9. It is not clear what constituted explicit information, how detailed and understandable it was, how it was delivered, and whether the design of the experiment incorporated some comprehension check concerning the understanding of the grammar rules. All of these points varied greatly from study to study.
10. Usually there is one structure per study. When studies have included structures of varying complexity and difficulty, they have used very different criteria for motivating the choice of a simple versus complex structure.

From the above, the following **remaining questions** can be outlined**:**

1. How should rules and example be integrated? Questions such as when the explicit instruction is delivered, that is before, while, or after activities or training sessions have not been considered important and hence have not received much attention. More specifically, what is the exact role of pre-practice GE, when experimentally contrasted with during-practice GE? While evidence from existing studies is contradictory, there has not been an empirical study in SLA to isolate the effects of these specific timings of GE provision.
2. Can the benefits of explicit instruction observed so far be generalizable across L2 structures that pose different processing problems?
3. What will be the effect of practice for the initial versus subsequent stages of language development?
4. Is there an interaction between different integrations of rules and examples, on one hand, and language aptitudes, on the other hand?

We now turn to literature in CP, to provide some potential answers concerning the issues outlined above. The following section will first provide a summary of empirical studies in CP investigating the effects of rule and example learning and will then move on to outline the principles of skill acquisition theory as instantiated by Anderson's ACT-R cognitive model of learning.

Synergy between rule and example learning in CP

In contrast to the field of SLA, studies in CP have explored the interaction between top-down and bottom-up learning since the 80's and have provided accumulating evidence in support of their synergetic effect in various complex paradigms such as finite-state and biconditional artificial grammars (Mathews et al., 1989; A. S. Reber, Kassin, Lewis, & Cantor, 1980; Sallas et al., 2007), serial reaction time tasks (Willingham et al., 1989), control of a complex task – minefield navigation task (Sun, 2006; Sun et al., 2001; Sun et al., 2005; Sun, Zhang, Slusarz, & Mathews, 2007), as well as a dynamic control task (Lane et al., 2008; Stanley et al., 1989). The general conclusion is that learners who receive mixed training outperform those in single conditions. Moreover, in several studies, participants' knowledge improved when they were asked to verbalize while performing the task at hand, pointing to a beneficial role for explicit knowledge.

For instance, using a finite-state artificial grammar, in their second experiment Reber et al. (1980) had both pure learning and mixed groups. Besides the "pure" groups, i.e., implicit and explicit only, one group had implicit learning from the stimuli followed by explicit learning of the diagram representing the grammar, the second explicit followed by implicit learning, and the third implicit training, explicit instruction and again implicit training. Measured by a string discrimination test, results demonstrated that the mixed groups outperformed the pure groups, with a more facilitative effect when participants were briefly exposed to the artificial grammar diagram prior to receiving the "implicit learning" (E-I > I-E = I-E-I > I =E).

Mathews et al. (1989 third and fourth experiments) also explored the sequential interaction between implicit and explicit learning. In their third experiment they used finite-state grammar and did not find any significant effect for mixed training, although the pure implicit and the implicit-then-explicit group performed slightly better than the pure explicit and "explicit-then-implicit" group. Their fourth experiment tested the same interaction using a biconditional grammar. Contrary to what was observed in Reber et al., (1980), results in this study showed that the implicit-then-explicit group performed significantly better than all of the other groups. Moreover, all groups performed significantly better than the purely implicit group. The exemplars from the finite-state grammar were constructed to have high family resemblance, lending themselves to automatic implicit learning mechanisms capable of identifying common patterns. The exemplars from the biconditional grammar, on the other hand, had few family resemblances, so that participants had to go beyond similarity among the exemplars. These findings accord well with the CLARION model, briefly reviewed below. Implicit learning becomes more prominent with more complex relations, with a large number of family resemblance patterns or sequences with a high degree of statistical structure. If, however, the relations to be learned are more salient and the input dimensions are relatively few, then learning will benefit from explicit learning or simultaneous interaction between the levels (Reber et al., 1980; Sun et al., 2005). Similar observations have been made in the SLA literature (DeKeyser, 1995; de Graaff, 1997; Robinson, 1996).

Further research has suggested that manipulating the synergetic effect of implicit and explicit learning may lead to different advantages regarding participants' speed and accuracy. For instance, Domangue et al. (2004) used the same artificial language paradigm to train participants with hybrids of model-based and experience-based learning, corresponding to the terminology of explicit and implicit learning used in other studies. Model-based learning makes use of the actual rules from which the exemplars are generated, whereas experience-based learning involves acquiring the target rules from mere exposure to exemplars. Results from their experiments demonstrated that whereas experience-based learning produced fast but not so accurate performance and the model-based learning produced accurate but slow performance, the integrated conditions (both model and experience-based learning simultaneously) led to intermediate levels of accuracy and speed, and interleaved learning (mixture of model and experience-based learning) irrespective of the order, led to increased accuracy and slower performance. Domangue et al. (2004: 1010) highlight that if both speed and accuracy are important, then mixed training is the optimal choice.

A more recent study comes from Lane et al. (2008) who used a dynamic control task to manipulate whether participants received model-based training prior to task experience and the possible varieties of such training. There were three groups: participants in the experiential group were only told to try and maintain the system at the goal state; participants in the quiz-only group had the same instructions but received additional quizzing about their knowledge in terms of input-output pairs; the third group, in addition to the training and quizzing, reviewed a partial look-up table with three input-output pairs. The look-up table was given to the participants before the training. The results suggested that providing the model-based training with the look-up table led to superior performance as compared to the other two groups. In line with the Domangue et al.

(2004) results, the hint + quiz group improved in terms of accuracy, but at the cost of speed. These results are replicated in their second experiment, in which instead of the partial look-up table, participants were exposed to a full look-up table prior to receiving the training. It was demonstrated that the table group, which received explicit information of the target rule, outperformed the experiential group in terms of accuracy. Several additional tests suggest that both groups deteriorated from a non-speeded to a speeded performance test, with no difference between groups on a transfer test.

Following the call for more extensive practice with letter strings in the artificial grammar paradigm, as well as the call for a more sensitive outcome measure by Mathews and Cochran (1998), the interaction of implicit and explicit learning was further explored by Sallas et al. (2007). These authors demonstrated that learning in the artificial grammar paradigm could be facilitated by providing explicit instruction, as long as the instructions are provided exactly when learners need them. Whereas the provision of the artificial grammar diagram in Domangue et al (2004) was used as a mapping tool for the letter strings to the diagram, Sallas et al. (2007) focused participants on the entire letter strings, instead of on decomposing them. During training, participants were exposed to erroneous letter strings and had to find the illegal letters within the strings. To help them achieve this, participants were either provided with letter cues, a whole diagram cue, or no help at all. The first two types of aids were achieved by animation and highlighting of the relevant information. Participants tried to correct the letters strings until they reached 70% accuracy. Feedback was also provided. Hence, the training conditions in Sallas et al. (2007) differed with respect to whether they required the processing of a whole diagram, or individual letters, corresponding to high-level and low-level grammar knowledge, respectively. There were four training groups in the first experiment: diagram assistance, letter assistance, no assistance, and no training (control). The whole experiment lasted for six sessions, each one hour long. Upon the 20 minutes of training, a cued-generation test was administered in sessions 1, 2, 4, and 5. No training was administered during sessions 3 and 6. Instead, two versions of a cued generation test were administered. The duration of the first version was fixed, i.e., participants had only 10 minutes to generate acceptable letter strings. In the second version of the test, the fixed trial test, participants were told that they did not have a time limit, but that they were only allowed to generate 60 letter strings. Results showed that more acceptable letter strings were produced by the fixed time test than by the fixed trial test, with all experimental groups outperforming the control group. No other significant differences were observed. With respect to perfect strings, i.e., high levels of accuracy (operationalized as the number of perfect strings generated on the first attempt over the total number of first attempts), results showed that the diagram assistance group generated significantly more perfect strings than all the other training conditions. There were no significant differences between diagram and letter assistance conditions regarding speed. As the authors conclude, the learners in the diagram assistance condition generated significantly more perfect strings on their first attempts and were as fast as participants in the other conditions. This suggests that higher-level grammatical knowledge provided just when participants need it produces more accurate and detailed knowledge, which is reminiscent of the theoretical claim put forward by Anderson's Skill Acquisition Theory that declarative knowledge and the tasks used to utilize this knowledge should be close together for proceduralization to be effective. The animation in the diagram assistance task highlighted the relevant part of the diagram just when participants were editing the corresponding part of the strings. While experiment 2 confirmed these findings, it also showed that those participants who received the animated version of the diagram produced more accurate strings than those who received a static version of the diagram, suggesting that although model-based knowledge is beneficial for both accuracy and speed, the way it is delivered may be crucial for optimal results. Finally, results from the third group in the second experiment, which prediction training with an animated diagram, in which participants were asked to predict which letter should appear next, revealed large individual differences among the participants. While some individuals generated high levels of perfect strings, others did very poorly. In summary, acquiring high-level knowledge to produce perfect strings hinged on whether participants were focused on letter-by-letter, chunk-by-chunk, or whole diagram associations. The more detailed the model-based instruction, the more accurate and stable the performance.

A slightly different perspective emerges from extensive experimentation with the two-level learning model developed by Sun et al., (2001, 2005, 2007). If Anderson's model is an exclusively top-down model, CLARION's (Connectionist Learning with Adaptive Rule Induction ON-line) two levels allow more flexibility to accommodate both top-down and bottom-up learning. Work on cognitive architectures has mainly focused on top-down models and has not given an equal chance to bottom-up learning. The latter can either proceed simultaneously (learning both implicit and explicit knowledge) or sequentially (first implicit and then explicit) according to Sun et al. (2007). Without going into detail about all the model's subcomponents, CLARION is a dual-representation structure with a top and a bottom level corresponding to explicit and implicit knowledge representations. Depending on whether the top level is switched on or off, learning may proceed either top-down or bottom-up. That is, when declarative knowledge is available and in a format easily understood and applied, learning is what the ACT model reflects - top-down. When no declarative knowledge is available or the rule to be learned is too hard, then through trial-and-error interactions with the world, in general, and the stimuli in particular, the learner proceeds from exposure to rules. Finally, the model also allows learning to occur with the two levels interacting simultaneously.

Experimenting with a minefield navigation task with both human and model participants, Sun et al. (2001) used a complex minefield navigation task in which an agent has a limited time to get to the target. Experimental groups received either a standard training corresponding to the experiential training in other studies, a verbalization, over-verbalization, or dual-task training. Results demonstrated several important findings: single-task conditions led to significantly better performance than the dual-task condition in terms of learning, and a significant increase in performance was observed in the verbalization conditions, as well as a floor effect in the over-verbalization group. It seems that verbalization forced participants to approach the task more explicitly, which in turn enhanced their performance, suggesting that explicit processes and declarative knowledge at the top-level improve learning. Over-verbalization operating on the top-level, on the other hand, seems to have interfered with the implicit learning at the bottom-level. In terms of transfer, explicit knowledge obtained at the top level helped in the improvement of the learned skill transfer. A time lag between implicit and explicit knowledge was also observed, a finding that the authors interpret as evidence of explicit learning being prompted by implicit learning. The authors also note that in other tasks, like artificial grammar learning, explicit and implicit knowledge may be more closely related than previously thought.

The general finding from the reviewed studies demonstrates that groups that are exposed to both implicit and explicit learning outperform the pure learning conditions. More specifically, in line with the SLA research, there is evidence suggesting that providing explicit information - be it in the form of rules, hints, or quizzes - followed by examples yields superior performance. It was also shown that this beneficial effect of EI is contingent upon the exact time when it is delivered

(Sallas et al., 2007), as well as the specific structure for learning (cf. Matthews et al., 1989; Reber et al., 1980). Of direct interest for the present study is that Sallas et al. (2007) revealed that providing EI to participants just as they needed it produces the most accurate results. More specifically, this study provides some answers with regards to when rules should be presented during a training session. Results showed that (a) a combination of top-down and bottom-up processing provides the most optimal way to facilitate learning, (b) the diagram helped learners during the proceduralization stage, (c) learners did not need to commit the declarative rules to memory in order to generate legal grammar strings; instead the diagram that was provided to them just when needed enabled learners to form a stable representation of the grammar rules. Providing declarative knowledge when a behavior facilitated by a corresponding production is needed has a facilitative effect on the building of the production, suggesting that the way EI is delivered and whether it is revisited or not may be crucial for successful acquisition.

Similarly to the SLA field, the operationalization of what constitutes *explicit* learning or training greatly varies from study to study. Whereas Reber et al. (1980) and Domangue (2004) presented the whole diagram to the learners, participants in Matthews et al. (1989) edited erroneous letter strings, thus acquiring explicit knowledge through trial-and-error interactions. Lane et al.'s (2008) hint and hint + quiz groups can also be said to have acquired their explicit knowledge inductively.

The results obtained from studies in SLA and CP reviewed above align well with the theoretical predictions and empirical findings of Anderson's ACT-R cognitive model of skill acquisition. This model will be reviewed in the following sections, and its applications to SLA will be outlined.

Skill Acquisition Theory

Anderson's ACT-R model is a highly influential and one of the most recognized models of learning and cognition (DeKeyser 2015). The ACT-R theory provides a theoretical framework for and some tentative answers to the many questions in SLA regarding the timing and sequencing of rule and example learning for maximal efficiency (Anderson, 1982, 1996, 2005; DeKeyser, 2015). The present work adopts the ACT-R theory as the main theoretical framework and uses its predictions about the acquisition of any cognitive skill as a platform to further test the theory when it comes to learning a second language.

Declarative knowledge

ACT-R is based on one fundamental assumption regarding knowledge representation, namely the distinction between declarative and procedural knowledge as qualitatively different and distinct knowledge representations. Declarative knowledge, defined as "factual knowledge that people can report or describe", knowledge *that*, is organized in chunks comprised of declarative rules (Anderson, 1993:5). These rules are a representation of a set of facts that is open to conscious inspection, reasoning and modification. Declarative knowledge can either be formed by means of explicit-deductive or explicit-inductive learning mechanisms. Some evidence exists to suggest that depending on the task at hand, declarative knowledge is best formed through declarative memory for examples of how procedures should be executed (Anderson & Fincham, 1994; Anderson, Fincham, & Douglass, 1997; Taatgen & Wallach, 2002). During this first stage in skill acquisition, also known as cognitive (Fitts, 1964), declarative (Anderson, 1982), or early stage (VanLehn, 1996), the learner relies on declarative memory to perform a task. One benefit of representing the knowledge in declarative form is its inherent flexibility. Declarative rules do not have directionality in their statement, allowing one to use them in multiple directions, for instance interpretation and generation of a computer programming language. The downside, however, is that the application of a declarative representation is "cognitively intense and slow" (Kim et al., 2013, p. 25; Neves & Anderson, 1981). Within the ACT-R framework, declarative rules are retrieved and used through interpretative production rules, which consist of IF-THEN pairs (described below), but also refer back to the declarative rules, making them slower than the production rules (Anderson, Fincham, & Douglass, 1997; Neves & Anderson, 1981). In other words, each declarative rule or fact must be separately retrieved from memory, interpreted and applied to the current situation.

Proceduralized knowledge

Once a declarative rule is applied on regular basis and is well established, it can be compiled into a production rule, which is regarded as the unit of operation in the second and third stage of skill acquisition. During the transition between the first and the second stage in skill acquisition, the learner starts compiling production rules, but for problematic areas still relies on declarative memory for examples or rules. These production rules constitute proceduralized knowledge and are stored in the form of IF –THEN or CONDITION-ACTION pairs. The "IF" (condition) part defines the circumstance under which the rule applies, while the "THEN" (action) part defines what should be done in that circumstance (Anderson 1993: 5). Procedural knowledge is 'knowledge that people can only manifest in their performance', i.e., knowledge *how* (Anderson, 1993:18).

During the second, associative or transitional stage, knowledge *that* is transformed into knowledge *how* by means of extensive, but meaningful practice. In other words, declarative knowledge is turned into behavior, which leads to proceduralized knowledge (DeKeyser, 2015).

As mentioned earlier, when only declarative knowledge is available, the individual has to retrieve pieces of information from memory and insert them into production rules in order to execute a specific behavior. Once proceduralization has taken place, however, production rules for a specific behavior are available to be used whenever the if (condition) part is satisfied (DeKeyser, 2007b; DeKeyser & Criado-Sánchez, 2012a), eliminating the need for constant buffering of declarative knowledge in working memory. That is why procedural knowledge developed through practice can be applied more rapidly and reliably (Anderson, 1993, Anderson et al., 1997).

One disadvantage of a procedural representation is that this knowledge cannot be inspected. Although the learner has some understanding of the production rules' content, changes cannot be made to the productions. New productions can be created that will eventually delete or restrict the range of applicability of bad productions (Neves & Anderson, 1981). In addition to this, whereas the initial declarative knowledge can be generalizable to various situations, procedural knowledge is highly specific and certainly skill-specific. ACT-R assumes that procedural knowledge is committed to a specific use and cannot generalize to other uses. For instance, practicing with a computer language in 'evaluation,' going from code to result, does not lead to the skill required for generation, going from desired result to code, and vice versa (Anderson, Fincham, & Douglass, 1997). This phenomenon, observed in many empirical studies, has been called the *directional asymmetry* that characterizes skill acquisition. It should be noted that declarative and procedural knowledge are not orthogonal concepts. Both of these types of knowledge can coexist and interact with each other in the course of the L2 development (Anderson, 1980). In fact, Neves & Anderson (1981:62) state that keeping both knowledge representations can bring the most optimal performance. "When speed is needed, the procedural encoding is used. When analysis or change is needed, the declarative encoding is used". The final stage of the proceduralization process is fully automatized knowledge characterized by increased speed, decreased error rate and less susceptibility to interference from/with other tasks.

Rule and example learning within ACT-R

A crucial aspect of any skill learning is the role that examples play in the process of acquisition. Based on experimental data, Anderson, Fincham, & Douglass (1997) proposed a four-stage model accounting for the general course of skill acquisition, arguing that skill acquisition involves four overlapping stages. In the first stage, learners rely on analogy to solve problems. By encountering examples that illustrate the solution of a (similar) problem, the individual maps the solution of the example to the solution of the current problem by means of analogy or direct retrieval. Each time a learner encounters a problem, the possibility exists that a rule will be abstracted, in declarative form. The development of abstract declarative rules lays the foundation for the second acquisition stage. The third stage comes only after extensive practice, whereby learners no longer have to retrieve the declarative rules, but instead use procedural embodiment of the rule that allows more rapid responses, but only work in one direction (directional asymmetry). The fourth stage involves example retrieval, which is faster and more direct than the previous stage. During this stage a speed advantage for repeated examples is usually observed (Logan, 1988). These four stages are not strictly sequenced, but will emerge depending on the nature of the learning task and problem.

Results from Anderson, Fincham, & Douglass (1997) suggest that early in the learning process learners relied both on analogy to previously encountered examples and on declarative rules. With practice, the original examples and rules were slowly forgotten and asymmetric rules developed instead. Evidence for this asymmetry of production rules has been observed in the SLA field as well (DeKeyser 1997; DeKeyser & Sokalski, 1996).

Working memory in ACT

Working memory "is a cognitive structure in which conscious processing occurs" (Kirschner et al., 2006: 75). Working memory is assumed to be necessary to keep information in mind while performing complex cognitive tasks such as reasoning, language comprehension, language production and learning (Baddeley, 2010; Just & Carpenter, 1992; Lovett, Reder, & Lebiere, 1997). Research on working memory has demonstrated two important findings: (1) that working memory is limited in its capacity and duration (Baddeley, 1986), and (2) that working memory varies from individual to individual (e.g., Just & Carpenter, 1992). These two findings undoubtedly have implications for any models of cognitive learning. Models of general human cognitive learning, such as Anderson's ACT-R, have incorporated a parameter of functional limit on working memory to account for differences in working memory exhibited among individuals. While the first ACT-R models had included a general attentional resource parameter (W) common across individuals, Lovett, Reder & Lebiere (1997) empirically demonstrated that the W parameter, describing working memory capacity, can be adjusted such that it accounts for individual ability and predicts individual performance. In other words, varying the W parameter can tune the ACT-R model to an individual's ability. The ACT-R architecture provides a good framework to build and add additional cognitive models that will account for other individual differences.

Right conditions for proceduralization

One important aspect of skill acquisition is what constitutes a smooth and effective transition from declarative to proceduralized knowledge representation. Theoretical literature and empirical findings suggest that there are certain prerequisite conditions essential for increasing the likelihood of successful proceduralization. First, solid and accurate declarative knowledge, either obtained deductively, by means of explicit instructions provided to the learner, or inductively, through processes of analogy and abstraction, must exist. Second, there should be plenty of opportunities for learners to apply this knowledge representation consistently. During these opportunities the declarative knowledge must be available in digestible format at the moment of executing the target behavior (DeKeyser, 2007b, DeKeyser & Criado-Sánchez, 2012a). This does not mean that declarative knowledge needs to be stored in long-term memory; rather it needs to be active in working memory (Anderson & Fincham, 1994).

Anderson & Fincham (1994: 1323) propose that the best avenue for proceduralization is when the declarative representation is in the form of an example that is used in an analogy process. In a follow-up study, Anderson, Fincham, & Douglass (1997) demonstrate that the combination of examples and declarative rules is necessary to make the transition from declarative to production rules. Their results suggest a gradual transition from example-processing to production-rule processing. More empirical support comes from CP studies demonstrating that setting the right conditions for proceduralization by means of providing explicit grammar knowledge just at the moment when participants needed it, leads to more accurate and detailed knowledge (Sallas et al., 2007). In the next section, we discuss the implications of ACT-R for language learning.

SAT and implications for language learning

In the field of SLA, many researchers view L2 language development, especially in classroom settings, as being similar to any other form of cognitive skill acquisition (de Jong & Perfetti, 2011; DeKeyser 1997, 1998, 2001, 2007a,b; Lyster 1994, 2004, 2007; McLaughlin 1990; Ranta & Lyster 2007; Lyster & Sato, 2013). Learners in the L2 classroom typically start with declarative knowledge about L2 grammar and lexis, which through extensive practice is proceduralized and finally automatized. So the initial declarative knowledge is the explicit knowledge a learner has of a particular grammatical construction or vocabulary item. The procedural knowledge, on the other hand, will be the knowledge visible in a person's behavior when using the target language. During the stage of proceduralization, learners learn to rely less on their declarative knowledge and more on the production behavior they have developed through practice. Proceduralization, therefore, can be viewed as providing crutches to learners to ease the transition between completely relying on declarative rules to formulate a sentence in a L2, to using L2 grammar and lexis fairly automatically, without having to retrieve from memory any declarative rules. As a result, cognitive resources are freed up to attend to other information online.

To illustrate with a language-specific example, declarative knowledge can consist of knowing THAT if a verb in English is regular and it needs to be used in the past, then the ending *–ed* is added to the regular verb and it is pronounced as /d/ if the verb ends in a voiced sound, as /t/ if the verb ends in a voiceless sound, and as /ɪd/ if the verb ends either in *d* or *t*. Procedural knowledge, on the other hand, operates with production rules, which are behavioral rules that take the form of IF/THEN pairs:

IF the situation to be described occurred in the past,

And the verb describing the action is regular,

And the verb ends in a voiced sound,

THEN add *–ed* to the verb,

And pronounce it /d/.

As mentioned above, it is crucial for declarative knowledge to be available throughout the initial execution of the target behavior for proceduralization to take place. In instructional practice, however, very often grammar rules are presented prior to example learning or practice and are seldom revisited during the proceduralization phase, contrary to what is suggested by ACT-R.

It follows from the above that for proceduralization to take place, (a) learners should have full initial access to the declarative knowledge (the rule), (b) the rule should be comprehensible to the learner, and (c) the rule should be accessible throughout the stage of proceduralization. The further apart declarative knowledge, consisting of rules and examples, is from practice with further examples, the greater the memory decay of declarative knowledge, and the weaker the resulting procedural knowledge are expected to be. Precisely the issue of setting the right conditions for proceduralization is often overlooked in language teaching, as pointed out by DeKeyser (2007a).

In addition, according to SAT, practice in SLA should be skill-specific; once knowledge has been proceduralized in one skill, for instance comprehension, it becomes more difficult for that knowledge to be generalized in another skill, for example production. In other words, in order to develop receptive knowledge, learners need practice comprehending input, and in order to develop productive knowledge, learners need to practice producing language (Anderson, 1983; DeKeyser & Sokalski, 1996). Although both comprehension and production are indicators of general grammatical and lexical competence, they differ in a number of ways. The listener's task is to map the incoming sounds onto existing words in the mental lexicon, recognize the syntactic structure of the string, assign meaning to the whole utterance and finally incorporate the utterance within a broader context. In contrast, in a production activity the speaker first decides on the meaning that needs to be conveyed and only then selects the appropriate form, retrieves the words from the mental lexicon and utters them in the order required by the syntax.

Literature form first language acquisition suggests that comprehension and production of language do not develop in parallel. It is well established that comprehension precedes production (Clark, 1993 in Hendriks, 2013: 18; Hendriks, 2013). Comprehension and production are differentially affected by an individual's attentional capacities. Just and Carpenter (1992) claim that comprehension is the area most highly affected by attentional limits.

A great number of studies in the SLA field have explored the effects of comprehension-based instruction (CBI) and production-based instruction (PBI) with regards to L2 grammar (e.g., Benati, 2005; Cadierno, 1995; DeKeyser and Sokalski, 1996; Shintani & Ellis, 2013; Qin, 2008; VanPatten, 1996), and single-word items (e.g., Ellis and He, 1999; Shintani, 2011; 2013; Webb, 2005). However, mixed results have been obtained. Most studies investigating the effect of comprehension and production on the acquisition of L2 grammar come from the Processing Instruction (PI) literature (Benati, 2005; Cadierno, 1995; VanPatten, 1996; VanPatten & Cadierno, 1993). The earlier general finding from these studies seems to be that exposing learners to structured input, CB practice, is a necessary and sufficient condition for successful language acquisition and that providing explicit information, as well as production practice, have little or no effect in the learning process. However, as DeKeyser and Prieto Botana (2005) show in their review on PI studies, when studies include a production-based outcome measures, the picture is reversed: the production-based group either outperforms the PI or show an equivalent performance to the PI groups. For instance, DeKeyser & Sokalski (1996) endeavored to replicate one of the earliest PI studies done by VanPatten & Cadierno (1993) with several improvements. There were three treatment groups, input practice, output practice, and no practice group, and two linguistic targets, direct object clitics in Spanish and the conditional form of verbs in Spanish. Contrary to VanPatten & Cadierno's (1993) finding that output practice does not add any additional benefit for the production of direct object clitics, results from the immediate posttest by DeKeyser & Sokalski (1996) demonstrated that input practice was significantly better for comprehension tasks, while the output practice was significantly better for production tasks, a finding which is not in line with VanPatten & Cadierno's (1993) claim that output practice does not add additional benefit for the production of direct object clitic in Spanish.; this finding is in line with the directional asymmetry predicted by SAT, however. As highlighted in DeKeyser & Prieto Botana (2015), very often those differences disappear when tested with a delayed posttest. The authors also point out that in the more recent research on PI, the status of production-based practice has shifted from being redundant to beneficial when the right conditions for practicing are met. In other words, recent PI studies have demonstrated that the relative effectiveness of PI in comparison to production-based practice hinges on whether the production-based practice is communicative or not.

A number of other SLA studies have provided additional support for the applicability of ACT-R to SLA in general, and for skill-specificity of practice in particular (e.g., DeKeyser, 1997; de Jong, 2005; Rodgers, 2011). The main implication for L2 learning is that both types of knowledge representation should be fostered: "highly specific procedural knowledge, highly automatized for efficient use in the situations that the learner is most likely to confront in the immediate future, and solid abstract declarative knowledge that can be called upon to be integrated into much broader, more abstract, procedural rules" (DeKeyser, 2015).

As highlighted by DeKeyser (2015), very little research in the area of SLA has explicitly set out to investigate the applicability of skill acquisition theory to L2 learning. While several very important questions have been addressed so far, the issues of the timing, availability, and precise role of declarative knowledge, as well as the exact relationship between rules and examples that produces the most optimal learning have not been thoroughly investigated in SLA. Skill acquisition theory, as embodied in Anderson's ACT-R architecture, can motivate and inform predictions about L2 learning, especially about certain issues that still remain controversial and for which no clear empirical findings exist.

Skill acquisition and individual differences

Acquisition of complex cognitive skills, including language learning, is inevitably connected with learner-internal cognitive capacities that stem from differences in individuals' cognitive architecture. It is therefore important to take into account a theory of individual differences when studying the acquisition of complex skills (Taatgen, 2001). Any learning theory that postulates different acquisition stages and at the same time gives prominence to cognitive variables to account for individual differences in acquisition, will predict that different cognitive variables may play a substantial and differential role during different acquisition stages. According to Ackerman's theory (1988, p. 270, 1990), "during skill acquisition the load on cognitive processes declines from novice attention-demanding processing to skilled automatic processing. Initial ability-performance associations are higher for more general cognitive abilities", which decline with consistent practice. Ackerman (1988, 1990) demonstrated that general intelligence, speed of proceduralization and psycho-motor speed differentially influence specific stages of skill acquisition. In particular, it was demonstrated that during the first stage of skill acquisition, general intelligence played the biggest role because during this stage a heavy load is placed on the cognitive-attentional system: conscious attention was required to understand the task in question and for a successful completion of the task. As practice progressed, attentional demands were reduced and so was the impact of general intelligence. Furthermore, Ackerman's (1988) study demonstrated that perceptual speed ability and psychomotor ability characterized participants' performance during the second and third stage of acquisition of the skill that primarily depended on a motor behavior. More recently, Ackerman's learning theory has been extended and incorporated within a cognitive model of skill acquisition. Instead of correlating performance on different tasks, Taatgen (2001) incorporated cognitive variable parameters in the ACT-R model to be tuned in order to account for individual differences during skill acquisition. Since both Ackerman (1988, 1990) and Taatgen (2001) experimented with a task that is primarily motor-based, i.e., the Air Traffic Controller Task, it is reasonable to expect that the abovementioned cognitive variables will mediate learning during the three stages. It can also be expected that other cognitive variables may play a differential role in the acquisition of skills that are not primarily motor-based.

Similar observations and questions have been posed in the SLA field. Researchers such as Robinson (2005) and Skehan (1998) point out that exploratory empirical research is needed to look at a relationship between different cognitive variables, on one hand, and different stages of development, on the other hand. Unfortunately, since then, virtually no empirical studies have looked at this intricate relation. An exception is a recent SLA study by Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong (2014), who examined how differences in cognitive abilities among learners, namely declarative and procedural learning abilities, interact with L2 syntactic development. Participants were exposed to an artificial language under incidental training conditions and were engaged in extensive comprehension and production activities in the L2. Participants' syntactic development was assessed both at the early (after treatment one) and late (after treatment four) stages of acquisition. In line with Ackerman's (1988) theory, results demonstrated a differential involvement of specific cognitive variables during different stages of acquisition. In particular, results showed that while declarative learning ability was positively related and predicted performance during the early stages of acquisition, procedural learning ability was positively related and predictive of learners' performance during the late stages of skill acquisition. The question remains to be answered of whether this complicated relationship will be observed in a more intentional learning set-up as well.

While the field of SLA has witnessed an increased interest in research investigating factors that may account for the variability and relative success of adults' L2 attainment, research has mainly focused on how individual differences predict learning outcomes, as opposed to how individual differences impact the process of learning. Specifically, evidence from experimental studies on Aptitude-by-Treatment Interaction (ATI) demonstrates that the benefits of a particular instructional treatment are constrained by language learning aptitude (e.g., Brooks, Kempe, & Sionov, 2006; Erlam, 2005; Li, 2013; Perrachione, Lee, Ha, & Wong, 2011; Robinson, 1997; Sheen, 2007; Suzuki, 2013; Wesche, 1981) and working memory among other cognitive variables (e.g., French & O'Brien, 2008; Goo, 2012; Juffs, 2004; Juffs & Harrington, 2011; Kormos & Safar, 2008; Linck et al. 2013; Mackey, Adams, Stafford & Winke, 2010; Martin & Ellis, 2012; O'Brien, Segalowitz, Collentine & Freed , 2006; Révész, 2012; Williams, 1999; Williams & Lovatt, 2003). One reason for mainly investigating the relationship between cognitive variables and learning outcomes is the nature of current language aptitude tests. Aptitude tests such as the MLAT are developmentally insensitive (Robinson, 2005, 2013). In other words, they do not measure abilities that contribute to learning over time, and are mainly targeting explicit, problem-solving abilities, such as explicit inductive learning, rote memory, and analytical ability (Granena, 2013, in press).

In addition, although not a new concept outside the field of SLA (see Reber, 1989, 1993; Snow, 1991), recent theories of cognitive aptitudes for language learning have called in for a more multifaceted view of language aptitude, proposing that individuals may have high ability in one aptitude component and low ability in other aptitude components (e.g., DeKeyser & Koeth, 2011; Granena, 2013, in press; Linck et al., 2013; Robinson, 2005; Skehan, 2012). For instance, it has been suggested that explicit and implicit learning are two distinct learning processes that both have separate cognitive variables associated with them (Granena, 2013b; Kaufman, DeYoung, Gray, Jiménez, Brown, & Mackintosh, 2010; Linck et al., 2013).

Finally, of particular interest to the present research is a study by Erlam (2005) who investigated the relationship between language aptitude, working memory and timing of GE. Erlam (2005) explored the role of language analytic ability and working memory in relation to the learning outcomes of three instructional treatments: deductive, inductive and structured input. Results demonstrated that there was no relationship between the individual differences measures and learning outcomes in the deductive treatment. This finding, the author concluded, may point to an equalizing effect of the deductive treatment that benefits all individuals in the language production activities. In other words, the presence of explicit GE prior to practice benefited all participants regardless of the type of instruction they received. Results also demonstrated that learners with higher language analytic ability and working memory benefited more from the structured input treatment, when producing the target structure. While both the deductive and the structured input groups received explicit GE prior to practice, in the structured input treatment participants were not required to produce the target structure, but were rather focused on input-based activities. It seems then that explicit GE without the traditional production activities cannot level out the effects of aptitude. The third main finding was that better language-analytic ability predicted better performance on the listening comprehension and written production test in the inductive treatment. Participants in this group were not given explicit GE, but were engaged in activities that facilitated explicit hypothesis testing regarding the target structure. Similarly to the previous finding, when an instructional treatment does not involve both explicit GE and production activities, the learning process draws more on language aptitude and working memory.

Based on the results obtained from Erlam (2005), it can be predicted that both language aptitude and working memory will mediate the effectiveness of a treatment that is primarily comprehension-based, regardless of the presence or absence of explicit GE. In other words, it can be predicted that learners with higher language aptitude and working memory across experimental groups will benefit more from the treatment than those individuals with lower language aptitude and working memory in the same group.

To summarize, what is clear from the above is that learning research should incorporate a multifaceted view of individual differences to account for a range of language learning cognitive abilities that may play a differential role at early versus late stages of acquisition. As outlined above, studies to date have mainly focused on exploring a single component of language learning aptitude in relation to a learning outcome, even though evidence exists to suggest that (1) language learning aptitude is not a unitary construct, but rather a composite of different aptitude components, (2) different aptitude components may play a differential role at different stages of language learning.

Purpose of the study

This dissertation research was motivated by several gaps in the literature on the role of explicit instruction in SLA. Very few studies in SLA have tested the theoretical claims of one of the most influential theories of learning, skill acquisition theory. The dissertation applies it to SLA by exploring the role of timing of explicit information. First, virtually no studies have looked at the precise role of grammar rules by isolating the effects of pre- and during-practice GE. Second, the present study will look at how the timing of explicit GE impacts the development of declarative knowledge and its proceduralization in early stages of language acquisition. Third, the study will attempt to reach a certain level of generalizability by including two L2 structures that place different processing demands on the learners: one emphasizing syntactic operations and the other relying on a semantic contrast not present in the L1. Finally, the present study will take into account the impact of cognitive variables such as explicit and implicit language learning aptitude and working memory on the effectiveness of the various treatments in the present study, and will explore their involvement both in the learning process and learning outcome, rather than concentrating merely on the learning outcome. To summarize, by accounting for different combinations of pre- and during-practice GE, qualitatively different L2 structures, as well as cognitive variables, the present study will be able to provide a better understanding of the precise role of explicit grammar rules in the process of L2 learning.

Chapter 3: Research Questions and Hypotheses

RQ1: What is the exact timing of GE with respect to practice that provides the optimal basis for the formation of solid declarative knowledge?

H1: Following SAT, it is hypothesized that the closer together rule and practice are, the better the learning of the L2 target structure. More specifically, it is hypothesized:

H1a: GEb+GEd+ > GEb- GEd+

H1b: GEb- GEd+ > GEb+ GEd-

H1c: GEb+ GEd- > GEb- GEd-

The first hypothesis will be addressed through the analysis of a metalinguistic knowledge test after session 1.

RQ2: Will the newly formed declarative knowledge be equally accessible for the different skills of comprehension and production?

H2: SAT assumes that declarative knowledge is flexible, in contrast to proceduralized knowledge, and can be used in the execution of any skill. It is hypothesized that after the first learning phase, learners will not have proceduralized their declarative knowledge and will be able to use this knowledge equally well for both comprehension and production activities. This prediction is based on the assumption and empirical observation in SAT that declarative knowledge is flexible and as such can be used in the execution of any skill in any direction (Anderson, 1993; Anderson et al., 2004; Anderson & Fincham, 1994; Anderson, Fincham, & Douglass, 1997).

The second hypothesis will be addressed through the analysis of the comprehension and production test after session 2, with both accuracy and reaction time data. First, the means of these measures will be descriptively compared, and then a comparison will be made of how each of these skills evolves over time.

RQ3: What is the exact timing of GE with regards to practice that provides the optimal basis for the transition from declarative to proceduralized knowledge?

H3: Following SAT, it is hypothesized that the closer together rule and practice are, the better the learning and retention of the L2 target structure. More specifically, it is hypothesized:

H1a: GEb+GEd+ > GEb- GEd+

H1b: GEb- GEd+ > GEb+ GEd-

H1c: GEb+ GEd- > GEb- GEd-

This question will be addressed through the analysis of both accuracy and reaction time data from the comprehension and production tests after the last session.

RQ4: What is the role of individual differences during the skill acquisition process?

H4a: It is predicted that language learning aptitude associated with explicit learning will play a bigger role in the initial stages of skill acquisition, in particular in the formation of declarative knowledge.

H4b: It is predicted that language learning aptitude associated with implicit learning will play a bigger role at the later stages of skill acquisition, in particular during the stage of final proceduralization.

H4c: It is predicted that general working memory capacity will play a bigger role at the initial stages of skill acquisition, in particular in the initial stage of proceduralization. In particular, the effect of working memory will be most visible in the GEd- conditions.

H4d: It is predicted that explicit language aptitude will mediate learning in the conditions in which rules are not provided concurrently with practice, i.e., GEd- conditions.

H4e: It is predicted that the effect of explicit language aptitude will be most visible in the group that will not receive any GE, i.e., GEb- GEd-.

Hypotheses 4a-d are based on theoretical as well as empirical findings demonstrating that there are different types of language learning aptitudes that are involved in different types of learning, primarily explicit and implicit (Granena, 2013, Morgan-Short et al., 2014). In addition, research on the determinants of individual differences during skill acquisition has also pointed to the same direction, namely that different abilities are involved at the earlier versus later stages of skill acquisition. Cognitive abilities, such as general intelligence, are more involved during the learning processes underlying stage one of skill acquisition, whereas perceptual speed ability as well as psychomotor ability play a bigger role during the last two stages (Ackerman, 1988). Finally, working memory has been an inherent part of ACT-R model, suggesting the importance of working memory in the acquisition of any cognitive skill (Daily et al., 1998; Rehling et al., 2003; Taatgen, 2001). These hypotheses will first be addressed by looking at the correlations between the IDs and the outcome measures. As a second step, simple linear regressions will be performed to look at the predictive power of LLAMA, WM and SRT on the outcome measures.

Chapter 4: Method

Participants

A total of 128 native speakers of English agreed to participate in the present study. Participants were beginner learners of Spanish enrolled in beginners' first-semester SPAN at the University of Maryland during the Fall 2015 semester. Twelve participants were excluded from the study because they either dropped the course or attended only the first or last session. The final sample consisted of 116 participants with average age of 21.69, minimum age 18 and maximum age 57, 51 males and 65 females. This population was chosen on the assumption that participants' knowledge of the two structures would be non-existent or at best minimal. The researcher introduced the study during the first week of the semester and collected the consent forms from those students who agreed to participate. The study took place during the third, fourth, fifth and sixth week of the semester. This was purposefully done so that (1) students had at least some minimal knowledge of Spanish necessary to process the practice materials, and (2) the practice of the two structures preceded the regular instruction students receive on these two structures. Since the study was a part of their regular course, participants received attendance credit for participating in the study. Those who did not want to participate were only physically present during the lab sessions and consequently were not penalized for not participating in the study.

Linguistic targets: OV structure and Ser/Estar

The L2 structures chosen for the present study were (1) Object Verb Subject (OV) sentences with direct object pronouns, and (2) *ser/estar* distinction in Spanish. The criteria for the choice of the target structures were both theoretical and practical. First, for both structures, literature exists that points to developmental stages of acquisition. The canonical word order in Spanish is SVO, but due to its rich morphology, Spanish exhibits relatively flexible word order such as SOV and OVS. Because of the relatively free word order in Spanish, subjects do not necessarily appear in sentence-initial position and hence word order is not a reliable cue for assigning subject and object roles within a sentence. In addition, being a null-subject language, subjects in Spanish can be omitted from a sentence. As a result, agreement provides native speakers of Spanish and L2 learners with cues for successful processing. Relevant to our target structure are the verbal agreement cues. In Spanish, verbs are marked for person and number (e.g., “pintó” vs. “pintaron”; ‘painted.3sg’ vs. ‘painted.3pl’) and as such verb morphology has been claimed to be the main cue for subject identification (MacWhinney, 2001). However, in the case where both the subject and the object are third-person singular, and the subject is also omitted, the verb morphology cue is not helpful and therefore the interpretation of sentences like “Lo fotografía la chica” is less obvious to non-native speakers. It has been documented that this OVS structure is particularly difficult for English native speakers who, relying on the consistent word order cue in English, tend to interpret the first noun phrases in SVO and OV(S) sentences as the subjects. (Bates et al., 1984; Bever, 1970; Gass, 1989; VanPatten & Oikkenon, 1996). For instance, in a sentence such as “Lo fotografía la chica” (Him photographs the girl), the first noun phrase, “lo” is interpreted as the subject of the sentence, thus creating the meaning of “He photographs the girl”. We will refer to this structure as OV rather than OVS, because, as explained above, an overt subject is not always present.

In the acquisition of *ser/estar*, learners begin with omitting the copula verb altogether and then overgeneralize *ser* (see Table 2). A specific source of difficulty, especially for English NS, is that both verbs *ser* and *estar* map onto one single verb in English, namely the verb “to be”. Although there are several different uses of *ser* and *estar*, the present study will only explore the last acquisition stage as highlighted by VanPatten (1987, Table 1), whereby both *ser* and *estar* are possible candidates in sentences containing adjectives of conditions, but the choice of one over the other will lead to a difference in meaning. As mentioned in Prieto Botana (2013), in these instances, instead of across-the-board rule application as is the case with direct object OV sentences, learners need to assess cases individually and make judgments regarding the nature of the predicates and the adjectives in terms of "inherent and circumstantial" and permanent and temporary" characteristics. For instance, the adjective *nervioso,* can be used to describe someone who is inherently nervous, in which case the verb *ser* will be used ("Es nervioso"), or to describe someone who became nervous under certain circumstances, but does not usually exhibit that characteristic ("Está nervioso").

Table 1 *Order of acquisition of SER and ESTAR*

VanPatten (1987)	
1.	Absence of copula
2.	Use of ser for most copula functions
3.	Appearance of estar with progressive
4.	Appearance of estar with locatives
5.	Appearance of estar with adjectives of conditions

Furthermore, whereas a number of SLA studies have examined the acquisition of OV by L2 learners, very few studies have examined the acquisition of ser/estar, in particular when adjectives of conditions are part of the predicate. Acknowledging the fundamental differences between these two structures, and the importance of taking into consideration structures that pose different acquisition problems to L2 learners, Prieto Botana (2013) was the first study to experiment with both of these structures within the PI framework. By including the same two structures, easier comparability across studies will be allowed.

Design

In order to test the specific hypotheses mentioned above, the present study manipulated the variables of GE before practice (GEb +/-), or during practice (GEd+/-) to create four experimental groups illustrated in Table 2 below:

Table 2 *Experimental design of the present study*

Group	**Treatment**
Group 1	GEb+ GEd+
Group 2	GEb+ GEd-
Group 3	GEb- GEd+
Group 4	GEb- GEd-

The four experimental groups differed in the timing as well as the frequency of GE. Group 1 received GE both prior and during practice. Group 2 only received GE prior to practice and was not allowed to consult any grammar rules during practice. Group 3 only received GE during practice, and Group 4 did not receive any GE. Immediate feedback, in the form of correct/ incorrect was provided for participants in all groups.

Materials and instruments

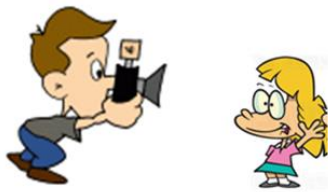
Materials for the present study were taken from Prieto Botana (2013). Materials consisted of explicit GE for both structures (GE) and both practice tasks. GE consisted of a couple of pages where the rules concerning the target structures were presented and furnished with examples. In order to ensure that participants paid attention to the explicit information, after each GE slide comprehension questions were included (see Appendix A and B).

Practice

Practice consisted of picture-matching and sentence-interpretation comprehension activities for each of the target structures. Practice was task-essential in the sense that, unless participants understood the direct object pronoun, or *ser* and *estar*, they would not have been able to choose the correct illustration of the sentence. Participants were allowed as much time as they needed to go through the practice activities. For the OV materials, participants had to pay attention to and understand the position of the direct object pronoun in order to decide whether the object pronoun in sentence-initial position refers to the doer or to the recipient of the action. For example, the following target sentence (1) was presented visually on a computer screen together with two pictures. Participants were asked to choose the picture that best illustrated the target sentence, by pressing either the left or right SHIFT button (See Appendix C for sample training items for Groups GEd+ and Appendix D for sample training items for Groups GEd-.)

(1) Lo fotografía la chica (fotografiar: to take a picture of)

a.



b.

Practice items from the sentence-interpretation activity were essentially the same as the picture-matching task, with the only difference that instead of the two pictures, there were two interpretations of the Spanish target sentence given in English (2).

(2) La asusta el gato (Asustar: to scare)

a. The cat scares her　　　　b. She scares the cat

Similarly, for the *ser/estar* distinction, the choice of one over the other picture depended on successful processing of the target verb. When presented with a picture-matching item, participants saw a sentence like the one presented in (3) and two corresponding pictures (a and b), or a sentence like (4) and two corresponding interpretation sentences in English (a and b)

(3) Está diferente (diferente: different)

a.

b.

(4) Manuel es tímido (tímido: shy)

a. He doesn't like to meet new people

b. He doesn't know anyone in this party

In order to avoid vocabulary comprehension issues, practice materials consisted of cognates and high-frequency lexical items. In addition, a gloss was provided for all of the verbs in the OV and all of the adjectives in the *ser/estar* practice and testing sessions. Finally, practice consisted of 40 picture-matching and 60 sentence-interpretation items for both structures.

Tests

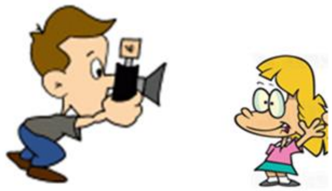
Comprehension

The comprehension assessment consisted of two comprehension tests. The first one mirrored the picture-matching practice activities. Participants were presented with a target sentence and were asked to choose one of the two offered pictures by pressing either the right or left SHIFT button. There were 20 new target sentences, not encountered during training for OV and 20 new sentences for *ser/estar*. One point was awarded for choosing the right picture, otherwise zero.

In order to test participants' comprehension of the target structure independent of the task they had practiced the structures with, a second comprehension test was administered. The second comprehension test was a form of a grammaticality judgment task (GJT), in which participants were presented visually with a target sentence in Spanish and presented with a corresponding picture. Upon reading the sentence and looking at the picture, participants were asked to judge the grammaticality of the sentence relative to the presented picture. In other words, given the pictorial depiction of an action, was the written sentence correct or not? An example is provided in (5). The GJT employed in this study departs from the typical administration of a grammaticality judgment task in that usually there are no pictures following a target sentence; participants are only required to judge the grammaticality of a sentence in the target language. In addition, for the items marked as incorrect, participants were asked to provide, i.e. type, the reasons for the picture-sentence mismatch. The last point was used as a metalinguistic knowledge test.

(5) Does this sentence accurately describe the situation in the picture?

Lo fotografía la chica (fotografiar: to take a picture)



Production

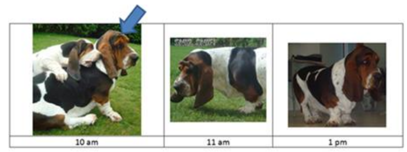
The production test consisted of one picture followed by an incomplete sentence. Participants were instructed to complete the gap with either the direct objects *lo* and *la* (6), or with the appropriate form of the verbs *ser* and *estar*, i.e. *es* and *está* (7) by typing the correct form.

(6) Fill in the gap with *lo* or *la*

__________ filma. (Filmar: to film)

(7) Fill in the gap with *es* or *está*

_________ tranquilo. (Tranquilo: calm, laid back)



Metalinguistic knowledge test (MTK)

The metalinguistic knowledge test was part of the GJT. Besides being required to indicate the accuracy of the sentence-picture pair, participants were asked to state what was the exact rule that made the sentence-picture pair accurate or inaccurate. There was no time pressure for any of the three tests.

LLAMA F

A subset of the LLAMA aptitude test (Meara, 2005), LLAMA F, was used as a measure of explicit language-analytic ability. In this task, participants saw a set of pictures and sentences describing the pictures, and tried to work out the grammar rules that operate in the language, and the priming task to measure how long it takes an individual to make the decision whether a string of letters forms a word or not.

OSPAN

Participants performed the operation span (OSPAN) working memory task (Turner and Engle, 1989; Unsworth et al., 2005). In this version of the OSPAN, a simple two-part mathematical equation is followed by a possible answer and an English letter in capital letters (e.g., (2 x 1) + 2

= 4 J). Participants were instructed to read the equation and the possible answer, decide whether the answer was correct or incorrect and finally read the letter aloud. At semi-random intervals, participants were prompted to recall all the words encountered since the last recall cue. There were 12 recall sets ranging from two to five letters. Participants were awarded one point for every letter remembered, yielding a total score of 40. The span for each individual was calculated as the average number of words remembered out of the total number of words (40). For an individual's WM span to be included in the analysis, s/he would have to achieve at least 80% accuracy on the math problems.

Implicit sequence learning ability: SRT task

The probabilistic SRT task was adopted from Kaufman et al. (2010). In the SRT task, participants saw a stimulus appearing at one of four locations on the computer screen and had to respond by pressing the corresponding key. Unknown to participants, the sequence of stimuli was actually generated by a probabilistic rule, and 85% of the sequences followed the rule (probable, training condition), but the other 15% of sequences were generated by another rule (improbable, the control condition). More specifically, sequence A (1–2–1–4–3–2–4–1–3–4–2–3) occurred with a probability of 0.85 and Sequence B (3–2–3–4–1–2–4–3–1–4–2–1) occurred with a probability of 0.15 in one block. This probabilistic nature of the SRT task made it difficult to learn the sequence explicitly. It is noted that these sequences are comprised entirely of second-order conditionals, so they could not be determined by first-order conditionals (Reed & Johnson, 1994); a second-order conditional sequence is determined by the previous two locations, not by the previous location, which makes the task more complex and implicit and minimizes chunk learning. There were eight blocks, and each block consisted of 120 trials, with 960 trials in total.

Procedure

Participants attended four sessions to complete the treatment of the present study. All practice and testing materials were computer-delivered through the presentation software DMDX in a computer lab on the University of Maryland campus. During the first session, participants completed the pretest, testing their knowledge of the two structures, a background questionnaire and individual differences measures (LLAMA, WM, SRT). The tests for the pretest were in the same format and order as the three posttests: production, comprehension, GJT. Following the first session, participants were randomly assigned to one of the four experimental conditions. During the second session participants started with the appropriate treatment. All participants first started with practice for the OV structure, followed immediately by the three tests for the OV structure. Then all participants proceeded with the SER practice followed by the three tests for this structure. During the practice treatment, participants in the GEb+ groups started by reading two slides of GE regarding the target rule, and answered comprehension questions to ensure that they had processed and understood the explicit information, and only then proceeded to practice. The practice treatment consisted of 40 picture-matching items and 60 sentence-interpretation items. The third session was identical to the second one. During the delayed posttest, participants were only administered the posttest outcome measures: production, comprehension and GJT. While the format of the tests was kept the same across the four sessions, the test items were different. The time elapsed between each of the four sessions was one week. Below is a table summarizing the experimental sessions (Table 3).

Table 3 *Overview of experimental sessions*

SESSION 1			
Pretests: Prod[1], Comp[2], GJT,			
ID: LLAMA, OSPAN,			
Background questionnaire			
SESSION 2			
Phase 1 Learning			**Posttest 1**
1.GEb+ GEd+	GE	Practice + GE	Prod, Comp, GJT
2. GEb+ GEd-	GE	Practice	Prod, Comp, GJT
3. GEb- GEd+		Practice + GE	Prod, Comp, GJT
4. GEb- GEd-		Practice	Prod, Comp, GJT
SESSION 3			
Phase 2 Learning			**Posttest 2**
1.GEb+ GEd+	GE	Practice + GE	Prod, Comp, GJT
2. GEb+ GEd-	GE	Practice	Prod, Comp, GJT
3. GEb- GEd+		Practice + GE	Prod, Comp, GJT
4. GEb- GEd-		Practice	Prod, Comp, GJT
SESSION 4			
Delayed Posttests: Prod, Comp, GJT			

Data analysis

The results from the current study are mainly based on accuracy. It was initially proposed that reaction times would also be analyzed in order to look at the effect of proceduralization on participants' reaction times, but due to several reasons the reaction time data for all of the measures was not analyzed. First, approximately one third of the students complained that moving from the comprehension measure that required the pressing of either the left or right shift, to the production measure that required typing, they were confused and it took them longer to respond. Secondly, it

[1] Prod= Test one Production
[2] Comp= Test one Comprehension

only makes sense to analyze reaction time data for measures that require immediate yes/no, correct/incorrect response. In the case of production, the reaction times may not be an accurate measure of the true reaction time. In addition, typing speed/ skill is also included in the reaction times. Therefore, only the reaction time data for the comprehension measure will be analyzed.

Before the main analyses the data were screened for outliers by standardizing the variables and coding all of the scores that were above or below three standard deviations from the mean as missing values. Because of elimination of outliers the N sizes of the groups differed for each of the outcome measures. All of the variables were then plotted and transformed if needed. The skewness and kurtosis values for the variables were within the range of normality, ranging from -2 to +2. The data was then fed into the imputation software NORM and SPSS. Based on the data augmentation values, the data was imputed from parameters by rounding to the nearest observed values. The imputed values were also examined to make sure they were in the range of possible values for the specific variables. Table 4 shows the summary results from NORM comparing means and standard deviations of the original and the imputed datasets.

Table 4 *Summary table with missing percentages per variable, Means and SD for original and imputed datasets*

	Number missing	%missing	Means OD	SD OD	Means IMP	SD IMP
LLAMA	8	6.84	45.04	29.07	45.22	28.87
SRT	26	22.22	2.5	1.6	2.52	1.7
WM	38	32.48	0.86	0.19	0.86	0.18
CovsT1	4	3.42	0.21	0.15	0.21	0.15
CovsT2	4	3.42	0.75	0.24	0.79	0.18
CovsT3	14	11.97	0.82	0.18	0.83	0.16
CovsT4	23	19.66	0.71	0.23	0.71	0.23
CserT1	5	4.27	0.61	0.14	0.61	0.14
CserT2	4	3.42	0.77	0.14	0.78	0.15
CserT3	15	12.82	0.75	0.14	0.75	0.14
CserT4	23	19.66	0.81	0.20	0.82	0.17
PovsT1	4	3.42	0.38	0.15	0.36	0.13
PovsT2	3	2.56	0.57	0.31	0.57	0.32
PovsT3	14	11.97	0.65	0.31	0.64	0.31
PovsT4	21	17.95	0.52	0.21	0.52	0.21

PserT1	5	4.27	0.53	0.06	0.54	0.07
PserT2	16	13.658	0.71	0.17	0.73	0.15
PserT3	27	23.08	0.77	0.12	0.78	0.12
PserT4	25	21.37	0.73	0.19	0.75	0.16
GovsT1	7	6.03	0.28	0.16	0.29	0.16
GovsT2	4	3.42	0.75	0.25	0.76	0.25
GovsT3	20	17.2	0.82	.19	0.81	0.19
GovsT4	26	22.2	0.77	0.21	0.76	0.22
GserT1	6	5.17	0.44	0.14	0.43	0.15
GserT2	8	6.84	0.71	0.15	0.71	0.16
GserT3	17	14.7	0.80	0.13	0.79	0.13
GserT4	26	22.2	0.75	0.16	0.74	0.16
GJT broken down by grammaticality: OV						
UNGovsT1	7	6%	.19	.14	.19	.15
GRovsT1	7	6	.38	.29	.39	.27
UNGovsT2	4	3.42	.72	.29	.71	.29
GRovsT2	4	3.42	.78	.25	.78	.24
UNGovsT3	20	17.2	.77	.24	.74	.24
GRovsT3	20	17.2	.88	.16	.86	.16
UNGovsT4	25	20.1	.67	.29	.64	.29
GRovsT4	25	21.1	.83	.21	.81	.22
GJT broken down by grammaticality: SER						
UNGserT1	6	5.17	.22	.25	.22	.25
GRserT1	6	5.17	.63	.20	.63	.20
UNGserT2	8	6.84	.68	.29	.67	.28
GRserT2	8	6.84	.73	.16	.74	.16
UNGserT3	17	14.7	.77	.17	.76	.17
GRserT3	17	14.7	.82	.14	.82	.13
UNGserT4	26	22.2	.69	.23	.67	.23
GRserT4	26	22.2	.79	.16	.78	.16

Note: OD = original data, IMP = imputed data

Reliability analysis using Cronbach's alpha was also conducted to inspect the reliability of the measures. Table 5 reports Cronbach's alpha for each variable. Although, in general, all of the outcome measures have acceptable reliability of .7 and above (Nunnally & Bernstein, 1994), six of our measures had reliability smaller than .70, mainly for the SER structure. With regards to the reliability of the individual differences measures, the WM OSPAN measure had a moderate reliability, within the lower range of the reported Cronbach's alpha values for OSPAN in previous studies (.70-.90, e.g., Conway et al., 2005). The SRT measure had a low reliability of .49, which is typical for reaction times measures in general, and for the SRT measure in particular (Kaufman et al., 2010). Finally, there was no way to compute reliability statistics for the language aptitude as measured by the LLAMA.

Table 5 *Cronbach's alpha for the outcome measures*

Outcome measure	Cronbach's Alpha
CovsT1	.85
CovsT2	.92
CovsT3	.89
CovsT4	.9
CserT1	.65
CserT2	.78
CserT3	.69
CserT4	.89
PovT1	.78
PovT2	.95
PovT3	.96
PovT4	.83
PserT1	.85
PserT2	.83
PserT3	.66
PserT4	.84
GovsT1	.64
GovsT2	.88
GovsT3	.88
GovsT4	.92
GserT1	.6
GserT2	.6
GserT3	.73
GserT4	.72
Individual Differences	
LLAMA	/
WM OSPAN	.72
SRT	.49

Chapter 5: Results

To address the first research question, about the impact of the different GE timings on the formation of declarative knowledge, univariate ANOVA with Group as the between-subject factor and the MTK test after session two as the dependent variable was conducted for the two structures separately. In addition, separate ANOVAs were also run with accuracy on the ungrammatical items on the GJT as the dependent variable.

The second research question aims to address indirectly the specificity of practice observed in numerous skill acquisition studies. Although direct comparisons between the comprehension and production measures are not possible, we can draw inferences based on the descriptive statistics across groups and times as well as any significant change over time and above-chance performance.

To address the third and main research question about the impact of the different GE timings on the most optimal transition from declarative to procedural knowledge, several ANOVA and ANCOVA models were tested. As a first step, repeated-measures ANOVAs were run with Time as a within-subject factor, Group as a between-subject factor and no covariates. The results from these models, looking at the effect of Group when no covariates are taken into account, demonstrate that there are no group differences. The results from these analyses can be found in Appendix E. As a second step, the covariates, LLAMA, WM and SRT, were added into the model one at a time. Adding the covariates in the models will not only provide a control for individual differences among participants, but will also indicate whether any interaction exists between the independent variables and the covariates. The ANCOVA models had a main effect for Time, Group, a two-way interaction term for Group and one of the covariates, and a three-way interaction between Time, Group and one of the covariates. The ANCOVA models will be the main analyses to answer the research questions and only these results will be reported in the main body of the dissertation. Of particular interest in ANCOVA is the assumption of homogeneity of regression slopes, which is important for the correct interpretation of the ANCOVA results (e.g. Huitema, 2011). When this assumption is violated, as evidenced by a significant two-way interaction between Group and one of the covariates, or a significant three-way interaction between Time, Group and one of the Covariates, the heterogeneous regression slopes may lead to misleading results and thus an alternative approach is recommended, namely the Johnson-Neyman procedure (Huitema, 2011). In particular, when heterogeneous regression slopes are present, the magnitude of the treatment effect is not the same at different levels of the covariate (Huitema, 2011:249). If ANCOVA evaluates the treatments effects at the covariate mean, the Johnson-Neyman procedure evaluates the treatment effects as a function of the level of the covariate. The Johnson-Neyman technique identifies regions of the covariate for which significant differences between the groups exist. In such cases, group differences will be interpreted from the Johnson-Neyman regions of significance values together with the interaction plots for each of the four times separately. As highlighted by Huitema (2011), a possibility exists that, even though the regression slopes are not homogeneous, there are no subjects for which differences are significant, i.e., there is no region of significance within the range of the sample data. In such cases, pairwise comparisons of group differences, if any, are reported.

Moreover, in the cases where the sphericity assumption was violated, that is, where Mauchly’s test was significant, the degrees of freedom were corrected using either the Greenhouse-Geisser or the Huynh-Feldt estimates. When the estimate of the sphericity, indicated by the epsilon value, was smaller than 0.75, the Greenhouse-Geisser correction was used. In cases where the epsilon value was larger than 0.75, the Huyhn-Feldt correction was used. The results from the final models will be interpreted, and the predictions for the research questions will be discussed in terms of two questions. The first one is regarding between-groups differences due to the differential treatment the four experimental groups received. The second question is related to within-group development, or how each group performed over time regarding knowledge retention.

Finally, the fourth research question aims to determine the nature of the relationship between the individual differences measures and the different experimental treatments. There are multiple results that can be gathered to answer this question. First, the results from the ANCOVA models together with the J-N regions of significance will be relevant for the research question at hand. In addition, where there were significant correlations between the outcome measures and one of the covariates, regardless of any interactions with the grouping variable, separate simple linear regression models were also conducted to look into the predictive power of the covariates for the outcome measures.

Univariate ANOVA

MTK and grammatical and ungrammatical GJT items

Univariate ANOVAs were conducted with Group as a between-subject factor and MTK after session two as well as accuracy on the ungrammatical GJT items as a dependent variable. The MTK test was part of the GJT. Participants were told that whenever they indicated that a sentence item is ungrammatical or incorrect, they should provide explanations in English with regards to the reasons for the mismatch between the picture and the sentence. After data transfer and coding, it was observed that there was a large portion of missing data. There are two possibilities why some participants did not provide any explanations for the incorrect sentences. First, participants may have lacked the conscious knowledge to articulate the reasons for ungrammaticality. Second, participants may have been aware of the reasons for ungrammaticality, but nevertheless may have decided not to provide any explanations. The large amount of missing data can be more problematic for the inferential repeated-measures analyses that exclude participants altogether even if they only missed one session. However, for the ANOVA analysis only after session 2, the missing data did not have such an impact: there were at least 17 participants in the groups. The means and standard deviations for the MTK test are presented in the Tables 6 and 7 below. It can be observed that while Group 3 [GEb-GEd+] has higher means than all other groups for both structures, Group 4 [GEb-GEd-] has the lowest means.

Table 6 *Unadjusted Means and Standard Deviations for MTK OV*

Group	T1	T2	T3	T4
GEb+GEd+ (N=21)	.13 (.24)	.70 (.28)	.73 (.31)	.74 (.37)
GEb+GEd- (N=18)	.02 (.04)	.71 (.23)	.75 (.24)	.74 (.32)
GEb-GEd+ (N=19)	.06 (.11)	.80 (.18)	.83 (.17)	.81 (.29)
Geb-GEd-	.07 (.17)	.63 (.27)	.61 (.28)	.68 (.28)

(N=17)

Table 7 *Unadjusted Means and Standard Deviations for MTK SER*

Group	T1	T2	T3	T4
GEb+GEd+ (N=24)	.06 (.11)	.69 (.24)	.72 (.24)	.71 (.25)
GEb+GEd- (N=22)	.08 (.21)	.61 (.23)	.64 (.20)	.65 (.24)
GEb-GEd+ (N=21)	.06 (.11)	.85 (.12)	.90 (.10)	.81 (.15)
GEb-GEd- (N=17)	.03 (.08)	.53 (.28)	.5 (.28)	.52 (.29)

The GJT data was broken down by grammatical and ungrammatical items. Besides serving to make inferences about participants' declarative knowledge from the ungrammatical items, the GJT may be informative by showing how participants in different groups responded to the grammatical versus ungrammatical items. Tables 8-11 show the means and standard deviations for both structures. It can be seen that the means for the ungrammatical items are consistently lower than for the grammatical items across both structures; this bias towards accepting GJT stimuli is often found. While at Time 2 the difference between the grammatical and ungrammatical items is smaller, at Time 3 and 4 it gets larger.

Table 8 *Unadjusted Means and Standard Deviations for GJT Ungrammatical items OV*

Group	T1	T2	T3	T4
GEb+GEd+ (N=30)	.22 (14)	.72 (.27)	.72 (.27)	.64 (.28)
GEb+GEd- (N=28)	.18 (.11)	.81 (.24)	.81 (.15)	.70 (.27)
GEb-GEd+ (N=30)	.17 (.16)	.69 (.33)	.80 (.2)	.70 (.27)
GEb-GEd- (N=28)	.17 (.16)	.64 (.31)	.61 (.29)	.52 (.3)

Table 9 *Unadjusted Means and Standard Deviations for GJT Ungrammatical items SER*

Group	T1	T2	T3	T4
GEb+GEd+ (N=30)	.21 (.21)	.74 (.19)	.75 (.17)	.75 (.13)
GEb+GEd- (N=28)	.23 (.28)	.74 (.27)	.79 (.14)	.72 (.19)
GEb-GEd+ (N=30)	.26 (.29)	.68 (.32)	.80 (.16)	.70 (.24)
GEb-GEd- (N=28)	.16 (.20)	.54 (.30)	.69 (.19)	.53 (.28)

Table 10 *Unadjusted Means and Standard Deviations for GJT Grammatical items OV*

Group	T1	T2	T3	T4
GEb+GEd+ (N=30)	.46 (.3)	.74 (.26)	.82 (.16)	.83 (.22)
GEb+GEd- (N=28)	.33 (.31)	.85 (.2)	.93 (.13)	.81 (.2)
GEb-GEd+ (N=30)	.39 (.25)	.80 (.23)	.87 (.14)	.84 (.2)
GEb-GEd- (N=28)	.37 (.2)	.73 (.26)	.85 (.2)	.74 (.22)

Table 11 *Unadjusted Means and Standard Deviations for GJT Grammatical items SER*

Group	T1	T2	T3	T4
GEb+GEd+ (N=30)	.60 (.2)	.71 (.17)	.87 (.11)	.78 (.14)
GEb+GEd- (N=28)	.64 (.19)	.75 (.1)	.78 (.14)	.74 (.15)
GEb-GEd+ (N=30)	.61 (.18)	.78 (.22)	.83 (.12)	.82 (.15)
GEb-GEd- (N=28)	.66 (.23)	.71 (.12)	.78 (.14)	.78 (.17)

In what follows the correlation matrices between the IDs and the GJT items are presented for both structures (in Tables 12-15).

Table 12 *Pearson's correlations between ID and GJT for Group 1 [GEb+GEd+]*

	LLAMA	SRT	WM
OV			
UNGT1	.168	-.038	-.069
GRT1	.123	-.031	.19
UNGT2		-.1	.217
GRT2	.363*	-.059	.306
UNGT3	.045	-.079	.167
GRT3	.391*	-.014	.393*
UNGT4	.247	.058	.389*
GRT4	.012	.017	.456**
SER			
UNGT1	-.079	.13	.174
GRT1	-.19	-.16	.026
UNGT2	.141	-.307	.312
GRT2	.071	-.043	.166
UNGT3	.263	.07	.41*
GRT3	.119	.335	.155
UNGT4	.14	.103	.537**
GRT4	.215	.175	.444*

*Note *$p<.05$, **$p<.01$

Table 13 *Pearson's correlations between ID and GJT for Group 2 [GEb+GEd-]*

	LLAMA	SRT	WM
OV			
UNGT1	-.194	-.006	-.253
GRT1	-.418	.214	-.498*
UNGT2	.087	.214	.292
GRT2	.048	.065	.357
UNGT3	.1	.232	.374*
GRT3	.28	.058	.565**
UNGT4	.265	.184	.277
GRT4	.05	.105	.155
SER			
UNGT1	-.08	.299	.065
GRT1	-.206	-.203	-.048
UNGT2	.267	.137	.591**
GRT2	.136	.045	-.243
UNGT3	-.081	.069	.288
GRT3	.173	.060	.272
UNGT4	.281	-.01	.302
GRT4	-.016	.199	.299

*Note **p*<.05, ***p*<.01.

Table 14 *Pearson's correlations between ID and GJT for Group 3 [GEb-GEd+]*

	LLAMA	SRT	WM
OV			
UNGT1	.089	-.307	.253
GRT1	-.046	-.092	.1
UNGT2	-.03	-.057	.199
GRT2	-.197	-.027	.098
UNGT3	.03	.019	.282
GRT3	.21	-.28	-.01
UNGT4	.109	-.109	.227
GRT4	.029	-.09	.107
SER			
UNGT1	-.153	-.12	.162
GRT1	-.013	.338	-.005
UNGT2	.085	.01	-.023
GRT2	.285	-.165	-.095
UNGT3	-.126	.122	.323
GRT3	-.205	.102	.366*
UNGT4	.102	-.162	.196
GRT4	-.018	-.087	.198

*Note **p*<.05, ***p*<.01

Table 15 *Pearson's correlations between ID and GJT for Group 4 [GEb-GEd-]*

	LLAMA	SRT	WM
OV			
UNGT1	-.205	.084	.206
GRT1	-.24	-.174	-.228
UNGT2	.473*	.1	.238
GRT2	.404*	-.109	.062
UNGT3	.438*	.192	.254
GRT3	.149	.212	-.038
UNGT4	.285	.053	.02
GRT4	.258	-.246	.084
SER			
UNGT1	.108	.167	.156
GRT1	-.013	-.213	.096
UNGT2	.434*	-.07	-.095
GRT2	-.236	.291	-.188
UNGT3	-.104	.213	.008
GRT3	-.015	.058	-.039
UNGT4	.03	.19	.027
GRT4	-.072	.117	-.058

*Note **p*<.05, ***p*<.01.

From the correlation matrices we can observe that LLAMA seems to have the strongest relationship with performance of Group 4 [GEb-GEd-], where it significantly correlated mainly with the performance of the ungrammatical items during Time 2 and 3 for the OV structure, and during Time 2 only for the SER structure. In addition, while LLAMA did not have a consistent effect in the performance of the other groups, WM significantly correlated with performance of Group 1 [GEb+GEd+] during the later stages of acquisition, Time 3 and 4, for both grammatical and ungrammatical items. For Group 2 [GEb+GEd-], WM significantly correlated with performance on both OV grammatical and ungrammatical items during Time 3, and only on the SER ungrammatical items during Time 2.

ANOVA models

Results for the MTK of the OV structure demonstrated that there was no significant main effect of Group, $F(3,71)=1.98$, $p=.12$, partial $\eta^2=.08$. Results for the SER structure did demonstrate a significant main effect of Group, $F(3,80)=11.17$, $p<.001$, partial $\eta^2=.29$. Post-hoc Bonferroni tests revealed that Group 3 [GEb-GEd+] was significantly better than Group 1 [GEb+GEd+], MD=.16, $p=.05$, Group 2 [GEb+GEd-], MD=.24, $p<.01$, and Group 4 [GEb-GEd-], MD=.32, p<.001.

Results from the ANOVA models with accuracy on the ungrammatical GJT sentences for OV demonstrated that there was no significant main effect of Group, $F(3,112)=1.57$, $p=.2$, partial $\eta^2=.04$. However, similar to the results from the MTK test, for the SER structure, the ANOVA model demonstrated a significant main effect of Group, $F(3,112)=3.09$, $p=.03$, partial $\eta^2=.07$. Post-hoc Bonferroni tests revealed only that Group 1 [GEb+GEd+] was significantly better than Group 4 [GEb-GEd-], MD=.19, $p=.05$.

Models with covariates (LLAMA, WM, SRT)

The choice of the covariates for the individual ANOVA models was based on the strength of the correlation between the individual difference measures and the outcome measures. In cases where more than one individual measure significantly correlated with the outcome measure, separate repeated-measures ANCOVA models were run. The correlation matrices between the outcome measures and the individual difference measures for each group are presented below (Tables 16-19). The correlation matrices were also used to run separate simple linear regression models that are reported at the end of the results' section. From the correlation matrices it can be seen that SRT has almost no significant relationship with the outcome measures, whereas LLAMA and WM do. As predicted, LLAMA had the strongest relationship with the outcome measures in the fourth group [GEb-GEd-] (Table 19), and a non-existent relationship in the third group [GEb-

GEd+] (Table 18). This aligns well with the Aptitude-by-Treatment interaction theory that would predict that the more the burden is on the learner, in this case the group did not receive any GE, the more important aptitude for explicit learning is. As for the third group, it seems that the treatment that the third group received, GE only during practice, evened out any differences in language aptitude. In addition, for the second group that didn't receive GE during practice but only before practice, WM seems to play a substantial role as seen by the significant correlations with almost all outcome measures across all four times (Table 17). Finally, while LLAMA significantly correlated with performance of Group 1 [GEb+GEd+] for OV during Time 3 and 4, and for SER only during Time 2, WM seems to have a stronger relationship than the other covariates, especially during times 3 and 4, as shown by all three measures (Table 16).

Due to the many correlations performed, a question arises with regards to Type I error, i.e., how many of the significant correlations at alpha smaller than .05 would appear by random chance. To address this potential problem, data was simulated using the Transform Variable and Generate random numbers in the range of 0 to 1 options in SPSS. Results demonstrated that for Group 1 [GEb+GEd+] there were no significant correlations, for Group 2 [GEb+GEd-] there were three significant correlations, for Group 3 [GEb-GEd+] two significant correlations, and for Group 4 [GEb-GEd-] three significant correlations. Given these results, we can say that the patterns observed in the original data are not likely to have appeared from random chance. Additional evidence for this claim comes from the simple linear regression models which demonstrate that in all of the cases where there was a significant correlation observed between the IDs and the outcome measures, those IDs were also significant predictors in the regression models.

Table 16 *Pearson's correlations between ID and outcome measures for Group 1 [GEb+GEd+]*

	LLAMA	SRT	WM
LLAMA			

SRT	.108		
WM	.194	-.044	
CovsT1	.097	-.127	.381*
CovsT2	.349	.051	.112
CovsT3	.458*	-.233	.483**
CovsT4	.367*	.098	.170
CserT1	.038	-.216	.372*
CserT2	.2	-.123	.337
CserT3	.153	-.082	.379*
CserT4	.2	.108	.665**
PovsT1	.068	-.066	-.282
PovsT2	.308	.201	.012
PovsT3	.134	-.014	.155
PovsT4	-.236	-.354	-.352
PserT1	-.131	-.071	-.058
PserT2	.369*	.133	.612**
PserT3	.212	.257	.095
PserT4	.19	.175	.629**
GovsT1	.232	-.027	.085
GovsT2	.425*	-.078	.287
GovsT3	.087	-.038	.121
GovsT4	.253	.022	.364*
GserT1	-.219	.168	.142
GserT2	.152	-.198	.349
GserT3	.311	.18	.439*
GserT4	.140	.239	.562**

*Note *$p<.05$, **$p<.01$

Table 17 *Pearson's correlations between ID and outcome measures for Group 2 [GEb+GEd-]*

	LLAMA	SRT	WM
LLAMA			
SRT	-.167		
WM	.389*	.267	
CovsT1	-.258	.123	-.394*
CovsT2	.102	.05	.401*
CovsT3	.444*	.025	.248
CovsT4	.734**	.089	.5**
CserT1	-.311	.078	.-.003
CserT2	.282	.165	.584**
CserT3	.286	.158	.548**
CserT4	.351	-.069	.535**
PovsT1	-.301	.306	.070
PovsT2	-.070	.570**	.394*
PovsT3	.143	.556**	.382*
PovsT4	.053	-.045	-.152
PserT1	.268	.264	.391
PserT2	.211	.265	.558**
PserT3	.312	.197	.615**
PserT4	.503**	-.019	.403*
GovsT1	-.420*	-.147	-.456
GovsT2	.069	.154	.348
GovsT3	.252	.112	.462*
GovsT4	.383*	.090	.401*
GserT1	-.220	.052	.042
GserT2	.320	.144	.429*
GserT3	.125	.022	.309
GserT4	.228	-.029	.334

*Note *p<.05, **p<.01

Table 18 *Pearson's correlations between ID and outcome measures for Group 3 [GEb-GEd+]*

	LLAMA	SRT	WM
LLAMA			
SRT	-.087		
WM	-.191	.276	
CovsT1	-.459	.054	.181
CovsT2	.070	-.247	-.163
CovsT3	.184	-.220	.125
CovsT4	.150	-.054	.243
CserT1	-.183	.116	.128
CserT2	.203	.140	.440*
CserT3	.236	.212	.524**
CserT4	.236	.117	.137
PovsT1	.098	-.097	-.006
PovsT2	.376*	-.004	-.039
PovsT3	.098	.171	.195
PovsT4	.094	-.02	-.109
PserT1	.083	-.177	.247
PserT2	.242	.108	.369*
PserT3	.047	.206	.405*
PserT4	.234	-.019	-.043
GovsT1	.017	-.214	.203
GovsT2	-.062	-.1	.143
GovsT3	.063	-.137	.073
GovsT4	.191	-.148	.155
GserT1	-.153	.114	.141
GserT2	.259	-.050	-.086
GserT3	-.13	.003	.296
GserT4	.011	-.180	.293
GRserT4	-.018	-.087	.198

*Note *p<.05, **p<.01

Table 19 *Pearson's correlations between ID and outcome measures for Group 4 [GEb-GEd-]*

	LLAMA	SRT	WM
LLAMA			
SRT	-.337		
WM	.064	.079	
CovsT1	-.261	-.178	-.115
CovsT2	.120	.218	-.108
CovsT3	-.101	.143	-.016
CovsT4	.687**	-.123	-.004
CserT1	-.119	-.132	.305
CserT2	.490**	-.026	.121
CserT3	.076	.359^	.259
CserT4	.338^	.356^	.355
PovsT1	-.079	-.308	-.069
PovsT2	.441	.156	-.080
PovsT3	.360	.191	.078
PovsT4	-.037	-.196	.149
PserT1	.397*	-.064	.195
PserT2	.570**	-.020	-.073
PserT3	.403*	.369^	.135
PserT4	.128	.281	.231
GovsT1	-.214	-.356	-.149
GovsT2	.465*	.002	.162
GovsT3	.445*	-.073	.180
GovsT4	.446*	-.024	.111
GserT1	.108	-.043	.150
GserT2	.387*	.040	-.048
GserT3	.137	-.075	.041
GserT4	-.048	.082	.027

*Note *$p<.05$, **$p<.01$.

In this section, we first outlined the rationale for choosing the covariates for the Repeated-Measures ANCOVA models and presented the correlation matrices between the IDs and the outcome measures. We are now proceeding to the data from the ANCOVA models and the Johnson-Neyman procedure for the OV and SER structure beginning with the comprehension, production and GJT measure.

OV

Comprehension

The means and standard deviations for the comprehension measure can be found in Table 20. The data from the comprehension measure of OV were subjected to Repeated-Measures ANCOVA with Time as a within-subject factor, Group as a between-subject factor and LLAMA as the only covariate. Results using the Huynh-Feldt correction for sphericity demonstrated a significant main effect of Time, $F(3,294)=69.05$, $p<.001$, partial $\eta^2=.41$, a non-significant main effect of Group, $F(3,98)=.27$, $p=.84$, partial $\eta^2=.008$, a significant interaction between Time and Group, $F(9,294)=2.31$, $p=.02$, partial $\eta^2=.07$, a significant interaction between Group and LLAMA, $F(4,98)=4.22$, $p=.003$, partial $\eta^2=.15$ and a significant three-way interaction between Time, Group and LLAMA, $F(12, 294)=5.77$, $p<.001$, partial $\eta^2=.19$. Following the main analyses, pairwise comparisons using the Bonferroni correction were also conducted to look at the difference between times across the four groups. The pairwise comparisons output the mean differences in percentage correct. The pairwise comparisons for Time are shown in Table 21 from which it can be seen that only Group 3 [GEb-GEd+] did not have a significant loss of knowledge from Time 3 to Time 4, suggesting a steady increase and retention of the acquired knowledge. For the other groups, the knowledge loss from Time 3 to Time 4 was significant.

Table 20 *Means and standard deviations for Comprehension OV*

Group	T1	T2	T3	T4
GEb+GEd+ (N=26)	.24 (.15)	.79 (.17)	.88 (.09)	.74 (.23)
GEb+GEd- (N=26)	.21 (.16)	.84 (.13)	.88 (.14)	.77 (.21)
GEb-GEd+ (N=28)	.25 (.17)	.76 (.21)	.78 (.17)	.69 (.19)
GEb-GEd- (N=26)	.17 (.12)	.80 (.16)	.81 (.17)	.68 (.22)

Table 21 *Significant mean differences across Time for Comprehension OV ANCOVA model with LLAMA*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.558***		.643***	-.129**		.515***
Group 2 GEb+GEd-	.608***		.627***	-.141**		.486***
Group 3 GEb-GEd+	.53***		.557***			.466***
Group 4 GEb-GEd-	.628***		.646***	-.139**		.507***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Since there was evidence that the regression slopes were heterogeneous, results from the J-N procedure will be interpreted instead of the ANCOVA pairwise comparisons for group differences. Results indicated that at Time 3 Groups 1 [GEb+GEd+] and 2 [GEb+GEd-] were significantly better than Group 4 [GEb-GEd-] for values of LLAMA ranging bigger than 46.2 (Figure 1), and at Time 4, Group 1 [GEb+GEd+] and 3 [GEb-GEd+] performed significantly better than Group 4 [GEb-GEd-] for those participants who had a value of LLAMA between 20.64 and 41.16 (Figure 2).

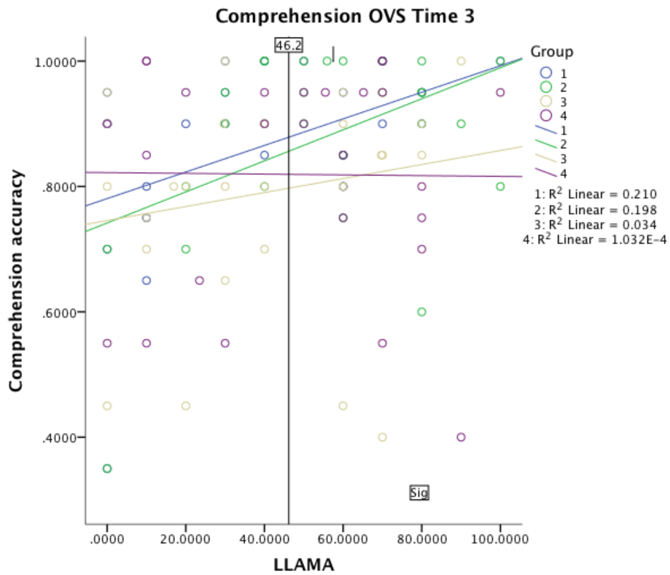


Figure 1 Group by LLAMA interaction plot for comprehension of OVS during Time 3

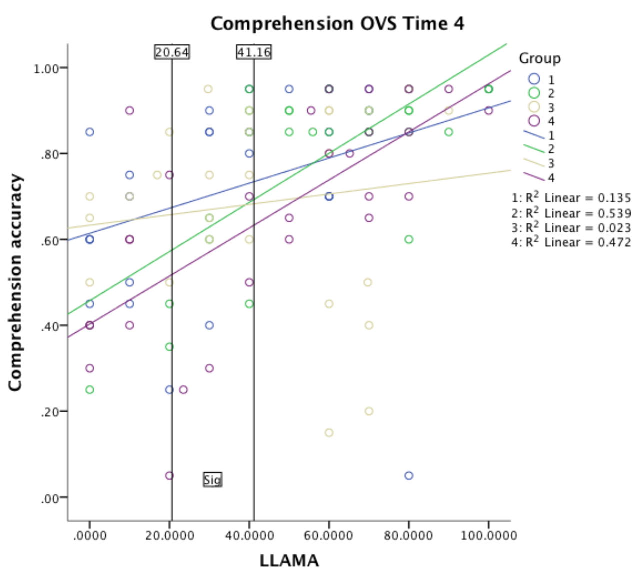


Figure 2 Group by LLAMA interaction plot for comprehension of OVS during Time 4

Production

For production as the outcome measure (see Table 22 for means and SD), the ANCOVA model with LLAMA as covariate, using the Greenhouse-Geisser correction for sphericity, demonstrated a significant main effect of Time, *F*(2.1, 219.3)=3.15, *p*=.04, partial η^2=.03, a main effect of Group approaching significance, *F*(3,102)=2.18, *p*=.09, partial η^2=.06, a significant interaction between Group and LLAMA, *F*(4,102)=2.5, *p*=.05, partial η^2=.09, a non-significant interaction between Time and Group, *F*(6.4, 219.3)= 1.01, p=.43, partial η^2=.03, and a non-significant three-way interaction between Time, Group and LLAMA, *F*(8.6,219.3)=1.3, *p*=.29, partial η^2=.05. Pairwise comparisons demonstrated that all groups' performance during Time 2, 3 and 4 was significantly better than Time 1. In addition, for Group 2 [GEb+GEd-] performance during Time 3 was better than Time 4 (MD=.258, *p*=.03). Table 23 shows all significant differences between times.

Table 22 *Means and Standard Deviations for Production OV*

Group	T1	T2	T3	T4
GEb+GEd+ (N=27)	.36 (11)	.53 (.30)	.64 (.28)	.54 (.20)
GEb+GEd- (N=26)	.38 (.12)	.64 (.32)	.74 (.26)	.47 (.20)
GEb-GEd+ (N=29)	.37 (.13)	.49 (.33)	.54 (.31)	.61 (.23)
GEb-GEd- (N=28)	.32 (.13)	.54 (.31)	.60 (.32)	.49 (.21)

Table 23 *Significant mean differences across Time for Production OV ANCOVA model with LLAMA*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.187*		.281***			.167**
Group 2 GEb+GEd-	.258***		.327***	-.26*		
Group 3 GEb-GEd+			.173*			.232***
Group 4 GEb-GEd-	.21**		.267***			.171**

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Although there was a significant interaction between Group and LLAMA, the Johnson-Neyman procedure indicated that there were no statistical significance transition points within the observed range of the covariate for Times 2 (Figure 3), 3 (Figure 4) and 4 (Figure 5). There were no significant group differences that were observed from the ANCOVA pairwise comparisons either.

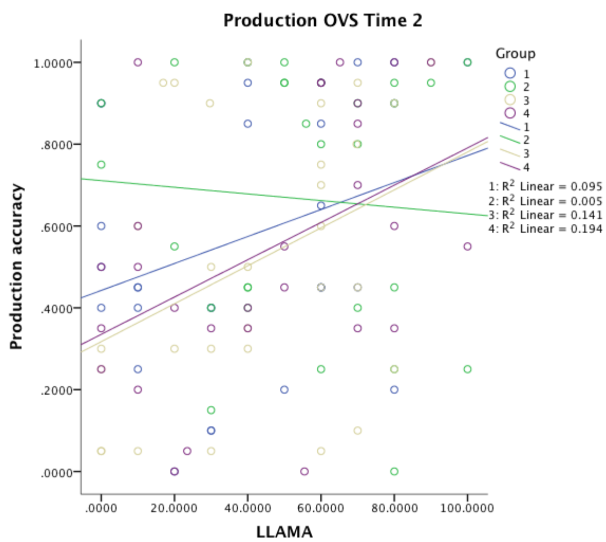


Figure 3 Group by LLAMA interaction plot for Production of OVS during Time 2

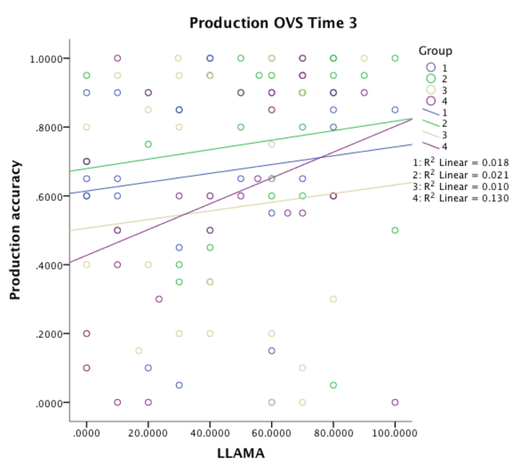


Figure 4 Group by LLAMA interaction plot for Production of OVS during Time 3

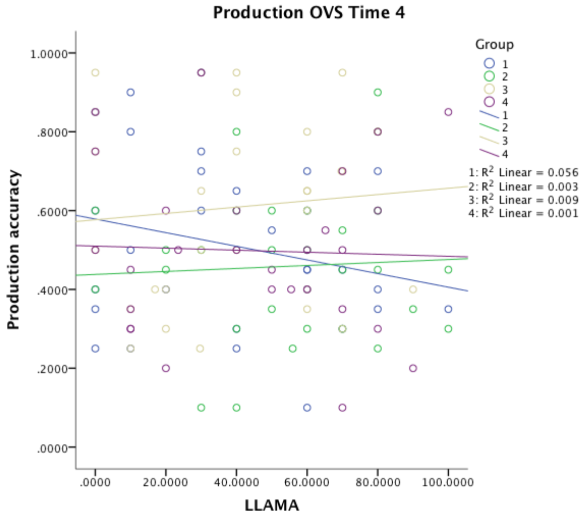


Figure 5 Group by LLAMA interaction plot for Production of OVS during Time 4

A second Repeated-Measures Model was run with SRT as the covariate. Results using the Greenhouse-Geisser correction for sphericity from this model demonstrated a significant main effect of Time, $F(2.1, 223,1)=4.7$, $p<.01$, partial $\eta^2=.04$, and a significant interaction between Group and SRT, $F(4,102)=3.13$, $p=.018$, partial $\eta^2=.11$. Pairwise comparisons for Time revealed the same mean differences as the ANCOVA model with LLAMA that can be seen in Table 24.

Table 24 *Significant mean differences across Time for Production OV ANCOVA model with SRT*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.179*		.277***			.172**
Group 2 GEb+GEd-	.273***		.362***	-.281**		
Group 3 GEb-GEd+			.15*			.229***
Group 4 GEb-GEd-	.224**		.281***			.171**

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Although there was a significant interaction between Group and SRT, the Johnson-Neyman procedure indicated that there were no statistical significance transition points within the observed range of the covariate for Times 2 (Figure 6), 3 (Figure 7) and 4 (Figure 8).

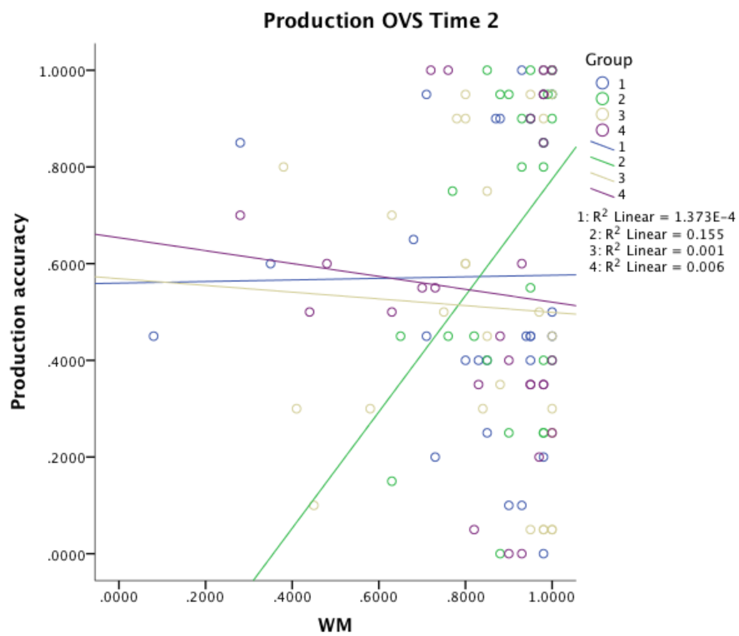


Figure 6 Group by WM interaction plot for Production of OVS during Time 2

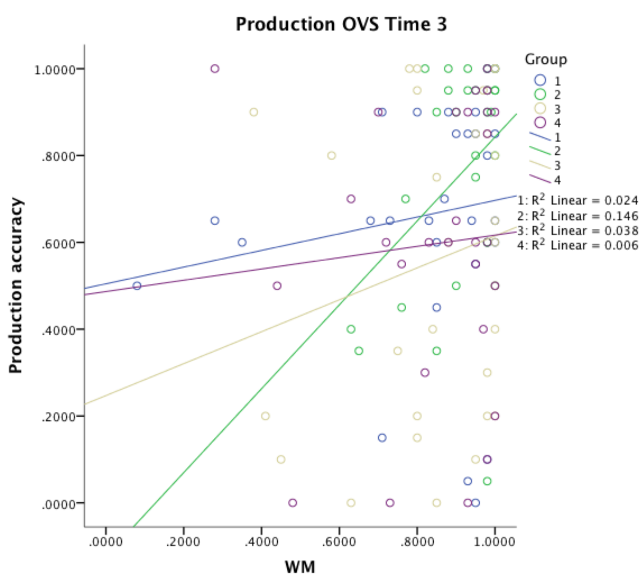


Figure 7 Group by WM interaction plot for Production of OVS during Time 3

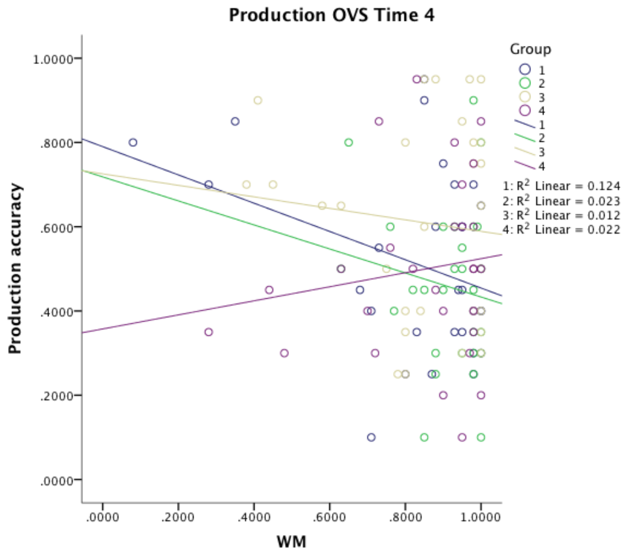


Figure 8 Group by WM interaction plot for Production of OVS during Time 4

GJT

For the GJT data, there were two ANCOVA models, with LLAMA and WM as the individual covariates. Means and SD are reported in Table 25. Results from the model with LLAMA as a covariate, using the Greenhouse-Geisser correction, demonstrated a significant main effect of Time, $F(2.2,8.5)=43.79$, $p<.001$, partial $\eta^2=.29$ a significant main effect of Group, $F(3,108)=3.6$, $p=.01$, partial $\eta^2=.09$, a significant interaction between Group and LLAMA, $F(4,108)=3.5$, $p=.01$, partial $\eta^2=.11$, and a significant interaction between Time, Group and LLAMA, $F(8.9,8.5)=2.6$, $p=.006$, partial $\eta^2=.09$. Although there was a significant interaction between Group and LLAMA, the Johnson-Neyman procedure indicated that there were no statistical significance transition points within the observed range of the covariate for Times 2 (Figure 9), 3 (Figure 10) and 4 (Figure 11), and therefore the results from the ANCOVA pairwise comparisons will be presented. Pairwise comparisons for group differences demonstrated that during the delayed posttest, Time 4, Group 3 [GEb-GEd+] was significantly better than Group 4 [GEb-GEd-], MD=.166, $p<.01$ (Figure 11). The significant mean differences for Time are shown in Table 26 from which it can be seen that all groups performed significantly better at Times 2, 3 and 4 than at Time 1. In addition, knowledge loss was also observed for Group 4 [GEb-GEd-].

Table 25 *Means and Standard Deviations for GJT OV*

Group	T1	T2	T3	T4
GEb+GEd+ (N=30)	.33 (.17)	.73 (.25)	.77 (.20)	.76 (.20)
GEb+GEd- (N=28)	.26 (.17)	.83 (.20)	.86 (.13)	.81 (.15)
GEb-GEd+ (N=30)	.27 (.17)	.77 (.25)	.85 (.15)	.83 (.19)
GEb-GEd- (N=28)	.25 (.11)	.68 (.26)	.77 (.24)	.66 (.23)

Table 26 *Significant mean differences across Time for GJT OV ANCOVA model with LLAMA*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.410***		.43***			.423***
Group 2 GEb+GEd-	.533***		.559***			.494***
Group 3 GEb-GEd+	.489***		.575***			.553***
Group 4 GEb-GEd-	.433***		.516***	-.105**		.411***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

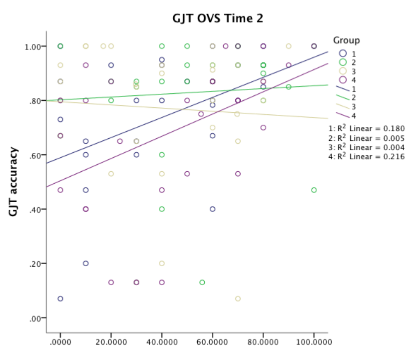


Figure 9 Group by LLAMA interaction plot for GJT of OVS during Time 2

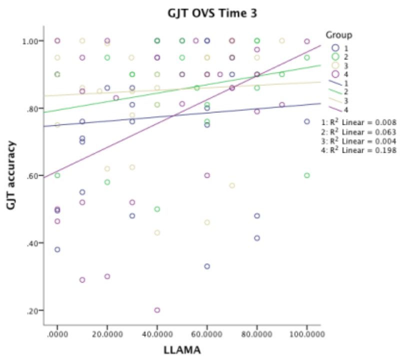


Figure 10 Group by LLAMA interaction plot for GJT of OVS during Time 3

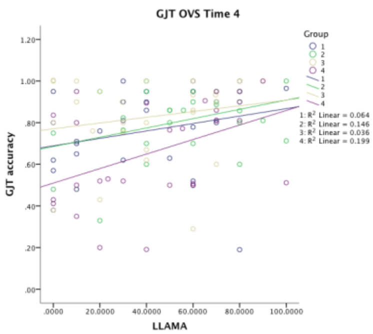


Figure 11 Group by LLAMA interaction plot for GJT of OVS during Time 4

The second model for the GJT data included WM as the covariate. The results demonstrated only a significant three-way interaction between Time, Group and WM, $F(9,243.9)=.1.8$, $p=.05$, partial $\eta^2=.06$. As was the case in the previous model, there were no statistical significance transition points within the observed range of the covariate for Times 2 (Figure 12), 3 (Figure 13) and 4 (Figure 14), and therefore the results from the ANCOVA pairwise comparisons will be presented. This model demonstrated a significant difference between Group 3 [GEb-GEd+] and 4 [GEb-GEd-] for the delayed posttest, MD=.161, $p=.02$. The same significant mean differences for time are observed as in the previous GJT model as seen from Table 27.

Table 27 *Significant mean differences across Time for GJT OV ANCOVA model with WM*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.404***		.44***			.427***
Group 2 GEb+GEd-	.49***		.527***			.471***
Group 3 GEb-GEd+	.492***		.573***			.548***
Group 4 GEb-GEd-	.437***		.52***	-.106**		.414***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

From Figures 12-14 we can additionally observe that the effect of WM is the strongest in Group 2 [GEb+GEd-].

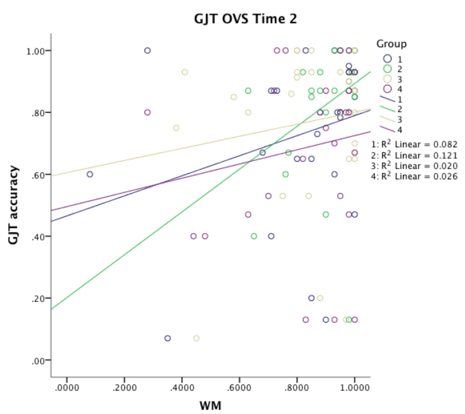


Figure 12 Group by WM interaction plot for GJT of OVS during Time 2

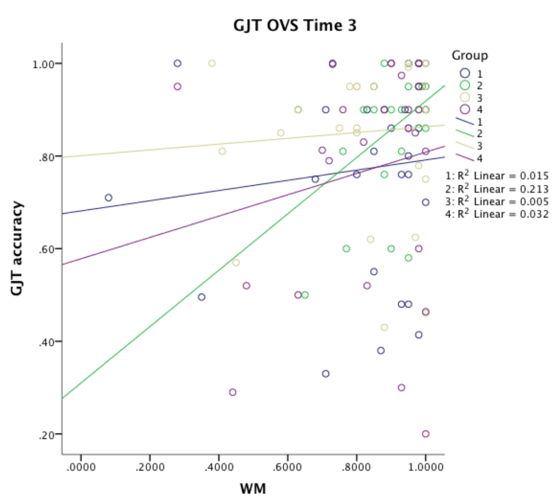


Figure 13 Group by WM interaction plot for GJT of OVS during Time 3

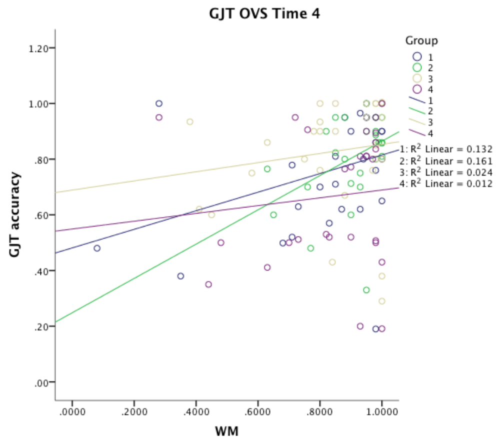


Figure 14 Group by WM interaction plot for GJT of OVS during Time 4

Summary for OV results

Table 28 summarizes the significant findings from the ANCOVA model with OV. The most important are the three-way interactions between Time, Group and one of the covariates which show that at certain time, and for certain values of the covariate there are group differences. In cases where there were either significant three-way interactions or significant two-way interactions between Group and one of the covariates, instead of the ANCOVA results the J-N ones were presented. It was noted that even though there may be a significant interaction that would indicate heterogeneity of regression slopes, there may not be significant J-N regions specifying group differences, if either the magnitude of the F ratio is small or the regions of significance fall outside the range of the observed values of the covariate (Huitema, 2011). Table 29 summarizes the main group differences for the OV structure. It can be seen that at Time 1 and 2 all groups had comparable knowledge of OV as evidenced by no group differences. During Time 3, J-N results from the comprehension measure suggested that Groups 1 [GEb+GEd+] and 2 [GEb+GEd-] significantly outperformed Group 4 [GEb-GEd-], and that during Time 4 the only difference that remained was between Groups 1 [GEb+GEd+] and 4 [GEb-GEd-]. In addition, the GJT measure revealed that Group 3 [GEb-GEd+] too significantly outperformed Group 4 [GEb-GEd-] during the delayed posttest.

In terms of within-group differences across time, based on the comprehension data for the OVS structure Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 4 [GEb-GEd-] had a significant decrease from Time 3 to Time 4. The production data additionally demonstrated that for Group 2 [GEb+GEd-] there was a significant decrease from Time 3 to Time 4, and the GJT data demonstrated that Group 4 [GEb-GEd-] had a significant decrease from Time 3 to Time 4. It appears that only Group 3 [GEb-GEd+] significantly retained the acquired knowledge based on all three outcome measures for the OV structure.

Table 28 *Summary table for Repeated-Measures ANCOVA models for the OV structure*

	Time	Time*Group	Time*Group*Cov	Group*Cov	Group
	Only significant F values are reported; *p<.05, **p<.01,***p<.001				
1. Comp OV with LLAMA	F(3,294)= 69.5***	F(9,294)= 2.31*	F(12, 294)=5.77***	F(4,98)=4.22**	
2. Prod OV with LLAMA	F(2.1, 219.3)=3.15*			F(4,102)=2.5*	
3. Prod OV with SRT	F(2.1, 223,1)=4.7**			F(4,102)=3.13*	
4. GJT OV with LLAMA	F(2.2,8.5)= 43.79***		F(8.9,8.5)= 2.6**	F(4,108)=3.5*	F(3,108)= 3.6**
5. GJT OV with WM			F(9,243.9)= 1.8*		

Table 29 *Summary table of significant group differences for the OV structure*

	Time* Group* Covariate	Group* Covariate	J-N	Pairwise comparisons
Comprehension LLAMA	√	√	(1) Time 3: Gr 1,2 >Gr4 for LLAMA 46.2 -100 (2) Time 4: Gr 1,3>Gr4 for LLAMA 20.64-41.16	
Production LLAMA		√		
Production SRT		√		
GJT LLAMA	√	√		Time 4: Gr3>Gr4
GJT WM	√	√		Time 4: Gr3>Gr4

SER

Comprehension

The data from the comprehension, production and GJT measures of the SER structure were subjected to repeated-measures ANCOVA. The means and SD can be found in Table 30. There were two ANCOVA models with the comprehension data, one with LLAMA and the other with WM as the covariate. Results from the ANCOVA model with LLAMA, using the Greenhouse-Geisser correction for the assumption of sphericity, demonstrated a significant main effect of Time, $F(2.3,\ 242.2)= 4.33$, $p<.01$, partial $\eta^2=.04$, a non-significant main effect of Group, $F(3,101)=1.002$, $p=.39$, partial $\eta^2=.03$, a significant three-way interaction between Time, Group and LLAMA, $F(9.5,\ 242.2)=1.84$, $p=.05$, partial $\eta^2=.07$ a non-significant interaction between Time and Group, $F(7.1,\ 242.23)=1.7$, $p=.1$, partial $\eta^2=.05$, and a non-significant interaction between Group and LLAMA, $F(4,101)=1.1$, $p=.34$, partial $\eta^2=.04$. The significant mean differences for Time are presented in Table 31 from which it can be seen that while all groups significantly improved during Times 2, 3 and 4 in comparison to Time 1, it was only Group 1

[GEb+GEd+] that also showed a significant increase in performance at Time 4 compared to Time 3.

Table 30 *Means and Standard Deviations for Comprehension SER*

Group	T1	T2	T3	T4
GEb+GEd+ (N=29)	.57 (.13)	.82 (.13)	.74 (.13)	.87 (.15)
GEb+GEd- (N=27)	.63 (.12)	.81 (.12)	.77 (.09)	.84 (.15)
GEb-GEd+ (N=26)	.62 (.15)	.80 (.16)	.80 (16)	.84 (.20)
GEb-GEd- (N=27)	.62 (.14)	.70 (.12)	.72 (.14)	.77 (.15)

Table 31 *Significant mean differences across Time for Comprehension SER ANCOVA model with LLAMA*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.247***	-.073*	.175***	.127***		.302***
Group 2 GEb+GEd-	.149***		.11*			.168**
Group 3 GEb-GEd+	.179***		.175***			.22***
Group 4 GEb-GEd-						.149**

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Since there is evidence for heterogeneous regression slopes, results from the J-N procedure will be interpreted instead of the ANCOVA pairwise comparisons for group differences. The J-N results demonstrated that at Time 2, Group 1 [GEb+GEd+] was superior to Group 4 [GEb-GEd-] only for those individuals who had a value of LLAMA ranging from 0 to 67.14 (Figure 15). In addition, at Time 4 this group difference remained for individuals with LLAMA values ranging from 11.27 to 54.57 (Figure 16).

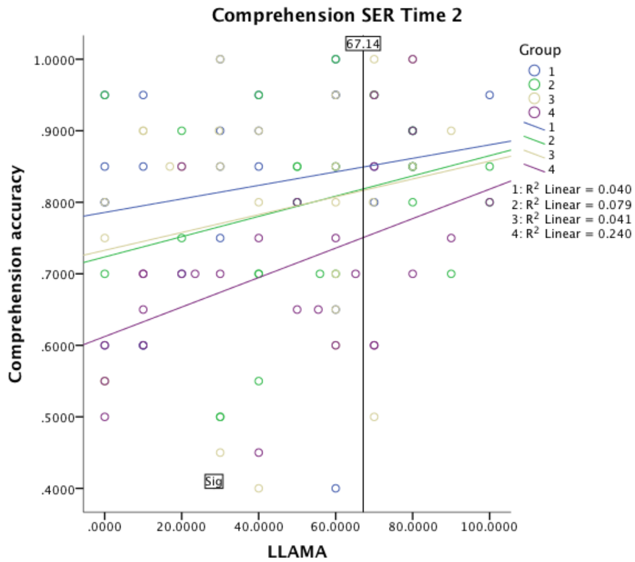


Figure 15 Group by LLAMA interaction plot for comprehension of SER during Time 2

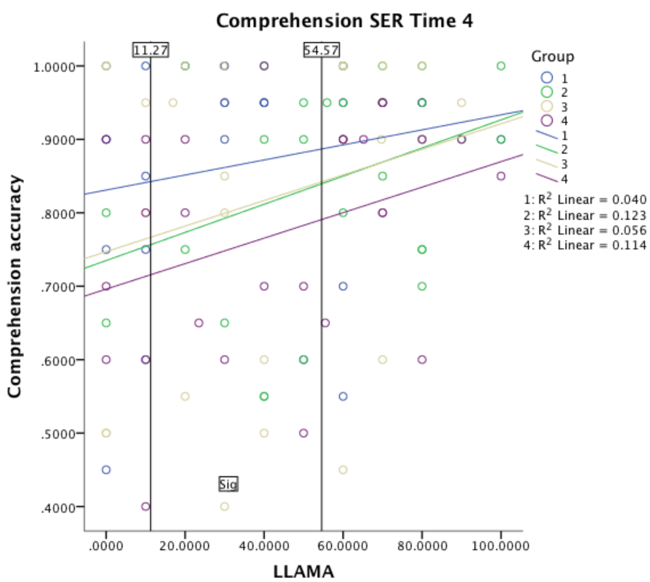


Figure 16 Group by LLAMA interaction plot for comprehension of SER during Time 4

The second model with WM as the covariate, using the Greenhouse-Geisser correction for sphericity, demonstrated a non-significant main effect of Time, $F(2.3, 236)=.37$, $p=.7$, partial $\eta^2=.02$, a non-significant main effect of Group, $F(3, 101)=1.04$, $p=.37$, partial $\eta^2=.03$, a significant interaction between Group and WM, $F(4, 101)=7.5$, $p<.001$, partial $\eta^2=.23$, a non-significant interaction between Time and Group, $F(7, 236)=1$, $p=.41$, partial $\eta^2=.03$, and a non-significant interaction between Time, Group and WM, $F(9.3, 236)=1.4$, $p=.183$, partial $\eta^2=.05$. Pairwise comparisons for Time revealed the same significant patterns as for the comprehension LLAMA model as seen in Table 32 reported below.

Table 32 *Significant mean differences across Time for Comprehension SER ANCOVA model with WM*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.244***	-.07*	.174***	.127***		.302***
Group 2 GEb+GEd-	.154***		.114*			.179**
Group 3 GEb-GEd+	.171***		.165***			.21***
Group 4 GEb-GEd-						.149**

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

In terms of group differences, even though there was a significant two-way interaction between Group and WM, there were no statistical significance transition points within the observed range of the covariate for Times 2 (Figure 17), 3 (Figure 18) and 4 (Figure 19), and therefore the results from the ANCOVA pairwise comparisons will be presented. Pairwise comparisons demonstrated that at Time 2, Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] were significantly better than Group 4 [GEb-GEd-], MD=.124, p<.01, MD=.097, p=.05, respectively. In addition, it was observed that at Time 4, Group 1 [GEb+GEd+] was significantly better than Group 4 [GEb-GEd-], MD=.103, p=.05 (Figure 19).

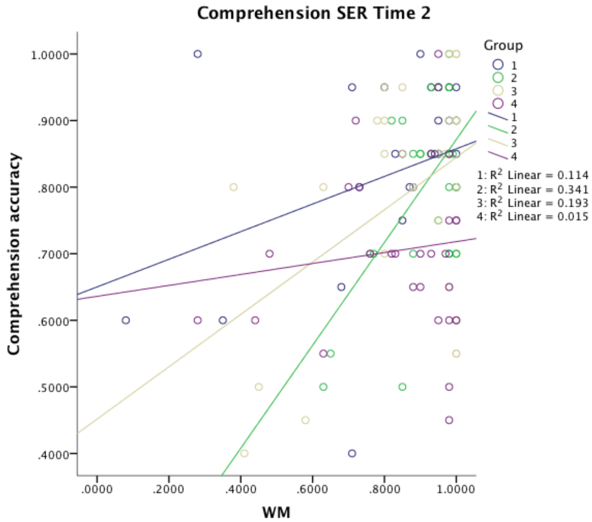


Figure 17 Group by WM interaction plot for Comprehension of SER during Time 2

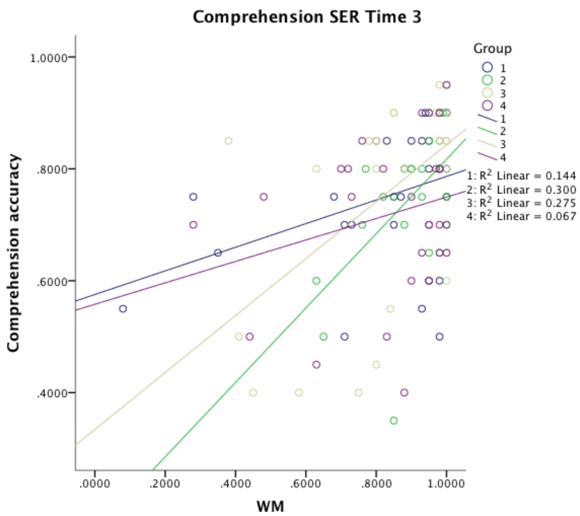


Figure 18 Group by WM interaction plot for Comprehension of SER during Time 3

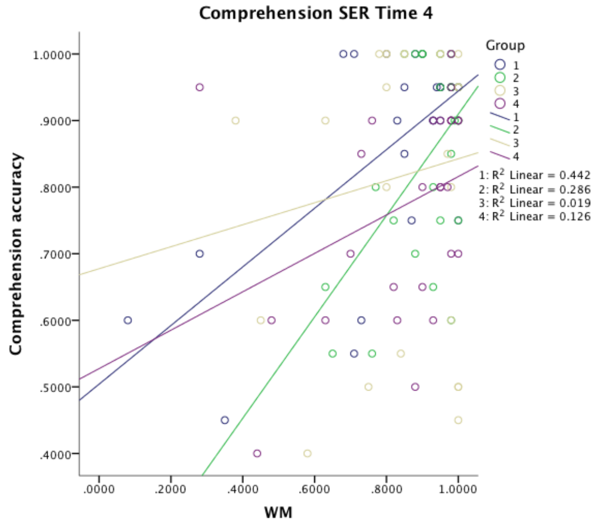


Figure 19 Group by WM interaction plot for Comprehension of SER during Time 4

Production

For the production data for SER, there was a model for each of the three covariates. The means and SD can be found in Table 33. The first model which included LLAMA as the covariate, using the Huynh-Feldt correction for sphericity, demonstrated a significant main effect for Time, *F*(2.9, 288.24)=28.9, *p*<.001, partial η^2=.23, a significant main effect of Group, *F*(3,99)=4.4, *p*<.01, partial η^2=.12 a significant interaction between Group and LLAMA, *F*(4,99)=4.3, *p*<.01, partial η^2=..15, an interaction between Time and Group approaching significance, *F*(8.7, 288.24)= 1.7, *p*=.08, partial η^2=.05 and a three-way interaction between Time, Group and LLAMA approaching significance, *F*(11.6, 288.24)=1.6, *p*=.07, partial η^2=.06. Table 34 shows the significant mean differences for Time, which point to a comparable and significant improvement of all groups at Times 2, 3 and 4 in comparison to Time 1.

Table 33 *Means and Standard Deviations for Production SER*

Group	T1	T2	T3	T4
GEb+GEd+ (N=29)	.55 (.06)	.78 (.13)	.81 (.09)	.80 (.12)
GEb+GEd- (N=23)	.57 (.07)	.73 (.08)	.77 (.09)	.75 (.15)
GEb-GEd+ (N=27)	.51 (.05)	.76 (.14)	.80 (.11)	.74 (.17)
GEb-GEd- (N=28)	.52 (.09)	.64 (.14)	.74 (.12)	.69 (.17)

Table 34 *Significant mean differences across Time for Production SER ANCOVA model with LLAMA*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.246***		.271***			.263***
Group 2 GEb+GEd-	.16***		.2***			.154***
Group 3 GEb-GEd+	.25***		.281***			.232***
Group 4 GEb-GEd-	.119***	.089**	.209***			.161***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Results from the J-N procedure demonstrated that for Time 2, the most obvious group difference is between Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], on one hand, and Group 4 [GEb-GEd-], on the other hand, for learners with language aptitude ranging from 0 to 68.4 (Figure 20). During Time 3 Group 1 [GEb+GEd+] is significantly better than both Group 2 [GEb+GEd-] and Group 4 [GEb-GEd-] for learners with language aptitude ranging from 0 to 53.83 (Figure 21). Finally, Figure 22 shows that Group 1 [GEb+GEd+] remained superior to Group 4 [GEb-GEd-] but only for learners with language aptitude between 14.6 and 76.4.

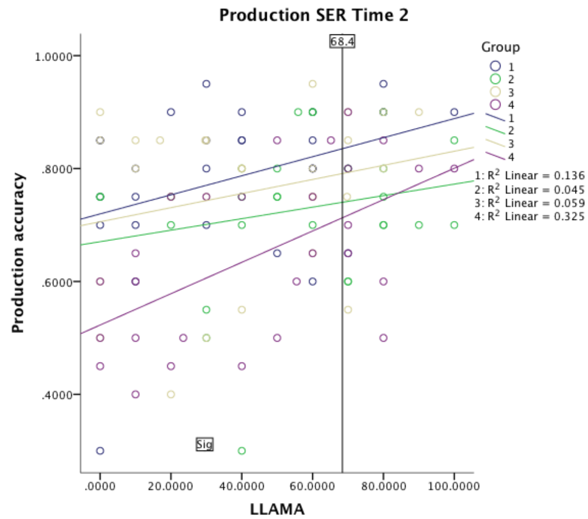


Figure 20 Group by LLAMA interaction plot for Production of SER during Time 2

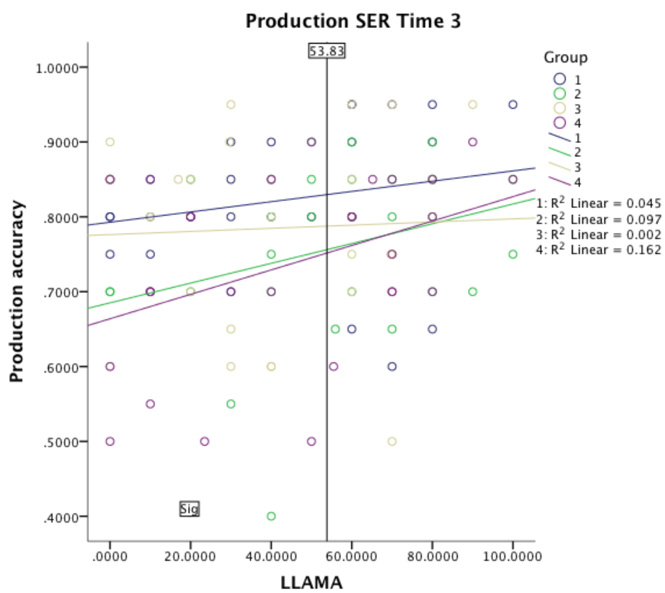


Figure 21 Group by LLAMA interaction plot for Production of SER during Time 3

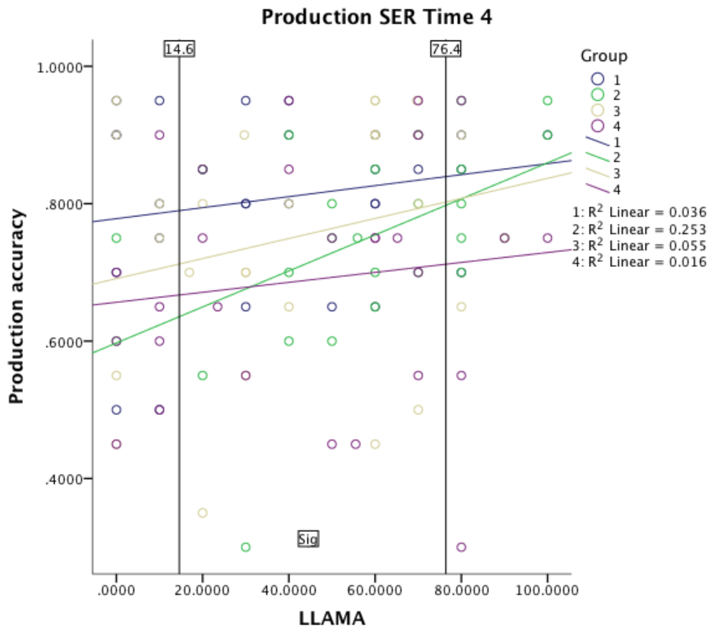


Figure 22 Group by LLAMA interaction plot for Production of SER during Time 4

The second model included SRT as the covariate. Using the Huynh-Feldt correction for sphericity, the model revealed a significant main effect of Time, *F*(2.8, 286.3)=29.4, *p*<.001, partial η^2=.5, a non-significant main effect for Group, *F*(3,99)=2.1, *p*=.1, partial η^2=.06, a non-significant interaction between Time and Group, *F*(8.6, 286.3)=.94, p=.40, partial η^2=.03, a non-significant interaction between Group and SRT, *F*(4,99)=.88, *p*=.47, partial η^2=.03, and a non-significant interaction between Time, Group and SRT, *F*(11.5, 286.3), *p*=.33, partial η^2=.04. Since evidence exists that the homogeneity of regression assumption has not been violated group differences using the pairwise comparisons from the ANCOVA model will be provided. Note that the group difference results from the ANCOVA model with SRT are the same as the one from the ANOVA model (Appendix E). During Time 2, Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] were significantly better than Group 4 [GEb-GEd-], MD=.138, *p*<.001, MD=.113, *p*=.01, respectively. During Times 3 and 4 only Group 1 [GEb+GEd+] remained significantly superior to Group 4

[GEb-GEd-], MD=.075, *p*=.05, MD=.114, *p*=.04, respectively. The significant time differences are presented in Table 35 below suggesting that all groups evolved over time to a comparable extent.

Table 35 *Significant mean differences across Time for Production SER ANCOVA model with SRT*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.236***		.268***			.259***
Group 2 GEb+GEd-	.154***		.2***			.176***
Group 3 GEb-GEd+	.243***		.275***			.228***
Group 4 GEb-GEd-	.12***	.094**	.214***			.166***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Finally, the third model with WM as the covariate, using the Huynh-Feldt correction for sphericity, demonstrated a non-significant main effect of Time, *F*(2.9, 287.5)=1.2, *p*=.3, partial η^2=.01, a non-significant main effect of Group, *F*(3,99)=.8, *p*=.5, partial η^2=.02, a significant interaction between Time and Group, *F*(8.7, 287.5), p=.05, partial η^2=.05, a significant interaction between Group and WM, *F*(4,99)=3.9, *p*=.005, partial η^2=.14, and a significant interaction between Time, Group and WM, *F*(11.6, 287.5), *p*=.003, partial η^2=.09. Table 36 below shows the significant time differences for each of the groups.

Table 36 *Significant mean differences across Time for Production SER ANCOVA model with WM*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.242***		.268***			.264***
Group 2 GEb+GEd-	.153***		.194***			.165***
Group 3 GEb-GEd+	.246***		.281***			.229***
Group 4 GEb-GEd-	.116***	.094**	.21***			.164***

Note: The mean difference (MD) will be reported only where significant differences appear.

Since there was a significant Group * Covariate interaction, the J-N results will be presented. During Time 3, as presented in Figure 24, Group 1 [GEb+GEd+] was significantly better than both Group 2 [GEb+GEd-] and Group 4 [GEb-GEd-] for individuals with WM between .59 and .91. Finally, during the delayed posttest, as presented in Figure 25, Group 1 [GEb+GEd+] is significantly better than Group 4 [GEb-GEd-] only for individuals with WM of .73-1. It should be noted that results from all three models and analyses with the production data point in the same direction, except the significant superiority of Group 1 [GEb+GEd+] to Group 2 [GEb+GEd-] during Time 3 for individuals with WM values between .59 and .91.

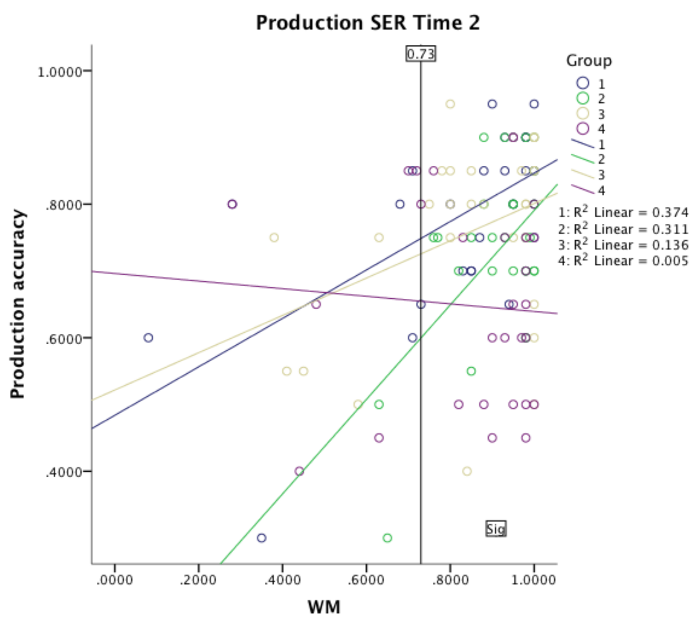


Figure 23 Group by WM interaction plot for Production of SER during Time 2

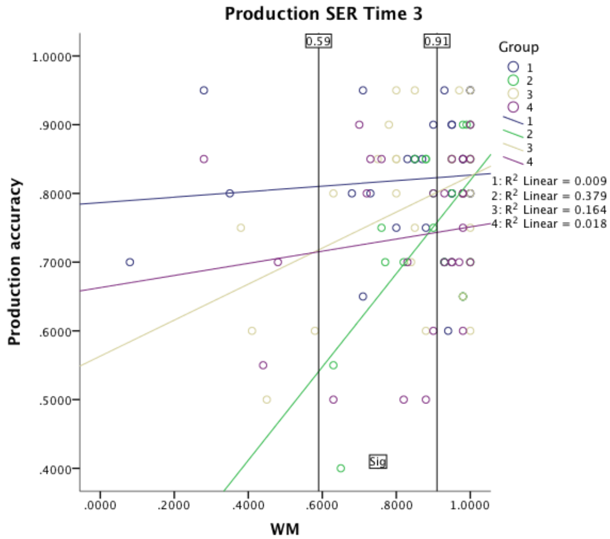


Figure 24 Group by WM interaction plot for Production of SER during Time 3

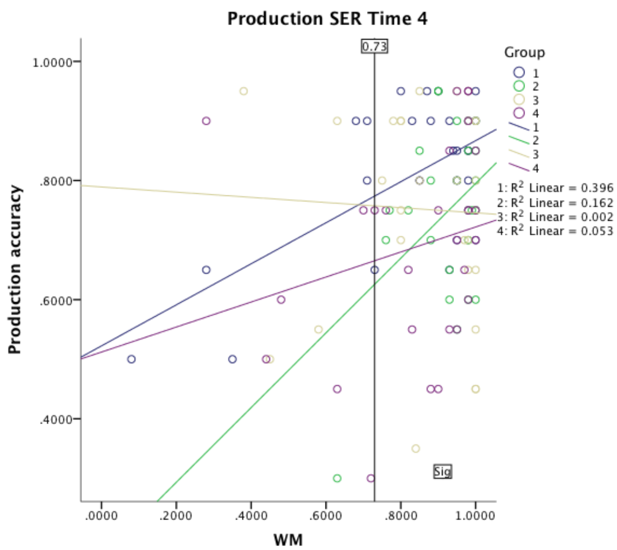


Figure 25 Group by WM interaction plot for Production of SER during Time 4

GJT

The means and SD can be found in Table 37 below. There were two models for the GJT measure. The first model included LLAMA as the covariate and using the Greenhouse-Geisser correction for sphericity demonstrated only a main effect of Time, $F(2.2,5)=32.25$, $p<.001$, partial $\eta^2=.23$. Table 38 shows the significant mean differences for Time. It can be seen that participants in all groups gained significantly more knowledge at Time 2, 3 and 4 compared to Time 1. In addition, participants in Group 4 [GEb-GEd-] had also significant loss of knowledge from Time 3 to Time 4. Since there is no indication that the regression slopes assumption has been violated, results from the ANCOVA model with respect to group differences will be interpreted. Pairwise comparisons demonstrated that Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] were significantly better than Group 4 [GEb-GEd-] during Time 3 (MD=.1, *p*=.03, MD=.1, *p*=.05, respectively) and during Time 4 (MD=.119, *p*=.03, MD=.19, *p*<.01, respectively).

Table 37 *Means and Standard Deviations for GJT SER*

Group	T1	T2	T3	T4
GEb+GEd+ (N=30)	.43 (.14)	.72 (.13)	.81 (.11)	.76 (.12)
GEb+GEd- (N=28)	.44 (.15)	.74 (.11)	.78 (.12)	.75 (.15)
GEb-GEd+ (N=30)	.45 (.15)	.73 (.18)	.82 (.12)	.79 (.15)
GEb-GEd- (N=28)	.44 (.15)	..63 (.13)	.72 (.15)	.64 (.18)

Table 38 *Significant mean differences across Time for GJT SER ANCOVA model with LLAMA*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.296***	.091**	.387***			.334***
Group 2 GEb+GEd-	.273***		.323***			.283***
Group 3 GEb-GEd+	.292***	.083**	.375***			.344***
Group 4 GEb-GEd-	.193***	.086**	.279***	-.074*		.205***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

The second model, using WM as the covariate, demonstrated only a significant two-way interaction between Group and WM, $F(4,108)=3.3$, *p*=.01, partial $\eta^2=.11$. Table 39 shows the significant mean differences for Time, which present the same pattern as the ones from the GJT with LLAMA model.

Table 39 *Significant mean differences across Time for GJT SER ANCOVA model with WM*

	T2>T1	T3>T2	T3>T1	T4>T3	T4>T2	T4>T1
Group 1 GEb+GEd+	.289***	.089**	.377***			.33***
Group 2 GEb+GEd-	.278***		.326***			.286***
Group 3 GEb-GEd+	.28***	.094**	.375***			.342***
Group 4 GEb-GEd-	.19***	.087**	.278***	-.074*		.204***

Note: The mean difference (MD) will be reported only where significant differences appear.
*p<.05, **p<.01,***p<.001

Results from the J-N analysis both corroborated and extended the findings from the ANCOVA model with GJT and LLAMA. Figure 26 illustrates that during Time 2, Group 1 [GEb+GEd+] is significantly better than Group 4 [GEb-GEd-] for individuals with WM of .81 and above. In addition, Figure 27 illustrates that during Time 3 and Time 4 both Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] are significantly better than Group 4 [GEb-GEd-], but only for individuals with WM of .81 and above.

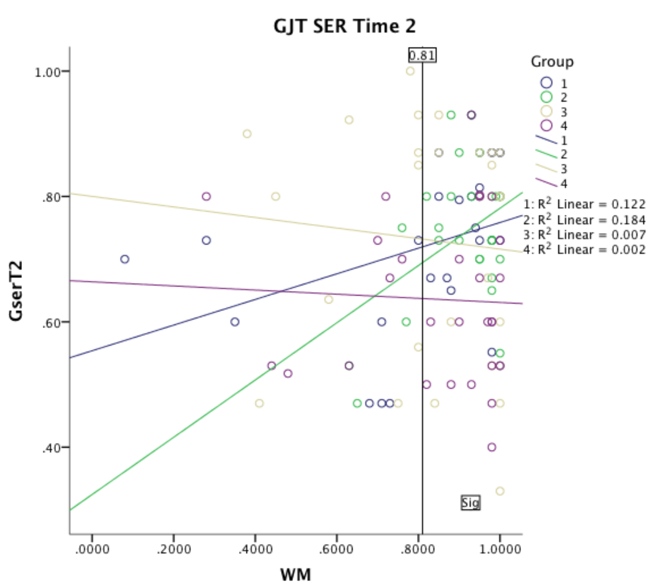


Figure 26 Group by WM interaction plot for GJT of SER during Time 2

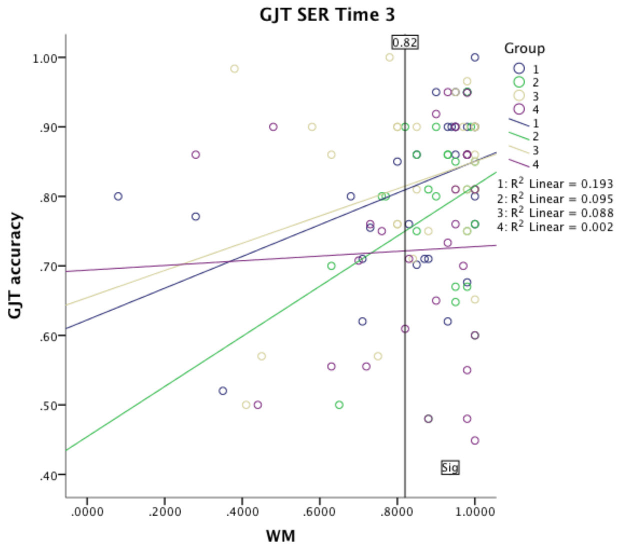


Figure 27 Group by WM interaction plot for GJT of SER during Time 3

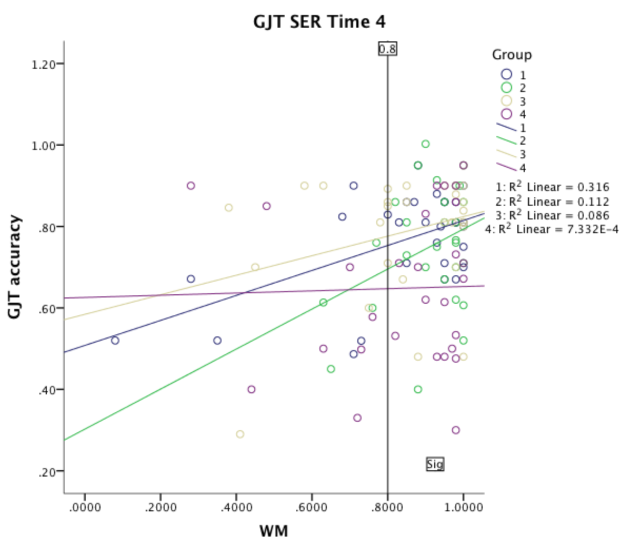


Figure 28 Group by WM interaction plot for GJT of SER during Time 4

Summary of SER results

Table 40 summarizes the significant findings from the ANCOVA model with SER. Table 41 summarizes the main group differences for the SER structure. It can be seen that at Time 1 all groups had comparable knowledge of SER, and that at Time 2 Group 1 [GEb+GEd+] significantly outperformed Group 4 [GEb-GEd-] as shown by comprehension with LLAMA, production with LLAMA, production with WM, GJT with LLAMA as well as GJT with WM. In addition, Group 3 [GEb-GEd+] was also significantly superior to Group 4 [GEb-GEd-] as shown by production with LLAMA, production with SER, production with WM and GJT with LLAMA. At Time 3 measures mainly indicated that Group 1 [GEb+GEd+] remained significantly superior to Group 4 [GEb-GEd-]: comprehension with LLAMA, production with LLAMA, production with WM, GJT with LLAMA, as well as GJT with WM. Measures also revealed that at Time 3 Group 1 [GEb+GEd+] was significantly better than Group 2 [GEb+GEd]: production with LLAMA and production with WM and that Group 3 [GEb-GEd+] was significantly better than Group 4 [GEb-GEd-]: GJT with LLAMA and GJT with WM. Finally, during the delayed posttest, Group 1 [GEb+GEd+] remained significantly better than Group 4 [GEb-GEd-] as demonstrated by comprehension with WM, production with LLAMA, production with SRT, production with WM, GJT with LLAMA and GJT with WM. Group 3 [GEb-GEd+] also remained significantly better than Group 4 [GEb-GEd-] as shown by GJT with LLAMA and GJT with WM.

In terms of within-group differences across time for the SER structure, the comprehension data revealed that the only increase from Time 3 to Time 4 was for participants in Group 1 [GEb+GEd+]. In addition, the GJT data revealed that only Group 4 [GEb-GEd-] had a significant loss from Time 3 to Time 4. Finally, the main group differences discussed after the presentation of the inferential statistics for each of the two structures are once again synthesized in Table 42.

Table 40 *Summary table for Repeated-Measures ANCOVA models for the SER structure*

	Time	Time*Group	Time*Group*Cov	Group*Cov	Group
	Only significant F values are reported; *p<.05, **p<.01,***p<.001				
1. Comp SER with LLAMA	$F(2.3, 242.2)= 4.33$**		$F(9.5, 242.2)= 1.84$*		
2. Comp SER with WM				$F(4, 101)=7.5$***	
3. Prod SER with LLAMA	$F(2.9, 288.24)= 28.9$***		$F(11.6, 288.24)= 1.6$*	$F(4,99)=4.3$**	$F(3,99)= 4.4$**
4. Prod SER with SRT	$F(2.8, 286.3)= 29.4$***				
5. Prod SER with WM		$F(8.7, 287.5)=2.1$*	$F(11.6, 287.5)= 4.2$**	$F(4,99)=3.9$**	
6. GJT SER with LLAMA	$F(2.2,5)= 32.25$***				
7. GJT SER with WM				$F(4,108)= 3.3$**	

Table 41 *Summary table of significant group differences for the SER structure*

	Time* Group* Covariate	Group* Covariate	J-N	Pairwise comparisons
Comprehension LLAMA	√		(1) Time 2: Gr 1>Gr4 for LLAMA 0-67.14 (2) Time 3: Gr 1>Gr4 for LLAMA 11.27-54.57	
Comprehension WM		√		(1) Time 4: G1>G4
Production LLAMA	√^	√	(1) Time 2: Gr1,3>Gr4 for LLAMA 0-68.4 (2) Time 3: Gr1>Gr2,4 for LLAMA 0-53.83 (3)Time 4: Gr1>Gr4 for LLAMA 14.6-76.4	
Production SRT				(1) Time 2: Gr1,3>Gr4 (2) Time 3: Gr1>Gr4 (3) Time 4: Gr1>Gr4
Production WM	√	√	(1) Time 2: Gr1,3>Gr4 for WM .73-1 (2) Time 3: Gr1>Gr2, 4 for WM .59-.91 (3) Time 4: Gr1>Gr4 for WM .73-1	
GJT LLAMA				(1) Time 3: Gr1,3>Gr4 (2) Time 4: Gr1,3>Gr4
GJT WM		√	(1) Time 2: Gr1>Gr4 for WM .81-1 (2) Time 3: Gr1,3>Gr4 for WM .81-1 (3) Time 4: Gr1,3>Gr4 for WM .81-1	

Table 42 *Summary of results for group differences*

	OV		SER	
	Group differences	Outcome measures	Group differences	Outcome measures
Time 1				
Time 2			a) G1>G4 b) G1,3>G4 c) G1,3>G4 d) G1,3>G4	a) J-N Comp (LLAMA 0-67.14); J-N GJT (WM .81-1) b) J-N Prod (LLAMA 0-68.4) c) Prod with SRT, GJT with LLAMA d) J-N Prod (WM .73-1)
Time 3	a) G1,2>G4	a) J-N Comp (LLAMA 46.2-100)	a) G1>G4 b) G1>G4 c) G1>G2,4 d) G1,3>G4	a) J-N Comp (LLAMA 11.27-54.57); J-N GJT (WM .81-1) b) Production SRT c) J-N Prod (LLAMA 0-53.83); JN Prod (WM .59-.91) d) GJT LLAMA
Time 4	a) G3>G4 b) G1,3>G4	a) GJT no Cov, GJT with LLAMA, GJT with SRT. b) J-N Comp (LLAMA 20.64-41.16)	a) G1>G4 b)G1,3>G4 c) G1>G4 d) G1,3>G4	a) Comp WM; Prod SRT b) GJT LLAMA c) J-N Prod (LLAMA 14.6-76.4); J-N Prod (WM .73-1) d) J-N GJT (WM .81-1)

Simple Linear Regression models following significant ID and outcome measure correlations

Following the significant correlations between the covariates and the outcome measures for each group separately, simple linear regressions were calculated to test if the ID measures significantly predicted participants' performance on the outcome measures (Appendix F). Following from the interactions between LLAMA and Group plotted in the figures above, the regression results additionally confirm that LLAMA consistently and significantly explains large portions of the variance in the performance of Group 4 [GEb-GEd-] for the OV structure based on the comprehension data for the delayed posttest, production data for Times 2, 3, and 4, and GJT data for Times 2, 3, and 4. In addition, LLAMA significantly predicted participants' performance on the comprehension and GJT at Time 4 for Group 2 [GEb+GEd-], and on the GJT performance at Time 2 for Group 1 [GEb+GEd+]. WM on the other hand had the strongest relationship with the treatment in Group 2 [GEb+GEd-]. Moreover, WM seems to play a role in the performance of Group 1 [GEb+GEd+] during the delayed posttest. Apparently Groups 1 [GEb+GEd+] and 2 [GEb+GEd-] had not sufficiently proceduralized the rules for OV and needed to fall back on their declarative knowledge. For Group 2 [GEb+GEd-] this was the case for the immediate and the delayed posttest, for Group 1 [GEb+GEd+] only for the delayed posttest.

Similarly, for the acquisition of the SER structure (Appendix F), LLAMA was most important in predicting outcomes for Group 4 [GEb-GEd-], and WM for Group 2 [GEb+GEd-]. In addition to these consistent relationships, WM seemed to be involved in the performance of participants in Group 1 [GEb+GEd+] during the comprehension of SER at Times 3, and 4, production of SER during Times 2, and 4, suggesting that although Group 1 [GEb+GEd+] was presented with the GE both before and during practice, participants were still engaged in the phase of proceduralization and were working with the grammar rules while practicing and performing on the tests. The relationship between knowledge loss and WM is more evident in the performance on the SER structure. While by the time of the delayed posttest Group 3 seems to have proceduralized the grammar rules, as shown by the non-significant relationship between WM and performance at Time 4 for Group 3 [GEb-GEd+], Group 1 [GEb+GEd+] heavily relies on WM to perform at Time 4.

Additional analysis with grammatical and ungrammatical GJT items

OV

ANCOVA with LLAMA on ungrammatical GJT items

ANCOVAs with LLAMA and WM were conducted for the grammatical and ungrammatical items from the GJT separately. The ANCOVA model with the ungrammatical items for OV and LLAMA as the covariate, demonstrated a significant main effect of Time, $F(2.6,284.8)=33.96$, $p<.001$, a significant main effect of Group, $F(3,108)=2.9$, $p=.03$, and a significant interaction between Group and LLAMA, $F(4,108)=2.6$, $p=.04$. Pairwise comparisons for time demonstrated that while all groups' performance at Times 2, 3 and 4 was significantly better than at Time 1, for Group 2 [GEb+GEd-], performance during Time 3 was significantly better than Time 4, suggesting a significant knowledge loss for Group 2 [GEb+GEd-]. Following the significant interaction between Group and LLAMA, the J-N results revealed that group differences exist for performance on ungrammatical items during Time 3 for participants who have LLAMA from 0 to 37.9. In particular, Group 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] show a significant superiority to Group 4 [GEb-GEd-] for individuals with LLAMA ranging from 0 to 37.9 (Figure 29).

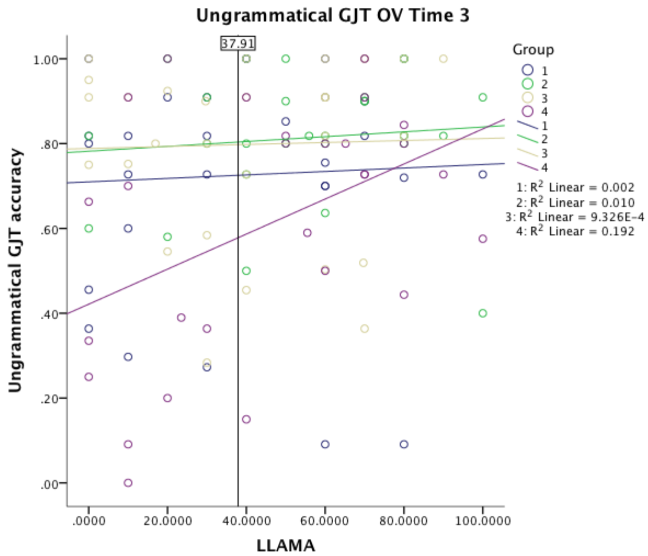


Figure 29 Group by WM interaction plot for Ungrammatical GJT of SER Time 3

ANCOVA with LLAMA on grammatical GJT items

The ANCOVA model with the grammatical items for OV and LLAMA as the covariate, demonstrated a significant main effect of Time, $F(2.2,241.4)=15.72$, $p<.001$. Pairwise comparisons revealed that there were no group differences across time: for all groups performance at Times 2, 3 and 4 was significantly better than at Time 1.

ANCOVA with WM on ungrammatical GJT items

Two additional models with WM as the covariate were run for the OV structure. The ANCOVA model with the ungrammatical items for OV and WM as the covariate demonstrated that there was a significant interaction between Group and WM, $F(4.108)=2.6$. $p=.04$. Since the J-N did not specify any significant regions of significance, pairwise comparisons will be reported. Pairwise comparisons revealed that the only group differences were during Time 3 between Group

3 [GEb-GEd+] and 4 [GEb-GEd-], MD=.189, p<.01. Pairwise comparisons for Time demonstrated that all groups' performance during Times 2, 3 and 4 was significantly better than Time 1.

ANCOVA with WM on grammatical GJT items

The ANCOVA model with the grammatical items for OV and WM as the covariate demonstrated that there was only a significant interaction between Group and WM, $F(4,108)=2.8$, $p=.03$. The J-N procedure specified that there were no regions of significance for the observed data. There was a main effect for time: for all groups performance during Times 2,3 and 4 was significantly better than at Time 1.

SER

ANCOVA with LLAMA on ungrammatical GJT items

The ANCOVA model with the ungrammatical items for SER and LLAMA as the covariate demonstrated that there was a significant main effect of Time, $F(2.3,258.35)=37.85$, $p<.001$, and a main effect of Group approaching significance, $F(3,108)=2.4$, $p=.06$. Pairwise comparisons demonstrated that during the delayed posttest, all groups that received GE, i.e., Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] were significantly better than Group 4 [GEb-GEd-], MD=.219, $p<.01$, MD=.168, $p=.04$, MD=.167, $p=.03$, respectively. Pairwise comparisons for time revealed that while all groups' performance during Times 2, 3, and 4 were significantly better than 1, for Group 2 [GEb+GEd-] and 4 [GEb-GEd-] performance during Time 3 was significantly better than Time 4, suggesting knowledge loss at the delayed posttest. J-N analysis revealed that during Time 4 there were group differences only for participants with a score of

LLAMA of 20.3 and above. From Figure 30 it can be seen that Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] performed significantly better than Group 4 [GEb-GEd-].

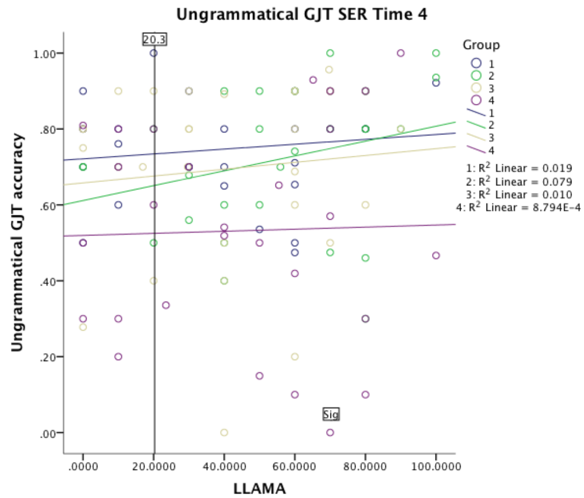


Figure 30 Group by LLAMA interaction plot for Ungrammatical GJT of SER at Time 4

ANCOVA with LLAMA on grammatical GJT items

The ANCOVA model with the grammatical items SER and LLAMA as the covariate demonstrated only a significant main effect of Time, $F(2.3, 251.59)=5.01$, $p<.01$. Pairwise comparisons for Time, revealed that for Groups 2 [GEb+GEd-] and 4 [GEb-GEd-] there were no significant differences across the four times. For Group 1 [GEb+GEd+], performance during Time 3 was significantly better than Time 1, 2 and 4 (MD=.279, p<.001, MD=.156, p<.001, MD=.09, p=.02, respectively), and that Time 4 was significantly better than Time 1, MD=.19, p<.01. Only for Group 3 [GEb-GEd+], performance at Times 2, 3 and 4 was significantly better than at Time 1.

ANCOVA with WM on ungrammatical items

Finally, the ANCOVA mode with the ungrammatical items for SER and WM as the covariate demonstrated a significant main effect of Time, *F*(2.4, 261.7)=3.47, *p*=.02, and significant interaction between Group and WM, *F*(4,108)=3.01, *p*=.02. J-N analysis revealed that during Time 3 group differences exist only for individual with WM of .75 and above and during Time 4 only for individuals with WM of .65 and above. From Figures 31 and 32 it can be seen that Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] are significantly better than Group 4 [GEb-GEd-].

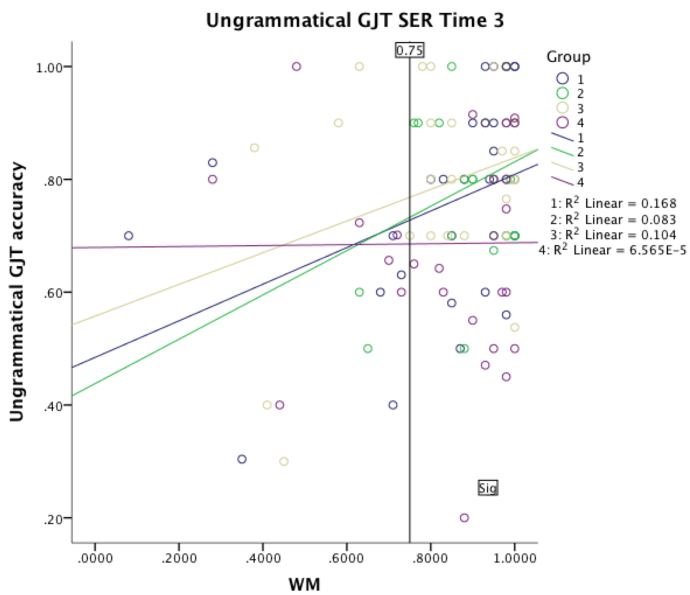


Figure 31 Group by WM interaction plot for Ungrammatical GJT of SER at Time 3

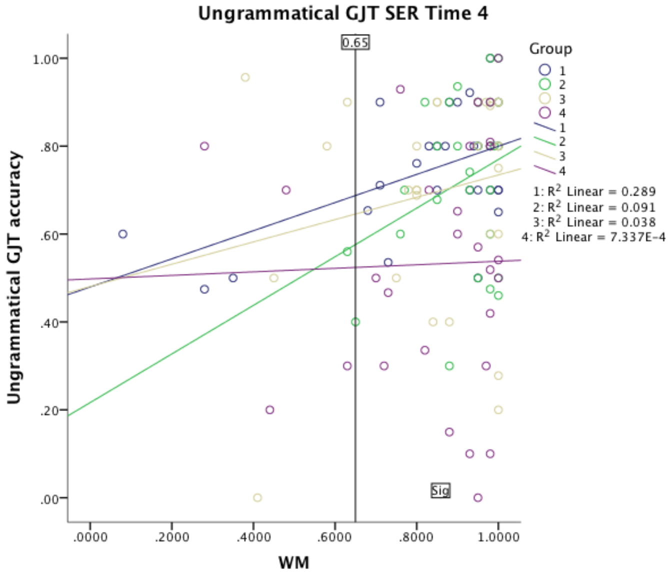


Figure 32 Group by WM interaction plot for Ungrammatical GJT of SER at Time 4

ANCOVA with WM on grammatical GJT items

The ANCOVA model with the grammatical items for SER and WM as the covariate demonstrated no significant findings. Pairwise comparisons revealed that during Time 3, Group 1 [GEb+GEd+] was significantly better than Group 2 [GEb+GEd-], MD=.108, *p*=.02. Similar to the results from the ANCOVA model with LLAMA, pairwise comparisons demonstrated that for Groups 2 [GEb+GEd-] and 4 [GEb-GEd-] there were no significant differences across the four times. For Group 1 [GEb+GEd+], performance during Time 3 was significantly better than Time 1, 2 and 4, and that Time 4 was significantly better than Time 1. Only for Group 3 [GEb-GEd+] was performance during Times 2, 3 and 4 significantly better than at Time 1.

Summary of analysis from grammatical and ungrammatical GJT items

To summarize, the results from the grammatical and ungrammatical items of the GJT demonstrate that for the OV structure there were no significant group differences during Time 1, 2 and 4. During Time 3, only participants with a LLAMA score ranging from 0 to 37.9, in Groups

1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] gained significantly more knowledge than participants in Group 4 [GEb-GEd-] as demonstrated by their performance on the ungrammatical items (See Table 42). Moreover, pairwise comparisons from the ANCOVA model with ungrammatical items and WM as the covariate demonstrated that only Group 3 [GEb-GEd+] was significantly better than Group 4 [GEb-GEd-]. There were no significant differences on participants' performance on the grammatical items.

For the SER structure it was found that during the delayed posttest participants with a score on LLAMA ranging from 20.3 and above as well as WM score of .65 and above in Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] were significantly better than participants in Group 4 [GEb-GEd-] in their performance on ungrammatical items. In addition, during Time 3 Group 1 [GEb+GEd+] scored significantly higher than Group 2 [GEb+GEd-] on the grammatical sentences.

Table 42 Summary of results for group differences on grammatical and ungrammatical GJT items

	OV		SER	
	Group differences	Outcome measures	Group differences	Outcome measures
Time 1				
Time 2				
Time 3	a) G1,2,3,>G4 b) G3>G4	a) J-N ungrammatical items (LLAMA 0-37.9) b) Pairwise comparisons ungrammatical items with WM	a) G1>G2 b) G1,2,3 >G4	a) Pairwise comparisons grammatical items with WM b) J-N ungrammatical items (WM .75 and above)
Time 4			a) G1,2,3 >G4	a) J-N ungrammatical items (LLAMA 20.3 and above, and WM .65 and above)

Reaction time data analysis

The reaction time data was only analyzed for the comprehension measure, as this was the only measure with time pressure. Participants were told to choose the right picture as quickly and accurately as possible upon reading the target sentence.

OV

The means and standard deviations for the reaction times of comprehension OV can be found in Table 43.

Table 43 *Means and standard deviations for RT Comprehension OV*

Group	T1	T2	T3	T4
1 [GEb+GEd+] (N=30)	7008.31 (2041.43)	5866.48 (1884.84)	4626.29 (1719.88)	4756.35 (1674.67)
2 [GEb+GEd-] (N=28)	6484.53 (1518.73)	5271.80 (1608.05)	4069.11 (1246.47)	4381.03 (1393.55)
3 [GEb-GEd+] (N=30)	7349.22 (2801.72)	5934.99 (1590.15)	5105.86 (1833.96)	5366.26 (2002.77)
4 [GEb-GEd-] (N=28)	6456.18 (1925.69)	5584.16 (1982.80)	4032.32 (1755.93)	4071.81 (1818.49)

The results from the Repeated-measures ANOVA demonstrated a significant main effect of Time, $F(2.8,319.46)=78.46$, $p<.001$, and a significant main effect of Group, $F(3,112)=2.7$, $p=.05$. Additional pairwise comparisons using the Bonferroni correction revealed that at Time 4 Group 3 [GEb-GEd+] was significantly slower than Group 4 [GEb-GEd-], MD=1294.448, $p=.03$. In terms of reaction times across the four sessions per group, for Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+], performance at Time 2, 3 and 4 was significantly faster than Time 1. In addition, performance during Time 3 and Time 4 was significantly faster than Time 2. For Group 4 [GEb-GEd-] performance during Time 3 and 4 was significantly faster than Time 1 and 2.

SER

The means and standard deviations for the reaction times of comprehension SER can be found in Table 44.

Table 44 *Means and standard deviations for RT Comprehension SER*

Group	T1	T2	T3	T4
1 [GEb+GEd+] (N=30)	6054.99 (1978.67)	4447.67 (1587.81)	3829.59 (1304.69)	3294.42 (851.22)
2 [GEb+GEd-] (N=28)	5893.98 (1693.52)	4167.66 (1157.36)	3767.42 (1139.55)	3188.98 (912.57)
3 [GEb-GEd+] (N=30)	6187.25 (2503.18)	4339.54 (1665.14)	3606.53 (1577.74)	3491.42 (1375.36)
4 [GEb-GEd-] (N=28)	6181.46 (1983.24)	3925.97 (1687.29)	3462.39 (1494.87)	3314.45 (1167.84)

The results from the Repeated-measures ANOVA demonstrated only a significant main effect of Time, $F(2.04,228.98)=110.361$, $p<.001$. Based on these results (a) all groups performed similarly with regards to the speed of answering and (b) all groups decreased their reaction times to a comparable degree from Time 1 to Time 4.

The results from the repeated-measures ANOVA with the reaction time data for OV and SER do not provide any significant information with regards to the differential proceduralization of the target rules as all groups decreased their reaction times to similar extents.

Chapter 6: Discussion

The main purpose of the present dissertation is to investigate the role of particular timings of GE in the acquisition of declarative and procedural knowledge of two Spanish structures: OV and SER. In addition, the dissertation aims to explore how individual differences in language learning, such as language aptitude and working memory, interact with the different experimental treatments. By taking into account the different combinations of pre- and during-practice GE, qualitatively different L2 structures, and individual differences, this dissertation aims to provide a better understanding of the precise role of explicit grammar rules in the process of L2 learning. The first hypothesis concerns the effect of different GE timing on the acquisition of declarative knowledge for the two structures. It was predicted that the closer together rules and practice are, the more successful the acquisition of declarative knowledge would be. The second hypothesis indirectly explored the question of specificity of practice, and predicted that the skill in which participants practiced the two structures would not transfer to the reverse skill, production. The third hypothesis concerns the acquisition of procedural knowledge predicting that the closer together GE and practice are, the more successful the transition to procedural knowledge would be, leading to improved retention at the delayed posttest. Finally, the fourth hypothesis, which is really a set of hypotheses, aimed to investigate the relationship between the IDs and the outcome measures and predicted (a) a stronger relationship between LLAMA and the outcome measures during the initial stages of learning, (b) that SRT would have an impact on the later stages of learning, (c) that WM would play a bigger role at the initial stages of learning and would be most visible in the groups that did not receive GE during practice, (d) that LLAMA would have a stronger effect in the groups that did not receive GE during practice, and (e) that LLAMA would have the strongest effect in the GEb-GEd- group. In what follows, each of the hypotheses is discussed in light of the results.

Discussion of Hypotheses

Hypothesis 1

H1: The closer together rule and practice occur, the better the acquisition of declarative knowledge of the L2 target structure will be. More specifically, it is hypothesized:

H1a: GEb+GEd+ > GEb- GEd+ (Group 1 > Group 3, 2, 4)

H1b: GEb- GEd+ > GEb+ GEd- (Group 3 > Group 2, 4)

H1c: GEb+ GEd- > GEb- GEd- (Group 2 > Group 4)

Results from the MTK and the accuracy on the ungrammatical sentences provided both similar and contrasting implications. First, both outcome measures demonstrated that for the OV structure there seems to be no difference across groups. All of the groups attained a comparable level of declarative knowledge of the OV structure after the second session. Second, both outcome measures demonstrated that there are group differences with respect to the metalinguistic knowledge of SER. However, while results from the MTK test demonstrated that Group 3 [GEb-GEd+] was significantly better than all other groups, results from the accuracy on the ungrammatical GJT items demonstrated that only Group 1 [GEb+GEd+] was significantly better than Group 4 [GEb-GEd-]. Given these results, it can be said that Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] were superior to the other groups in the declarative knowledge the learners acquired after session two. It seems that receiving GE while practicing helps learners to solidify the declarative knowledge that was presented to them: it was exactly the groups with GEd+ that showed superiority to the other groups in terms of (a) being significantly more accurate than Group

4 [GEb-GEd-], (b) being able to retain the acquired knowledge at the one-week delayed posttest, and (c) being only minimally affected by individual differences.

Moreover, the results from the MTK and the accuracy on the ungrammatical GJT items are in line with the results from the other outcome measures, in that group differences seem to be present for the SER but not the OV structure. It seems that participants were more ready to understand the GE for the OV structure and consequently acquire this structure at similar levels regardless of the treatment they received. In contrast, the SER structure seemed to pose more difficulty for learners which may be due to the fact that the difference between inherent and circumstantial traits is not morphologically expressed in English.

Therefore, it can be concluded that while H1a and H1b were partially supported in that Group 1 [GEb+GEd+] and Group 3 [GEb-GEd+] were superior to Group 4 [GEb-GEd-], but not to Group 2 [GEb+GEd-]. The results do not provide support for H1c: Group 2 [GEb+GEd-] was not found to be superior to Group 4 [GEb-GEd-].

Hypothesis 2

H2: After the first learning phase, learners will not have proceduralized their declarative knowledge and will be able to use this knowledge equally well for both comprehension and production activities.

In order to address this hypothesis, the means of the comprehension and the production measure were descriptively compared. If the knowledge used in comprehension was proceduralized, we would expect, based on the theoretical assumption of SAT and previous empirical studies, that this knowledge will not be easily available in the skill of production. In addition to looking at accuracy scores for each group on the comprehension and production measure, we can compare whether comprehension evolves differently from production across the experimental groups.

OV

For OV, the predictions for this hypothesis are partially supported. The means from the production measures across all groups are substantially lower than the means for the comprehension measure during Times 2, 3 and 4. It can be said that proceduralization for all groups must have started even after the first session of training as evidenced by the difference in performance on the two measures for Session 2. Learners in Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] were using their proceduralized knowledge, preventing them from transferring their knowledge to another skill, i.e., production. All of the groups except Group 2 [GEb+GEd-] were at chance on the production measure after session two. It is interesting to observe that while Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 4 [GEb-GEd-] were at chance level during the delayed posttest, only Group’s 3 [GEb-GEd+] performance was significantly above chance (mean=.61). In this group, when GE was provided only during practice, the amount of declarative knowledge acquired was higher than for the other groups as observed from the MTK test. Consulting the GE while practicing must have contributed to a stronger transition to procedural knowledge, which in turn enabled participants in this group to better retain their knowledge at the one-week delayed posttest.

In terms of how these skills evolved over time, there seems to be a relatively parallel development of both skills. Results from the comprehension measure revealed that performance at Times 2, 3 and 4 was significantly better than at Time 1. In addition, all of the groups, except Group 3 [GEb-GEd+] also had significant losses from Time 3 to Time 4. For the production measure of OV, Groups 1 [GEb+GEd+], 3 [GEb-GEd+], and 4 [GEb-GEd-] performed significantly better at Times 2, 3 and 4 than at Time 1. Group 2's [GEb+GEd-] performance was slightly different: while Time 2 and 3 were significantly better than Time 1, performance at the delayed posttest was significantly worse than at Time 3, suggesting loss of knowledge during the delayed posttest.

SER

For the SER structure the results are slightly different. Participants in all groups performed equally for both skills, suggesting that even after two sessions of task-essential practice, participants could not fully proceduralize their declarative knowledge. During the delayed posttest, the means for the comprehension measure were higher than the means for the production measure for all groups.

In terms of development of these skills over time, for the comprehension measure participants in Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] scored significantly higher at Times 2, 3 and 4 than at Time 1. Group 4 [GEb-GEd-], however, performed significantly better at Time 4 than at Time 1 only. In production, all groups' performance was significantly higher at Times 2, 3 and 4 than at Time 1.

Specificity of practice was, therefore, observed for the OV structure and only for the groups that started proceduralizing their declarative knowledge during session 2, but not for the SER structure, which was harder to proceduralize and might have required more practice. Discussing the results in light of previous findings, the current experiment demonstrated that for the OV structure, which was easier to proceduralize, comprehension-based practice was not enough for learners to develop productive knowledge of this structure, a finding that is contrary to what was observed in VanPatten & Cadierno's (1993) study, which experimented with the same syntactic structure. For the SER structure, no difference between the comprehension and the production task was observed suggesting that participants even at the delayed posttest must have relied on their declarative knowledge to perform on the tests. Additional evidence for this claim comes from the relationship between WM and the outcome measures across groups. While WM only clearly predicted learners' performance in Group 2 [GEb+GEd-] for both structures, the relationship between WM and outcome measures for Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] was mainly present for the SER structure. Based on the correlation matrices and regression models, it was shown that WM had a stronger and more consistent relationship with the SER structure as opposed the OV one. Even participants in the groups that received GE during practice had to rely on the declarative knowledge of the SER structure during the later stages of testing. For example, WM explained 14% of the variance on the comprehension of SER during Time 2, and 44% variance on the comprehension of SER during the delayed posttest. For the production measure, WM explained 37% of the variance for Group 1 [GEb+GEd+] during Time 2 and 40% of the variance during the delayed posttest. Similarly, but less strongly, WM explained 19% of the variance of comprehension of SER and 14% of the variance in the production of SER for Group 3 [GEb-GEd+] during Time 2, as well as 27% for comprehension and 16% for production during Time 3. The same participants were using their proceduralized knowledge when performing on the OV items, as evidenced by the virtually non-existent relationship between WM and performance on the OV tests (and by the lack of difference between production and production tests for OV). While WM significantly predicted performance on the comprehension of GJT during Time 3 for Group 1 [GEb+GEd+], WM was not a significant predictor for performance on the OV structure for Group 3 [GEb-GEd+].

Hypothesis 3

H3: The closer together rule and practice are, the better the learning and retention of the L2 target structure will be. More specifically, it is hypothesized:

H3a: GEb+GEd+ > GEb- GEd+ (Group 1 > Group 3)

H3b: GEb- GEd+ > GEb+ GEd- (Group 3 > Group 2)

H3c: GEb+ GEd- > GEb- GEd- (Group 2 > Group 4)

The hypotheses will be evaluated by taking into account (a) group differences in percentage correct accuracy from the three outcome measures, and (b) retention of knowledge over time. (The differential involvement of language aptitudes across the four experimental groups will be discussed under Hypothesis 4.)

To help the reader evaluate Hypothesis 3, the significant group differences across measures are presented in Table 45.

Table 45 *Summary of results for group differences*

	OV		SER	
	Group differences	Outcome measures	Group differences	Outcome measures
Time 1				
Time 2			a) G1>G4 b) G1,3>G4 c) G1,3>G4 d) G1,3>G4	a) J-N Comp (LLAMA 0-67.14); J-N GJT (WM .81-1) b) J-N Prod (LLAMA 0-68.4) c) Prod with SRT, GJT with LLAMA d) J-N Prod (WM .73-1)
Time 3	a) G1,2>G4	a) J-N Comp (LLAMA 46.2-100)	a) G1>G4 b) G1>G4 c) G1>G2,4 d) G1,3>G4	a) J-N Comp (LLAMA 11.27-54.57); J-N GJT (WM .81-1) b) Production SRT c) J-N Prod (LLAMA 0-53.83); JN Prod (WM .59-.91) d) GJT LLAMA
Time 4	a) G3>G4 b) G1,3>G4	a) GJT no Cov, GJT with LLAMA, GJT with SRT. b) J-N Comp (LLAMA 20.64-41.16)	a) G1>G4 b)G1,3>G4 c) G1>G4 d) G1,3>G4	a) Comp WM; Prod SRT b) GJT LLAMA c) J-N Prod (LLAMA 14.6-76.4); J-N Prod (WM .73-1) d) J-N GJT (WM .81-1)

OV

(a) Group differences:

During Time 3, Groups 1 [GEb+GEd+] and 2 [GEb+GEd-] obtained significantly higher gains than Group 4 [GEb-GEd-] as measured by the comprehension test. However, during the delayed posttest, only the groups that received GE during practice, i.e., Group 1 [GEb+GEd+] and Group 3 [GEb-GEd+] were found to be significantly superior to Group 4 [GEb-GEd-], as demonstrated by the comprehension and GJT test.

In addition, an analysis of the participants' performance on the ungrammatical items on the GJT revealed that the treatments that Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] received generated gains significantly higher than the ones for Group 4 [GEb-GEd-] during Time 3, but no significant differences remained either for the grammatical or ungrammatical items during the delayed posttest.

Furthermore, the comprehension test and the GJT measure revealed that all groups had above-chance performance at Times 2, 3 and 4. Performance on the production differed across groups. While only Groups 1 [GEb+GEd+] and 2 [GEb+GEd-], achieved above chance accuracy at Time 3, during the delayed posttest only the treatment that Group 3 [GEb-GEd+] received generated gains that were significantly above chance, $t(29)=2.6$, $p=.01$. Group 4 [GEb-GEd-] was at chance level for all production tests.

Based on the overall performance on the comprehension measure and the GJT, as well as the above chance performance on the production measure at Time 4, it appears that only providing GE during practice enabled learners to perform successfully on these measures. It is not clear why Group 1 [GEb+GEd+], which also received GE during practice, did not achieve above-chance performance on the production measure and was not found to be superior to Group 3 [GEb-GEd+]. One explanation is that participants in this group may have relied more on the GE that was provided to them before the practice material and did not feel the need to go through the GE during practice. The experimenter did not have any control over whether participants in this group read the GE during practice or not. Evidence for this explanation comes from the regression model results showing that WM was a significant predictor of comprehension of OV during Time 3 for Group 1 [GEb+GEd+]. Initially, based on the theoretical underpinnings from SAT is was predicted that Group 1 [GEb+GEd+], which received GE both prior and during practice would generate superior learning gains than all other groups, the assumption being that it is important to focus learners' attention on the grammatical rule at hand so that they fully comprehend it and start the proceduralization stage before the beginning of practice. However, the current results do not support this assumption. Instead they suggest that for the OV structure there is no additional benefit for providing GE before practice, but rather that presenting GE only at the time when it is needed the most, i.e., while engaged in practice activities whose successful completion depends on the grammatical rules, was enough for high and durable accuracy. This finding is in line with Sallas et al.'s (2007) study as well as the theoretical claims put forward by Anderson's SAT that the provision of declarative knowledge and the tasks whose completion hinges upon this knowledge should be as close to each other as possible for proceduralization to be effective. In addition, the performance of Group 2 [GEb+GEd-], was not significantly different from that of Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], and while participants in this group achieved high accuracy on the comprehension and the GJT measures, they nevertheless did not gain enough knowledge to outperform Group 4 [GEb-GEd-] at any time, as shown by all three outcome measures. Finally, although participants in Group 4 [GEb-GEd-] managed to induce the target OV rule to some degree as shown by the means on the three outcome measures, accuracy was consistently lower than for participants in other groups, and as discussed below, largely depended on language aptitudes. In the next subsection, the performance of the groups will be evaluated in terms of retention of knowledge.

(b) retention of knowledge:

The comprehension measure demonstrated that all groups gained significantly more knowledge after the treatments, Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 4 [GEb-GEd-] but also had significant losses from Time 3 to Time 4. Only Group 3 [GEb-GEd+] showed steady performance across tests. The data from the production and the GJT measure showed a slightly different pattern. For the production, only Group 2 [GEb+GEd-] showed significant loss of knowledge from Time 3 to Time 4, suggesting that presenting GE only before practice was not enough for participants in this group to retain their knowledge. In fact, accuracy for this group dramatically dropped from 74 % correct at Time 3 to 47 % correct at Time 4. As for Group's 4 [GEb-GEd-] performance, although there was no significant loss from Time 3 to Time 4, performance during Time 4 dropped to chance: mean of .49. In addition, the GJT measure demonstrated that only Group 4's [GEb-GEd-] performance was significantly worse at Time 4 in comparison to Time 3. Finally, results from the ANCOVA model with the ungrammatical GJT sentences, similarly to the production measure, showed that only Group 2's [GEb+GEd-] performance was significantly worse at Time 4 than Time 3.

Overall, results suggest that for retention of knowledge too it was necessary for GE to be provided concurrently with practice as shown by the steady performance of Group 3 [GEb-GEd+] in all three measures. In contrast, two measures demonstrated that although Group 2 [GEb+GEd-] achieved high accuracy during Times 2 and 3, it also had a significant drop in performance at the one-week delayed posttest.

SER

For the SER construction, the pattern was slightly different and more robust. Based on the results from all outcome measures, it seems that for a successful acquisition of this structure it was essential for participants to receive concurrent GE with practice.

(a) Group differences:

The data from all three measures demonstrated that Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], which received GE during practice, performed consistently and significantly better than Group 4 [GEb-GEd-], during Times 2, 3 and 4. While these group differences were independent of LLAMA and WM in some measures, in others they depended on the level of the covariate. For instance, during Time 2 group differences on the comprehension measure were only observed for values of LLAMA between 0 and 67; group differences on the production measure were only observed for levels of LLAMA between 0 and 68 and WM above .73; and group differences on the GJT measure were only found for levels of WM above .81. This pattern is found for more or less the same values for LLAMA and WM at Times 3 and 4, with only one difference: during Time 4 the group differences demonstrated by the comprehension measure were not dependent on the level of LLAMA. Throughout Times 2, 3, and 4, group differences independent of the levels of the covariates were also found for the production models controlling for SRT as well as for the GJT models controlling for LLAMA.

In addition, while performance on the comprehension and production was above chance for all groups and all outcome measures, Group 4's [GEb-GEd-] performance on the ungrammatical items was once again at chance, suggesting that the above-chance performance on the overall GJT was driven by the accuracy on the grammatical items. In line with the results from the OV structure, only participants who received GE during practice were able to obtain significantly higher gains than participants in Group 4 [GEb-GEd-]. Results from performance on the ungrammatical items add one important aspect to these findings. Namely, as mentioned earlier, it was demonstrated that both during Time 3 and Time 4 all groups that received GE performed significantly better than participants who did not receive any GE on the ungrammatical SER GJT items as shown by the J-N analyses. These group differences, however, hinged on participants' language learning aptitude and WM. Only participants with LLAMA of 20.3 and above and WM of .75 and above, in Groups 1 [GEb+GEd+], 2 [GEb+GEd-] and 3 [GEb-GEd+], were superior to participants of the same profile in Group 4 [GEb-GEd-] in their performance on the ungrammatical GJT items.

(b) Retention of knowledge:

In terms of knowledge retention, the comprehension measure demonstrated that although all groups retained their acquired knowledge at the delayed posttest, only Group 1 [GEb+GEd+] performed significantly better at the delayed posttest in comparison to Time 3. For the SER structure then there was an additional benefit of providing GE prior to practice (Groups 1 [GEb+GEd+] and 3 [GEb-GEd+]).

While for the production measure there were no significant knowledge losses for any group, the GJT measure revealed that Group 4's [GEb-GEd-] performance was significantly lower at Time 4 than at Time 3. In addition, results from performance on the ungrammatical GJT items for SER showed that both Group 2's [GEb+GEd-] and Group 4's [GEb-GEd-] performance during the delayed posttest was significantly worse than at Time 3. The results from the grammatical items show a different pattern. While for Groups 2 [GEb+GEd-] and 4 [GEb-GEd-] there were no significant differences across the four times, Group's 1 [GEb+GEd+] performance during Time 3 was significantly better than Time 1, 2 and 4, and Time 4 was significantly better than Time 1.

It can be concluded, therefore, that for the retention of knowledge of the SER structure it was important not only to present the GE during practice, but also prior to practice. At least one of the measures demonstrated that not providing GE before practice leads to significant loss of knowledge.

The predictions for this research question were partially supported. In particular, while hypothesis 3a and 3b were partially supported in that there was evidence suggesting that Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] obtained gains significantly higher than Group 4 [GEb-GEd-], and Group 1 [GEb+GEd+] on some measures outperformed Group 2 [GEb+GEd-]. Evidence for 3c comes only from participants' performance on the ungrammatical GJT items. Namely for the OV structure it was demonstrated that Group 2 [GEb+GEd-] was significantly better than Group 4 [GEb-GEd-] during Time 3. For the SER structure, Group 2 [GEb+GEd-] outperformed Group 4 [GEb-GEd-] on ungrammatical items both during Time 3 and Time 4.

Hypothesis 4

H4: The interaction between individual differences and instructional treatments

OV

Results from all three outcome measures demonstrated that participants' accuracy was contingent on the Group by Language Aptitudes interaction. The most prominent ATI was observed in Groups 2 [GEb+GEd-] and 4 [GEb-GEd-], discussed in turn. Participants in Group 2 [GEb+GEd-] were only presented with the GE once before the beginning of the practice. In line with the predictions motivated by SAT, when declarative rules are not repeatedly presented to learners during the proceduralization stage, learners have to continue to retrieve the declarative rules from memory in order to perform successfully during practice and assessment tasks. This situation is not ideal for learning as individuals vary greatly in their WM capacity and thus successful learning will only be possible for individuals with high WM capacity. In addition, using WM while practicing takes away attentional resources that might be used elsewhere. The impact of WM was precisely the strongest in Group 2 [GEb+GEd-], the group without GE during practice, across the two structures. WM was a significant predictor in almost all of the regression models for Group 2 [GEb+GEd-], and explained 16% of the variance on the comprehension of OV during Time 2, and 25% during Time 4. For production, WM explained 15% of the variance during Times 2 and 3, as well as 16% of the variance on the GJT measure. A large portion of the variance in the performance of Group 2 [GEb+GEd-] was also explained by participants' language learning aptitude, especially during comprehension at Time 3 (19%) and Time 4 (54%), as well as for GJT at Time 4 (16%). Even though participants in this group received the same GE prior to practice, they had to rely on both their WM and language aptitude to perform successfully on all three measures. The impact of language learning aptitude during the performance of comprehension at Time 4 was exceptionally high. These findings stress the importance of providing learners with maximal guidance during the proceduralization stage, so that the reliance on language aptitude and the burden on WM is lessened.

Furthermore, it was predicted that language aptitude would be most visible in the GEb-GEd- condition. Although learners in Group 4 [GEb-GEd-] could figure out some rules by means of task-essential practice, their learning was only evident in the comprehension and GJT measures, and largely dependent on language learning aptitude. LLAMA consistently and significantly predicted performance on the outcome measures, explaining 47% of the variance in comprehension at the delayed posttest, 20% in the production during Time 3, 13% during Time 3, 22% in GJT Time 2, 20% during Time 3, and 20% during the delayed posttest. WM did not play any role in the performance of Group 4 [GEb-GEd-] for the OV structure.

With respect to Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], the impact of language aptitude and working memory was less in these groups, which received GE during practice. LLAMA was a significant predictor of participants' performance in Group 1 [GEb+GEd+] for the comprehension at Time 3 and 4, as well as GJT at Time 3. In this respect, Groups 1 [GEb+GEd+] and 2 [GEb+GEd-] showed a similar behavior, as LLAMA was also a significant predictor for Group 2 [GEb+GEd-] with respect to these measures. This reinforces our explanation of the reasons for why Group 1 [GEb+GEd+] was not significantly better than Group 3 [GEb-GEd+] in any of the three measures. Namely, participants in both Groups 1 [GEb+GEd+] and 2 [GEb+GEd-] relied on their language learning ability to perform on the comprehension tests during Times 3 and 4. Had participants in Group 1 [GEb+GEd+] paid close attention to the GE presented during practice, the involvement of language aptitude would be expected to be non-existent, as is the case with Group 3 [GEb-GEd+], which is discussed later. However, it should be emphasized that, even though to a lesser degree, language aptitude for explicit learning was also involved in the performance of the groups receiving GE during practice. As illustrated in the interaction plots participants across these groups who had high explicit learning aptitude benefitted more from the pedagogical treatment than the low aptitude learners, suggesting that GE during practice alone was not enough for successful learning.

Moreover, participants in Group 1 [GEb+GEd+] seem to have relied on WM during the later stages of learning. WM significantly explained 23% of the variance in participants' performance on the comprehension measure at Time 3, and 13% of the variance on the GJT measure at Time 4. This suggest that participants in this group did not proceduralize the declarative rule enough so that they still needed to keep the declarative OV rule in their WM while performing at Time 3 and Time 4.

Finally, the accurate performance of Group 3 [GEb-GEd+] was largely independent of language aptitude or working memory, rendering this group superior to the others in this respect. LLAMA and WM were not significant predictors for Group 3 [GEb-GEd+] at any point except for the production of OV during Time 2, at which point participants must have needed languagelearning aptitude to grasp GE and consolidate their declarative knowledge for this structure.

SER

The performance of all groups with respect to SER depended on language aptitudes. As was the case for the OV structure, the involvement of language aptitude and WM was most strongly felt in Groups 2 [GEb+GEd-] and 4 [GEb-GEd-].

Performance of Group 2 [GEb+GEd-] depended highly on both WM and LLAMA. WM explained large and significant portions of the variance in the accuracy of comprehension at Time 2 (34%), Time 3 (30%), Time 4 (29%); production at Time 2 (31%), Time 3 (38%), Time 4 (16%); and GJT at Time 2 (18%). In addition, this group's performance also depended on participants' language aptitude. LLAMA explained 12% of the variance in comprehension at Time 4, 25% of the variance in the production Time 4, as well as 10% of the variance in GJT at Time 2.

Similarly, the performance of Group 4 [GEb-GEd-] was largely contingent upon language aptitude, explaining 24% of the variance in comprehension at Time 2, 11% at time 3, 32% in production at time 2, 165 at Time 3 and 15% in GJT at Time 2. WM did not play any role in the performance of Group 4 [GEb-GEd-] for the SER structure.

With respect to Group 1's [GEb+GEd+] performance, results demonstrated that for the SER structure, in contrast to the OV one, WM had a very large impact in almost all outcome measures. WM explained 10% of the variance in comprehension at Time 2, 14% at Time 3, 44% at Time 4, 37% of the variance in production at Time 2, 40% at Time 4, and 12% of the variance in GJT at Time 2. This suggests that the SER construction was harder to proceduralize even for the group that received GE both prior and during practice. In addition to WM, LLAMA also significantly predicted performance on the production test at Time 2.

Finally, while Group 3 [GEb-GEd+] depended on WM to utilize their declarative knowledge during the second and third session, performance during the delayed posttest was not contingent upon their WM.

H4a

H4a: Language learning aptitude associated with explicit learning will play a bigger role in the initial stages of skill acquisition, in particular in the formation of declarative knowledge.

Evidence for this hypothesis comes from the significant correlations between LLAMA and the outcome measures, as well as the significant regression models predicting the outcome measures from LLAMA. This prediction is largely borne out, but the pattern varies across groups. While for Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], in which the strongest proceduralization of the two grammatical structures was observed, LLAMA only explained a significant portion of the variance in the outcome measures during Time 2, for Group 2 [GEb+GEd-] LLAMA only correlated during the delayed posttest. As mentioned earlier, LLAMA explained most of the variance in the performance of Group 4 [GEb-GEd-] and was not a significant predictor for Group 3 [GEb-GEd+]. It can be therefore concluded that when the right conditions for proceduralization are in place, as was the case for Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], language aptitude for explicit learning has the strongest effect in the early stages of the process of proceduralization. The relationship between LLAMA and the performance of Groups 2 [GEb+GEd-] and 4 [GEb-GEd-] is discussed later.

H4b

H4b: Language learning aptitude associated with implicit learning will play a bigger role at the later stages of skill acquisition, in particular during the stage of final proceduralization.

This hypothesis was not supported. It was shown that SRT, an indicator of language learning aptitude for implicit learning, did not have any effect on the outcome measures in this experiment. It was initially proposed that language learning aptitude for implicit learning would have an impact in the later stages of proceduralization, where implicit learning mechanisms may play a role. One of the limitations of the current experiment is that it did not provide enough practice sessions for participants to engage in the later stages of proceduralization. This effect was especially visible for the SER structure, which proved to be harder to proceduralize than the OV one. Therefore, since participants were only engaged in the very early stages of proceduralization, it is understandable that their performance was not contingent on implicit learning ability especially given the explicit nature of the learning at hand. However, this does not imply that implicit learning aptitude does not have any impact on learning under different conditions (with less focus on forms) or even at later stages of learning under the conditions of the present experiment. Furthermore, the low reliability of the SRT task necessarily limits its predictive validity.

H4c

H4c: General working memory capacity will play a bigger role at the initial stages of skill acquisition, in particular in the initial stage of proceduralization. In particular, the effect of working memory will be most visible in the GEd- conditions.

This hypothesis is supported from the current data. First, the impact of WM depended on the extent to which participants' knowledge was proceduralized or whether there was declarative knowledge to be proceduralized at all. In the case of Group 3 [GEb-GEd+], which started with proceduralization of the OV structure even after session 2, neither LLAMA nor WM played any role. In contrast, for the SER structure, whose proceduralization was harder than the OV one, WM also played a significant role in the groups that had GE during practice. For Group 3 [GEb-GEd+], WM predicted performance during Time 2 and 3, while for Group 1 [GEb+GEd+] WM predicted performance during the delayed posttest as well, suggesting that this group had to rely on their declarative knowledge to perform during the delayed posttest.

The effect of WM was most visible, however, in the performance of Group 2 [GEb+GEd-], where participants had to keep the rules in mind while performing the practice tasks. In line with SAT's predictions, WM was a consistent and significant predictor for both structures for all measures during some of the times for OV and for all of the times for SER in the performance of Group 2 [GEb+GEd-].

Finally, WM did not have any impact on the performance of Group 4 [GEb-GEd-]. Participants in this group were not given any declarative rules and consequently did not need WM while performing on the two structures.

To summarize, the treatment that Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] received reduced the need for participants to rely on their WM to remember the declarative rules, as GE was provided for each of the practice items individually. In the case of OV, which was easier to understand, WM did not play a consistent role in these groups. In the case of SER, which was harder to grasp, providing GE during practice was not enough and participants had to rely on their WM in order to accurately use the GE during testing. The treatment that Group 2 [GEb+GEd-] received capitalized on participants' WM ability and consequently the most visible correlation with WM was observed in Group 2 [GEb+GEd]. The strong effect of WM in the group that received GE prior to practice only is reminiscent of the literature in educational psychology emphasizing that "any instructional treatment that ignores the limits of working memory when dealing with novel information (…) is unlikely to be effective" (Kirschner et al., 2006, p.77).

H4d

H4d: The effect of explicit language aptitude will be most visible in the group that will not receive any GE, i.e., GEb- GEd-.

The present study provides strong evidence in support of this hypothesis. The effect of LLAMA was the strongest and most consistent in the performance of Group 4 [GEb-GEd-]. Results from the simple linear regressions demonstrated that for the OV structure, LLAMA significantly explained 47% of the variance in comprehension at the delayed posttest, 20% in the production during Time 3, 13% during Time 3, 22% in GJT Time 2, 20% during Time 3, and 20% during the delayed posttest. For performance on the SER structure, LLAMA significantly explained 24% of the variance in comprehension during Time 2, 24% during Time 3, 32% in production during Time 2, 16% during Time 3, and 15% in the performance on the GJT during Time 2. In line with previous results, in particular, Erlam (2005), participants who were not guided with GE before or during practice had to rely on their language learning aptitude to find patterns in the data. This complex relationship is very clearly illustrated in the interaction plots in the results section and quantified in the results from the regression models. The ATI pattern observed in the present study, and in other previous ATI studies, is exactly what is predicted in educational psychology with regards to the ATI: when learners are minimally guided during the process of learning, the success of rule discovery is contingent upon the individuals' language learning abilities (e.g., Kirschner et al., 2006). In other words, the more the burden is put on the learner, the greater the need for language aptitudes.

Chapter 7: Conclusions

The following will provide an overview of major findings, discussing them in light of previous theoretical and empirical research in the field of SLA and CP. Limitations of the study will then be discussed, and finally further directions will be outlined.

Summary of findings

There are three main findings from the present experiment: (1) the importance of right rule and example integration, (2) aptitude-by-treatment interaction, and (3) the specificity of practice. Each of these points will be discussed in turn.

First, the present experiment showed the importance of the manner in which GE was made available to the learners. Across the two structures, the different timings of GE that Groups 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+] received were not distinct enough to lead to differences among these groups during the delayed posttest. It was, however, observed that only the treatment received in Group 1 [GEb+GEd+], managed to outperform Group 2 [GEb+GEd-], for SER during Time 3. Looking at the results holistically, it seems that no matter when GE is provided, learners that receive GE will fare better than the ones who are left on their own to find grammatical patterns from the input. This was especially true for the knowledge participants acquired that was necessary for accurate rejection of ungrammatical OV GJT items during Time 3, and for ungrammatical SER items during both Time 3 and Time 4. Task-essential practice only as was the case for Group 4, whether for OV or SER, was not enough for the formation of solid declarative and proceduralized knowledge, and accurate performance largely hinged on language aptitude for this group. The different timings of GE did have an impact, however, on which of the GE+ groups was significantly superior to Group 4 [GEb-GEd-] in the sense of the extent to which they retained the acquired knowledge, and how the group difference interacted with the ID measures.

One key component in achieving accurate and durable performance, at least after one week, involved using GE to acquire procedural knowledge of the OV and SER grammar structures, rather than relying on memory for declarative rules or including highly structured grammatical patterns within a task-essential practice. Results demonstrated that providing rules and examples in an unstructured way is insufficient for explicit learning to yield its most beneficial effects. A second crucial factor was for the declarative rules to be presented in just-in time fashion. This specific combination of top-down and bottom-up processing yields the most beneficial results indeed, and is the least susceptible to individual differences[3]. All of our results indicate that participants whose practice was assisted by providing GE while being engaged with practice activities were more likely to perform successfully and retain their knowledge than participants who received GE only prior to practice or participants who did not receive any GE. We argue that the superiority

[3] It should be noted that the terms top-down and bottom-up are used in a slightly different sense here from the way they are typically used in psycholinguistics. While in psycholinguistics top-down refers to interpretation of a larger syntactic unit based on what is available to the learner and bottom-up processing refers to moving from smaller units such as phonemes to larger units, we refer to top-down processing as going from rules to examples, and bottom-up processing as going from examples to extracting grammar rules.

of the GEd+ groups is due to the fact that participants in these groups were provided with declarative rules every time they had a real need to use those rules, applied their declarative knowledge on regular basis, and consequently were more successful at proceduralizing it. In other words, the specific integration of top-down and bottom-up learning in the GEd+ groups set up the optimal conditions for proceduralization. This finding is relevant in several aspects.

This finding provides support for the theoretical assumptions of SAT that predict the most successful proceduralization when solid declarative knowledge is formed and when learners are engaged in activities capitalizing on that declarative knowledge (DeKeyser, 2007b, 2015; Sallas et al., 2007). Participants in Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] formed their declarative knowledge by being guided through the comprehension activities based on the explicit grammar rules. They received the metalinguistic declarative knowledge exactly at the time when they needed it to solve a problem and when they were ready to use it. As demonstrated from the MTK descriptive statistics, it was exactly Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] that had higher means than the other groups, providing support for studies in cognitive psychology which highlight the importance of developing declarative knowledge through memory for examples of how procedures should be executed (Anderson & Fincham, 1994; Anderson, Fincham, & Douglass, 1997; Taatgen & Wallach, 2002). In particular, Anderson, Fincham, & Douglass (1997) demonstrated that the most optimal transition from the declarative to proceduralized stage is enabled when practice combines examples and declarative rules throughout the stage of proceduralization.

Next, the study's finding that GE during practice is superior to GE prior to practice or no GE at all is also valuable in extending the findings supporting the integrative approach found in Sallas et al., (2007). In that study of artificial grammar learning, fast and accurate performance was only achieved by participants who received an animated version of the artificial grammar diagram just as it was needed during practice. The success of the GEd+ groups in the current study suggests that their results as well as the theoretical assumptions of SAT have validity for the acquisition of L2 Spanish structures, at least as measured by the three outcome measures.

As mentioned earlier, these results, then, are most supportive of the notion that not only are both top-down (rules) and bottom-up (examples) processes essential for the development of accurate L2 grammar knowledge; but that it is also important how and when these processes are integrated. Learning models that predict acquisition of grammar knowledge through bottom-up learning alone, i.e., from exposure to practice or examples, do not fit the patterns observed in our data. Rather, with extensive practice that combines top-down and bottom-up processes, higher-level grammar knowledge is more likely to be acquired. One can justifiably argue, however, that the practice participants received in the present study was not nearly as extensive as needed for such higher-level grammatical knowledge to be developed, and future studies should try to address this issue by allowing more training time.

Numerous SLA researchers have argued that the course of language development is parallel to the course of acquisition of any cognitive skill (de Jong & Perfetti, 2011; DeKeyser 1997, 1998, 2001, 2007a,b; Lyster 1994, 2004, 2007; McLaughlin 1990; Ranta & Lyster 2007; Lyster & Sato, 2013). The results from the present study not only attest to these claims but also complement previous findings in that it does not suffice for GE to be provided at any time during the learning process or for GE to be formed by learners by means of induction. Instead, the provision and timing of GE was demonstrated to be a crucial factor for proceduralization.

The issue of timing of GE has largely been neglected in empirical investigations, despite the recurrent emphasis on its importance with regards to explicit learning. Studies employing GE have not been systematic in how and when GE was provided, making it hard to isolate the specific features of explicit learning that lead to learning gains. While numerous studies attest to the superiority of deductive explicit learning over incidental, inductive and implicit learning, they fall short of providing solid explanations for what exactly facilitated their superiority. In the same vain, while there are studies that show a non-facilitative effect of explicit instruction when participants are exposed to task-essential structured input, the explanations of such findings do not do justice to the issue at hand. The present findings offer a window into the conditions under which explicit learning bears the most fruitful results as well as the conditions that do disservice to explicit learning.

This brings us to our next point: the findings from the present study offer some tentative explanations about why in certain cases providing explicit information did not add any additional benefit over and beyond what was acquired from task-essential structured input. As mentioned in the literature review, it is mainly the PI studies that have reached this conclusion (e.g., Benati, 2004; Farley, 2004a; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996; Wong, 2004b). All of these studies tried isolating the effect of explicit information by having several experiment groups. For instance, in Benati (2004) there were three groups: a PI group, which received GE followed by structured input, an

SI group, which only received the structured input portion, and an EI group, which only received the same GE, but did not receive any structured input. Based on predictions from SAT and previous experimental findings, as well as findings from the present study, we know that the combination and distribution of rules and examples is one of the key factors for successful proceduralization and hence the learning of target structures. It has been repeatedly emphasized and empirically demonstrated that providing explicit rules without the opportunity to apply those rules in meaningful contexts is simply not enough. Similarly, providing GE only prior to practice, pre-emptively, does not seem to bring any additional facilitative effect when compared to exposing learners to task-essential practice. On the contrary, such pedagogical treatments place a heavy burden on learners' WM and thus would only benefit high WM learners. We can say, therefore, that Benati's (2004) PI groups, but especially the EI group, were at a disadvantage from the start. Taking all of this into consideration, it is not surprising that the PI or the EI group did not outperform the SI-only group. The comparison between the groups does not do justice to the problem at hand. There is a resemblance between the treatments in Benati (2004) and the ones in the present study: the PI group received the same treatment as our Group 2 [GEb+GEd-], and the SI group received the same treatment as our Group 4 [GEb-GEd-]. Had we only included these two groups, we would have arrived at the same conclusions. However, we have to depart from such a conclusion, especially as the results from the other deductive groups did show a more sustainable and superior effect than the group that only received task-essential practice. In this sense, the comparison between the present findings and previous research on the role of GE within the PI framework gives us an additional insight into how learning can proceed if the right conditions for explicit learning are not met.

These findings can also provide additional insight into some observed discrepancies in previous results. For instance, while Rosa & O'Neill (1999) did not find any difference between the instructed and rule-search groups, Robinson (1996) demonstrated that the instructed condition in his study gained significantly higher accuracy than the rule-search one. One seemingly subtle difference between these studies, among others as well, was that the instructed participants in Rosa & O'Neill were only given the rules prior to practice, and participants in Robinson's (1995 and 1996 study) were allowed to consult the rules even during practice. In fact, Rosa & O'Neill (1999, p.539) mention that "had explicit instruction been more crucial either in their study (VanPatten & Oikkenon, 1996) or in ours, perhaps more significant differences would have been found".

Our findings also provide some answers with respect to the role of pre-practice GE. In line with Presson et al. (2014) and Stafford et al. (2011), it seems that the results from the current experiment do not provide any evidence for an additional benefit of providing GE prior to practice. None of the measures for either of the two structures demonstrated a superiority of Group 1 [GEb+GEd+] over Group 3 [GEb-GEd+]. In fact, evidence for the opposite can be gleaned from the results showing a steady development and retention of the acquired knowledge for participants in Group 3 [GEb-GEd+], as well as no relationship with LLAMA and WM. In contrast, the performance of Group 1 [GEb+GEd+] on the OV structure was partly explained by LLAMA for the GJT test at Time 2, as well as WM for the comprehension at Time 3 and GJT test at Time 4. In addition, for the SER structure, WM was a significant predictor in the comprehension and production measures during the delayed posttest.

Second, the present study provides a complex pattern of ATI. In line with predictions from SAT and educational psychology, the experimental treatments that provided minimal guidance to the learning problem, such as the GEd- groups, were the ones that drew most on LLAMA and WM. Participants in Group 2 [GEb+GEd-] were only given the rules before the practice problems and were not reminded about them throughout the training session. Participants had to keep the GE in their WM in order to solve the comprehension problem activities. As a result, those who had higher WM benefitted more from this treatment than those individuals with lower WM. Similarly, participants in Group 4 [GEb-GEd-] were not provided with any GE but were only engaged in a task-essential practice that included immediate feedback in the form of a correct or incorrect response, and it was demonstrated that successful induction and performance for this group depended highly on participants' language learning aptitude. It can be said that this treatment was somewhat beneficial, relative to the GEd+ groups, for high aptitude learners, but not for low aptitude learners. The high involvement of WM and LLAMA in the two GEd- groups was observed for both structures at all stages of the learning process.

Language aptitude and WM were also significant predictors for the GEd+ groups, but to a lesser degree. Their importance for these treatments seems to depend on the stage of acquisition. For the SER structure, which was harder to proceduralize, WM predicted performance for both Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] throughout the learning process. However, for the OV structure, the effect of WM was mainly present in the delayed posttest for Group 1 [GEb+GEd+]. In this group it was found that WM predicted performance during the later stages of learning. If the correlations with WM were driven by learners engaging in recall of the rules, that suggests that participants in this group still needed the declarative rule at the delayed posttest even after two sessions of training. On the contrary, for Group 3 [GEb-GEd+], WM was involved in the learning process of the SER structure and only during the learning process (Time 2 and 3), but not at the one-week delayed posttest. With respect to language aptitude and the GE+ groups, participants in Group 1 [GEb+GEd+] relied on their language analytic ability to consolidate their declarative knowledge mainly during the first session of training, but not participants in Group 3 [GEb-GEd+]. LLAMA was not predictive of Group 3's [GEb-GEd+] performance for either of the structures. The interaction results demonstrated, however, that learners with high language aptitude for explicit learning benefitted the most from the most explicit treatments, i.e., Group 1 [GEb+GEd+] and 3 [GEb-GEd+], suggesting how important GE is even for learners with high apatitude for explicit learning.

In almost all respects, then, our hypotheses concerning the interactions between our different experimental treatments and IDs are fully confirmed. The hypotheses were motivated by theoretical assumptions and empirical findings that hinge on the premise that some pedagogical interventions require a certain mental capacity that is facilitated by a particular ID variable (DeKeyser, 2012). For instance, the treatment that Group 4 [GEb-GEd-] received required participants to figure out rules by themselves in highly structured sets of examples meant to facilitate induction. The mental process that leads to successful learning from such a treatment is analytical aptitude, a finding which was consistent in our results for both of the structures.

Our ATI finings extend conclusions reached in previous studies, in particular Erlam (2005). In both Erlam (2005) and the present study the deductive treatment outperformed the inductive treatment, and language learning aptitude was observed to predict performance in the inductive but not the deductive treatment to the same extent. The current study departs from Erlam (2005) on one important point, however. One of the main research questions was whether the timing of GE is an important factor for acquisition and retention of knowledge, and while the present study did not find any significant difference among the deductive groups (Group 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+]), it demonstrated nevertheless that the behavior of these groups is contingent on the delivery of GE and modulated by individual differences. It was found that not all deductive groups were significantly different from the inductive group, but only those that received GE during practice, setting these two groups apart from the group that only received GE prior to practice. Results demonstrated that for neither of the structures did Group 2 [GEb+GEd-] outperform Group 4 [GEb-GEd-], except as observed in performance of the ungrammatical SER items. It seems that deductive treatments are superior to inductive ones, but only provided that GE is available throughout the process of proceduralization.

In addition, not all three deductive groups demonstrated the pattern found in Erlam (2005) with respect to individual differences. Erlam (2005) found that it was the deductive treatment that showed no relationship with language aptitude and WM, concluding that presenting grammar rules deductively to learners has an equalizing effect on ID, in particular, language aptitude and WM. The results of this study corroborate this finding, but also complement it in that only the treatment in the GEd+ groups seems to act as an equalizer of ID, as no consistent relationship was observed between the outcomes for Groups 1 and 3 on one hand, and WM and LLAMA on the other hand. As mentioned earlier, a significant relationship between WM and the outcome measures, however, was observed in Group 2 GEb+GEd- for both structures. It can be concluded, therefore, that deduction outperformed induction in the present study and proved less dependent on language aptitudes, but only provided that GE was presented to the learners during practice.

The ATI patterns of the present experiment are largely in line with previous SLA findings: Hauptman (1971) demonstrated that inductive treatment is more beneficial for high-aptitude learners, Wesche (1981) demonstrated that the effectiveness of different pedagogical treatments impacted learners with varying aptitude profiles differently; Robinson (2005) and Erlam (2005) found that language aptitude predicted performance when learning was more deductive.

Previous research in SLA has manipulated various factors for focusing learners' attention to form, ranging from input flooding, input enhancement, recasting, structured input, just to name a few. Results have been mixed. As Robinson (2005) points out, one plausible explanation for these contrasting results stems from the fact that based on individual differences in language aptitude and working memory, some learners may benefit more from one instruction treatment than others. In this respect, this study's findings further emphasize the importance of including IDs in any investigation of language learning. Without the language aptitude and WM variables in this study, explanations of the reasons for the superiority of one treatment over another would only be possible at the level of speculation. As highlighted by DeKeyser (2012), aptitude-treatment interaction research not only shows the importance of an aptitude or a treatment, but more importantly shows *why* treatments are sometimes beneficial and other times not.

Third, the current study provides indirect evidence for the directional asymmetry of proceduralized rules. According to SAT, while declarative knowledge is generalizable to various situations and its form can be used equally well for any skill, proceduralized knowledge is highly skill-specific. Results from the comprehension and production measure revealed that for the OV structure participants across groups had proceduralized their declarative knowledge, as demonstrated by their inability to transfer this knowledge from comprehension to production. The knowledge of the SER structure, however, could be used to a comparable extent in both comprehension and production, suggesting that it was declarative and procedural knowledge that participants were using when performing on the SER items. Consequently, the directional asymmetry was observed for the OV but not for the SER structure. For the OV structure, all groups had higher means on the comprehension than on the production. In contrast, the not-fully proceduralized knowledge of the SER structure could be equally used for performance on the comprehension and production tests across groups. In addition, we can also assume that when participants employ their declarative knowledge, as opposed to procedural, the involvement of WM will be bigger. It was exactly this pattern that was observed and shown in the correlation matrices: WM is more predictive for SER than for the OV structure.

Implications

DeKeyser (2012) has argued for the importance of investigating the process rather than the product of learning. One avenue for such investigations is made possible by taking into account the interactions between aptitudes and treatments. In this regard, the ATI results from the current experiment not only provide us with such insights into the learning process, but also provide us with important educational implications.

Our results show that the involvement of language aptitude and WM not only depends on the task demands inherent in the four experimental conditions, but also depends on the stage of acquisition. In cases where an individual's strength in language aptitude or WM matched the task demands of the experimental conditions, superior levels of learning were achieved on all three outcome measures. While matching students' individual preferences for learning from a specific instructional treatment may be the ideal situation, an alternative and less costly possibility may be to utilize a treatment that is effective irrespective of individuals' language aptitude or working memory capacity. The present study offers evidence that, at least for the present target structures, outcome measures and population, providing learners with grammar rules at the time of practicing, when the need for those rules can be most strongly felt, and when learners are most ready to utilize them, neutralizes any differences in language aptitude for explicit learning and working memory.

Although among language instructors there is a preference for pedagogical interventions that capitalize on the communicative use of language, there might be advantages, in particular for structures that pose serious acquisitional problems even for highly advanced learners, for explicit instruction that focuses learners on common strategies that facilitate or hinder the acquisition of grammar structures. In fact, our results clearly demonstrate that depriving learners from receiving GE, even in situations where practice training maximized chances of induction by having highly structured task-essential practice, was doing them a disservice. It can be concluded, therefore, that the treatment employed in the GEd+ groups provides a viable intervention for targeting problematic L2 grammar structures. One concern when employing this pedagogical treatment, however, is the extent to which participants really pay attention to and read the GE during practice. In the current experiment, some evidence exists that participants who were presented with GE both before and during practice may not have used the opportunity to reread and rely on the written GE during practice to the full extent.

In addition, previous empirical findings have provided positive evidence for the beneficial effect of explicit instruction, and such practices are common within typical pedagogical materials. So far, only inadequate and rare accounts in SLA exist that try to explain the particular components of explicit instruction and the right conditions under which explicit instruction is to be effective. This study's results offer tentative answers with regards to this question. It was observed that even though Group 2 [GEb+GEd-] received GE prior to practice, the timing was not ideal for accurate performance. Instead, providing GE during practice generating significantly higher gains on all three outcome measures and proved to be rather independent of language aptitude and WM, at least for the syntactic structure that was easier to proceduralize.

Limitations and Future Directions

One of the major limitations of the present study is the nature of the outcome measures. It can be said that our three outcome measures are all biased towards drawing on explicit knowledge and are not communicative in nature. Reaching high accuracy on the outcome measures employed in this study is by no means equal to obtaining native-like command and communicative competence of the L2 grammar structures. Future studies should employ a variety of outcome measures that do not primarily focus participants' attention on their explicit knowledge and that are not as restrictive in nature as was the case with the fill-in the blank production measure in the current experiment. An important question is whether the gains of the experimental conditions and the superiority of the GEd+ groups will transfer to measures such as elicited imitation, oral production and oral interviews.

It can also be argued that the training participants received was too little and did not present fair and equal opportunities for all groups to acquire the two target structures. This is particularly true for Group 4 [GEb-GEd-], as it is well established that implicit learning is slow and laborious and thus participants in these conditions take more time to arrive at a correct representation of grammatical structures. Had participants received more training sessions and had the outcome measures relied less on explicit knowledge, the observed difference between and superiority of the GE+ over GE- groups may have either diminished or taken a different turn, such that the more implicit treatment received in Group 4 [GEb-GEd-] would have shown more durable and stable gains than the other groups.

In the same vein, it is very likely that results may have taken a slightly different form had several delayed posttests been employed. This is especially true in light of previous findings by Li (2010) that demonstrated that explicit feedback yielded better results than implicit feedback on immediate and short-delayed posttests, but implicit feedback was more effective than explicit feedback in the long run. Future studies should, therefore, examine whether the GEd+ superiority to GEd- groups will in fact be observed when longer practice is provided and when participants' acquired knowledge is gauged by several delayed posttests.

Another limitation of the present study is the inability to guarantee that participants in the GEd+ conditions indeed read the GE for every practice problem. To eliminate speculation as to why Group 1 [GEb+GEd+] was not found to be superior to Group 3 [GEb-GEd+] both in terms of accuracy and in terms of the interaction with language aptitude and WM, further research should employ an online measure of participants' attention. Eye-tracking methodology would be particularly helpful in this regard as it will enable researchers to map accuracy scores to proportion of time spent reading the GE during practice. A separate question that is related to the treatment that Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] received is whether participants in these groups were drawing on their memory of GE or relearning the GE each time they were presented. This potential confound is built into the current experimental design and therefore it makes it hard to conclude that participants in these groups were not learning the same GE more thoroughly than participants in the other groups. Indirect evidence can be gleaned from the involvement of WM even in the performance of these two groups, supporting the claim that participants were not relearning the GE. However, more direct evidence would be to examine participants' performance during practice with online measures such as think-aloud protocols, as well as to investigate the extent to which participants' performance during practice hinges on WM.

Another outstanding question regarding the treatment received by the GEd+ groups would be to investigate a range of L2 structures and expose participants to all of them at the same time, instead of receiving blocks of training targeting only one feature. That would mirror more natural L2 acquisition and therefore provide more generalizable results.

Finally, the individual differences measures employed in the present study are not exhaustive, and future studies should employ several tasks tapping into working memory ability and language aptitude for explicit learning but also control for other cognitive individual differences that may interact with learning, such as the direct and indirect language learning strategies explained in Oxford (1990). Of those, metacognition, the awareness based on previous experience, is of a particular interest as studies have demonstrated that metacognition has an indirect effect on cognitive processing (Purpura, 1997); therefore future studies should look not only at participants' beliefs about learning and their impact on the actual learning, but also how their metacognitive processing interacts with their aptitude for explicit language learning.

In conclusion, the findings of the present experiment extend the scope of the literature on explicit L2 learning and aptitude-by-treatment interaction. In an experimental examination of four pedagogical treatments, which manipulated the provision and timing of GE on the acquisition of two target Spanish structures, it was demonstrated that two crucial factors maximize learning gains:

(a) top-down and bottom-up learning are both necessary, and (b) the way these learning processes are integrated makes a substantial difference for the learning of L2 grammar. Participants who received GE in a just-in-time fashion were more likely to perform and retain their knowledge successfully. It was additionally demonstrated that the various combinations of top-down and bottom-up processes imposed differential task demands on the learners and consequently drew on language aptitude and WM to a different extent. Since these interactions were the least observed in the conditions that received GEd+, we argue that the right integration of rules and practice ameliorated task demands that were burdensome for the learner, and thus mitigated the effect of participants' individual differences. Finally, some evidence was also observed that the comprehension practice participants received for the two structures was not enough for the formation of solid productive knowledge of the same. All of these conclusions, however, are tentative inasmuch as the outcome measures that were employed were explicit in nature and as such do not allow for generalizability of findings to language acquisition in general.

Appendices

Appendix A: Sheet of EI

Information about Ser/Estar in Spanish

Please carefully read this information because comprehension questions will follow.
English verb 'to be' has several equivalents in Spanish: Among them are SER and ESTAR. Generally, we use SER when we talk about inherent qualities. By inherent qualities we mean traits that are built-in, ingrained, an essential part of how someone or something really is. For example:
"El hombre ES serio"
The usage of SER in the example above indicates that the man is a serious individual. This is a part of his personality. This person is not prone to being boisterous, or frivolous. Rather, his usual demeanor is sober and stern.

In contrast to SER, ESTAR is used to express traits that are true in a particular circumstance. Although the trait may not be a part of the personality of the individual at hand, it happens to describe the state of the person under a particular circumstance. For example:
"El hombre ESTÁ serio"
The usage of ESTAR in the example above indicates that the man is serious under the present circumstances. Thus, ESTAR is generally used to describe circumstantial states, not a defining or inherent quality.

IMPORTANTLY ... because in English the verb to be is used for both inherent and circumstantial conditions, American leaners of Spanish often tend to confuse the two. In the following activities it will be very important for you to look at the verb (ES or ESTÁ) in order to know whether the sentences you read refer to an inherent (ES) or circumstantial (ESTÁ) trait.

Information about OVS in Spanish

Please carefully read this information because comprehension questions will follow.
Consider sentence 1) below:

1) Mónica compra un perro

In that sentence we could replace 'un perro' the following way:

2) Mónica LO compra

Because Spanish has flexible word order we can also have the following sentence:

3) LO compra Mónica

HIM buys Mónica (or *Mónica buys him*, if put in the English order) As you can see, 'LO compra Mónica' literally means '*HIM buys Mónica*' (NOT '*HE buys Mónica*', mind you!) and although this sentence is not possible in English it is both possible and very common in Spanish. PLEASE NOTE that while in 2) 'Mónica LO compra' the first word in the sentence (i.e., Mónica) is the DOER of the action of buying, sentence 3) 'LO compra Mónica' starts with the VICTIM of the buying. In grammar we refer to the DOER as the SUBJECT and the VICTIM as the OBJECT.

IMPORTANTLY because sentences in English can start with the DOER (subject) only, Americans tend to process sentences such as 4) incorrectly as shown below:

4) LA visita Juan

is processed as ... *She visits Juan* ** WRONG *'She visits Juan'* would be 'Ella visita a Juan'. Note that 4) says 'LA visita' and NOT 'Ella'. 3) 'LO compra Mónica' is processed as ... *He buys Mónica* ** WRONG *'He buys Mónica'* would be 'Él compra a Mónica' and as you see 3) says 'LO compra' and NOT 'Él'. In the next tasks it will be crucial to remember that LO is different from ÉL and LA is different from ELLA. LO and LA stand for the VICTIM/OBJECT of buying, visiting or whatever the verb, whereas ÉL and ELLA designate who does the buying, visiting, etc.

Appendix B: Comprehension questions following the EI slides

Comprehension questions about Ser/Estar

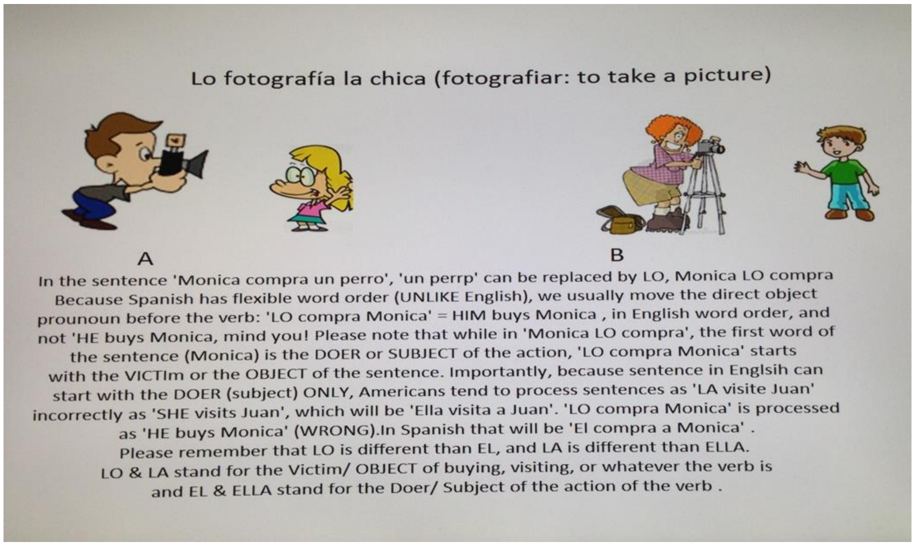
1. English TO BE has two equivalents in Spanish, SER and ESTAR. SER is used for inherent, built-in traits.   A) TRUE B) FALSE
2. SER and ESTAR can be used interchangeably. A) TRUE B) FALSE
3. The information as to whether a trait is inherent or not is found is the verb. 'Es' indicates a inherent trait and 'está' indicates a trait that is caused by an external circumstance. A) TRUE B) FALSE
4. 'El hombre ESTÁ serio' means that being serious is part of the man’s personality. A) TRUE B) FALSE
5. El hombre ESTÁ serio' means that a particular circumstance is causing the man to be serious but his personality is not necessarily that way. A) TRUE B) FALSE
6. 'Inherent' and 'circumstantial' mean the same thing. A) TRUE B) FALSE

Comprehension questions about OVS

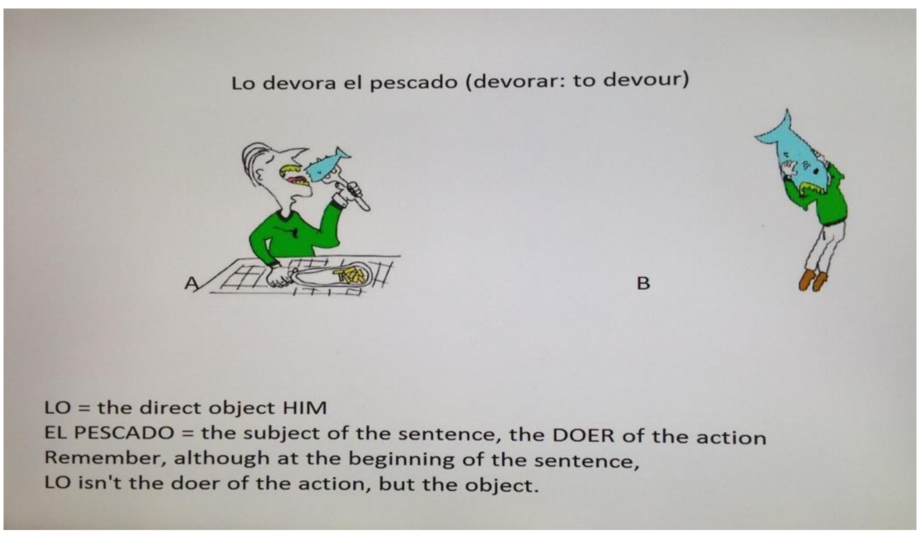
1. In grammar we refer to the DOER of an action expressed by a verb (eg., buying) as the SUBJECT and the VICTIM or recipient of the action is the OBJECT
A) TRUE B) FALSE
2. LO and LA designate the VICTIM/OBJECT, the recipient of the action in the verb but ÉL and ELLA are for the DOER/SUBJECT of the action in the verb. A) TRUE B) FALSE
3. 'Lo admira Sonia' means 'He admires Sonia' A) TRUE B) FALSE
4. 'Lo admira Sonia' means 'Sonia admires him' A) TRUE B) FALSE
5. In 'LA inspira Raúl' the first word 'LA' designates the recipient of the inspiration A) TRUE B) FALSE
6. LO and LA designate the VICTIM/OBJECT, the recipient of the action in the verb but ÉL and ELLA are for the DOER/SUBJECT of the action in the verb. A) TRUE B) FALSE

Appendix C: Sample training items for Groups GEd+

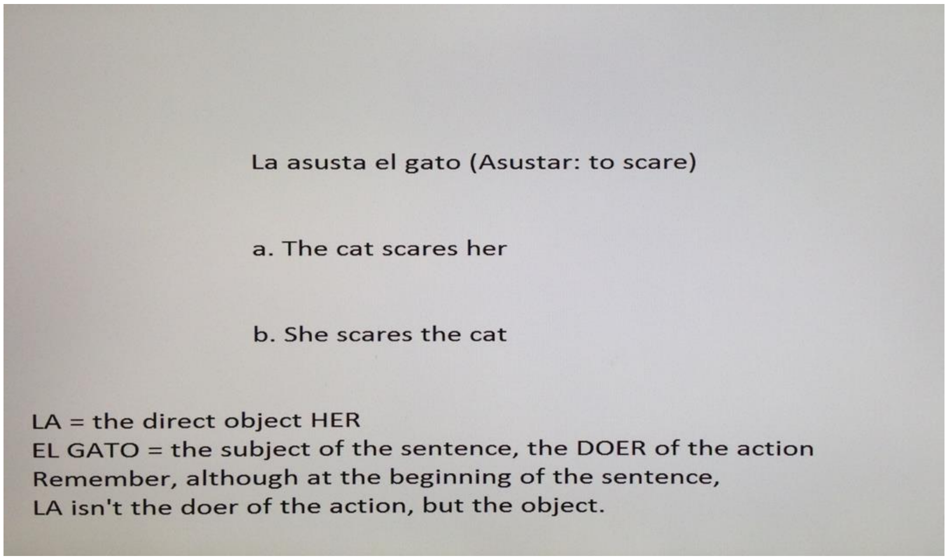
First practice slide for the picture-matching task



EI on all subsequent practice items for the picture-matching task

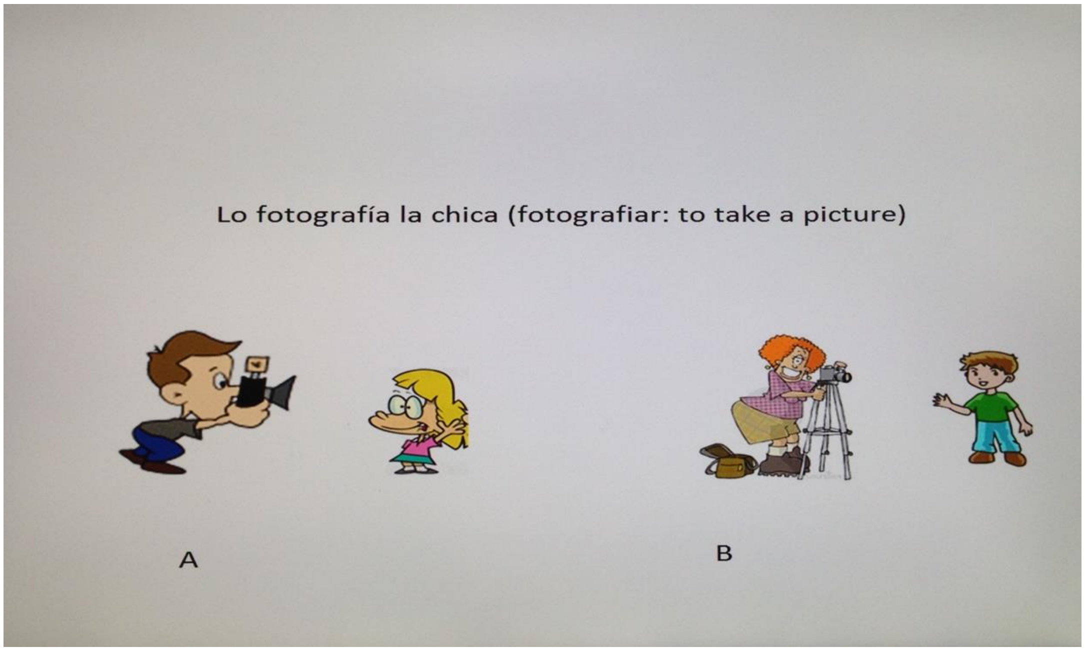


EI on all subsequent practice items for the sentence-interpretation task

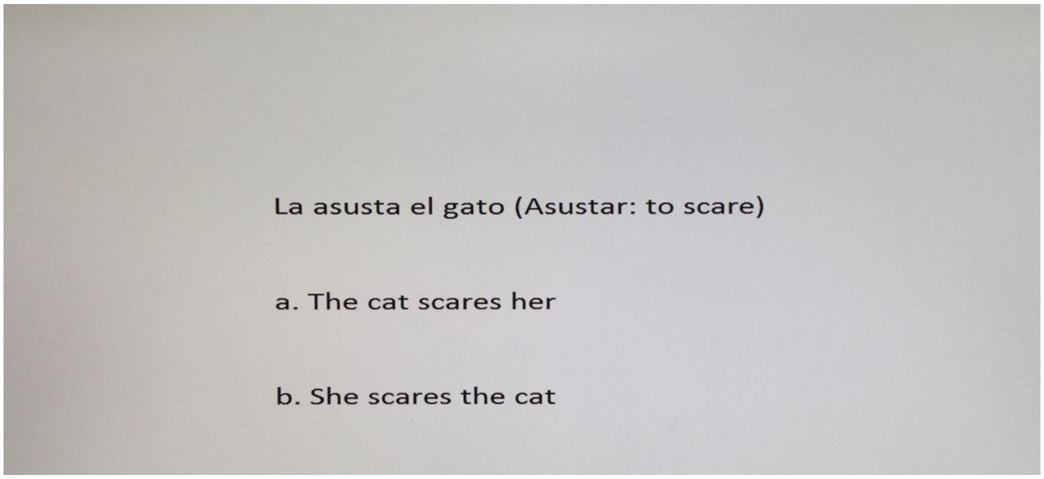


Appendix D: Sample training items for Groups GEd-

Practice items for the picture-matching task



Practice items for the sentence-interpretation task



Appendix E: ANOVA models for comprehension, production and GJT

OVS

Data were first analyzed with Repeated-Measures ANOVA with Group as a between-subject factor and Time as a within-subject factor. Repeated-measures ANOVA for the comprehension measure with Time as a within-subject factor and Group as a between-subject factor indicate that there was a significant main effect of Time, $F(2.8, 288.4) = 337.32$, $p<.001$, no significant main effect of Group, $F(3,102) = 2.21$, $p=.09$, and no significant interaction between Time and Group, $F(8.4, 288.4) = .949$, $p=.48$. Sphericity was violated so the Huynh-Feldt correction was used. Following the significant main effect of Time, pairwise comparisons using the Bonferroni adjustment for multiple comparisons, indicate that there were no significant differences between the groups at Time 1, i.e., the pretest. For Group 1 [GEb+GEd+], Time 2 was significantly better than Time 1 (MD=.55, $p<.001$), and Time 3 was significantly better than Time 1 (MD=.638, $p<.01$) and Time 4 (MD=.135, $p<.01$). Time 4 was significantly better than Time 1 (MD=.5, $p<.001$), but significantly worse than Time 3 (-.135, $p<.01$). For Group 2 [GEb+GEd-], performance during Time 2, 3 and 4 was significantly better than Time 1 (MD=.625, $p<.001$; MD=.665, $p<.001$; MD=.554, $p<.001$). In addition, performance during Time 3 was significantly better than Time 4 (MD=.112, $p<.03$). Group 3 performed significantly better at Time 2, 3,and 4 than at Time 1 (MD=.51, $p<.001$; MD=.537, $p<.001$; MD=.443, $p<.001$), with no significant differences between Time 4 on one hand and Time 2 and 3 on the other hand (MD=-.068, $p=1$; MD=-.095, $p=.09$), suggesting that they have retained their knowledge at the delayed posttest. Finally, Group 4 performed significantly better at Time 2,3 and 4 than Time 1 (MD=.631, $p < .001$; MD=.648, $p < .001$; MD=.517, $p < .001$). Similarly to Groups 1 and 2, Group 4 performed significantly worse during Time 4 in comparison to Time 3 (MD=-.131, $p = .01$).

The Repeated-Measures ANOVA model for the production data demonstrated that there was a significant main effect of Time, $F(2.14,227.41)= 22.63$, $p<.001$, non-significant main effect of Group, $F(3,106)= 1.44$, $p=.235$, and a non-significant interaction between Time and Group, $F(6.43, 227.42)= 1.69$, $p=.12$. Sphericity was violated so the Huynh-Feldt correction was used. Group 1 performance at Time 2, 3 and 4 was significantly better than Time 1 (MD=.172, $p = .05$; MD=.278, $p < .001$; MD= .178, $p<.01$). Group 2 performance at Time 2 and 3 was significantly better than at Time 1 (MD= .26, $p < .001$; MD=.352, $p<.001$). In addition Group 2 [GEb+GEd-]performed significantly worse on Time 4 in comparison to Time 3 (MD = .267, $p=.013$). Group 3 [GEb-GEd+] performance at Time 3 and Time 4 was significantly better than Time 1 (MD=.171, $p = .02$; MD=.231, $p < .001$). Group's 4 [GEb-GEd-] performance at Time 2, 3 and 4 was significantly better than Time 1 (MD = .214, $p = .005$; MD=.270, $p < .001$; MD= .171, $p < .01$).

The Repeated-Measures ANOVA model for the GJT data demonstrated a significant main effect of Time, $F(2.2,248.2)=256.45$, $p < .001$, non-significant interaction between Group and Time, $F(6.6,248.2)=1.76$, $p=.1$, and a significant main effect of Group, $F(3,112)=2.6$, $p=.05$. Sphericity was violated so the Greenhouse-Geisser correction was used. Pairwise comparisons demonstrated that the only significant group difference was between Groups 3 [GEb-GEd+] and 4 [GEb-GEd-], during Time 3, MD=.161, $p=.01$. In terms of learning over time, for Group 1

[GEb+GEd+] performance during Time 2, 3 and 4 was significantly better than 1, MD=.398.

SER

Repeated-measures ANOVA for the comprehension data with Time as a within-subject factor and Group as a between-subject factor indicate that there was a significant main effect of Time, $F(2.46,258.4)= 62.15$, $p<.001$, main effect of Group approaching significance, $F(3,105)= 2.51$, $p=.06$, and significant interaction between Time and Group, $F(7.38, 288.4)= 2.03$, $p=.04$. Sphericity was violated so the Huynh-Feldt correction was used. Following the significant interaction between Time and Group, pairwise comparisons using the Bonferroni correction for multiple comparisons revealed that there was no significant difference between the groups at Time 1, Time 3 and Time 4. In addition, during Time 2, Group 1 [GEb+GEd+], 2 [GEb+GEd-], and 3 [GEb-GEd+]performed significantly better than Group 4 [GEb-GEd-] (MD= .121, $p=.009$; MD= .115, $p=.017$, and MD= .102, $p=.05$). Following the significant main effect of Time, pairwise comparisons using the Bonferroni correction demonstrated that for all groups performance during Time 2, 3 and 4 was significantly better than Time 1. In addition, for Group 1 [GEb+GEd+], performance during Time 2 and 4 was significantly better than Time 3 (MD=.071, $p=.02$; MD=.122, $p<.001$).

Results from repeated-measures ANOVA on the production data demonstrated a significant main effect of Time, F (2.79, 288.31)= 127.33, $p<.001$, a significant interaction between Time and Group, F (8.39,288.31)= 2.15, $p=.029$, and a significant main effect of Group, $F(3,103)=5.097$, $p=.002$. Following the significant interaction between Time and Group, pairwise comparisons using the Bonferroni correction demonstrated that during Time 2, Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] were significantly better than Group 4 [GEb-GEd-] (MD= .136, *p*<.001; MD=.115, *p*<.01). During Time 3 and 4 the only difference was between Group 1 [GEb+GEd+] and Group 4 [GEb-GEd-] (MD=.078, *p*=.04, MD= .118, *p*=.034). Pairwise comparisons looking at the retention of knowledge across times, demonstrated that all groups performance during Time 2, 3 and 4 was significantly better than time 1, with no significant differences between Time 2, 3, and 4. In addition for Group 4 [GEb-GEd-], performance at Time 3 was significantly better than at Time 2 (MD= .089, *p*<.001).

Results from repeated-measures ANOVA on the GJT data for SER demonstrated a significant main effect of Time, $F(2.2,251.1)=172.3$, $p<.001$ and a significant main effect of Group, $F(3,112)=4.4$, $p=.006$. Pairwise comparisons revealed that during Time 2, Group 2 [GEb+GEd-] was significantly better than Group 4 [GEb-GEd-], MD=.104, p=.04, during Time 3, Group 1 [GEb+GEd+] and 3 [GEb-GEd+] were significantly better than Group 4 [GEb-GEd-], MD=.09, p=.05; MD=.098, p=.03, and that during Time 4, Groups 1 [GEb+GEd+] and 3 [GEb-GEd+] remained significantly better than Group 4 [GEb-GEd-], MD=.116, p=.03; MD=.14, p=.005. With regards to learning over time, all groups performance during Time 2, 3 and 4 was significantly better than Time 1. In addition, for Groups 1 [GEb+GEd+] and 3 [GEb-GEd+], performance during Time 3 was significantly better than Time 2, MD=.09, p<.01; MD=.09, p<.01. For Group 4 [GEb-GEd-], performance during Time 3 was significantly better than Time 4 indicating loss of knowledge during the delayed posttest, MD=.074, =<.01.

Table A1. Summary table for Repeated-Measures ANOVA models.

	Time	Time*Group	Group
Comprehension OVS	*F*(2.8,288.4)= 337.32, *p*<.001		
Comprehension SER	F(2.46,258.4)=62.15, *p*<.001	*F*(7.38, 288.4)=2.03, *p*=.04	*F*(3,105)= 2.51, *p*=.06
Production OVS	*F*(2.14,227.41)=22.63, *p*<.001		
Production SER	*F* (2.79,288.31)= 127.33, *p*<.001	*F* (8.39,288.31)=2.15, *p*=.029	*F*(3,103)=5.097, *p*=.002
GJT OVS	*F*(2.2,248.2)=256.45, *p*<.001		*F*(3,112)=2.6, *p*=.05
GJT SER	*F*(2.2,251.1)=172.3, *p*<.001		*F*(3,112)=4.4, *p*=.006

Appendix F: Summary of simple linear regression models

Table A2 Simple linear regressions for OVS predicting outcome measures from LLAMA and WM for each group separately

Regression Model	R^2				Regression model *F* value *p<.05, **p<.01,***p<.001				Unstandardized B *t* value *p<.05, **p<.01,***p<.001			
	Gr1	Gr2	Gr3	Gr4	Gr1	Gr2	Gr3	Gr4	Gr1	Gr2	Gr3	Gr4
LLAMA on CovsT3 (111)	.21	.19	.03	.0	7.1*	6.4*	.94	.002	.002 2.6*	.002 2.5*	.001 .97	0 .05
LLAMA on CovsT4 (115)	.13	.54	.02	.47	4.3*	30.37***	.64	23.21***	.003, 2.08*	.006, 5.5***	.001, .8	.006, 4.8***
LLAMA on PovsT2 (115)	.09	.005	.14	.20	2.9^	.128	4.6*	6.3**	.003, 1.7^	.001, .35	.005, 2.15*	.005, 2.5**
LLAMA on PovsT3 (115)	.01	.02	.01	.13	.5	.54	.27	3.87*	.001, .7	.001, .73	.001, .52	.004, 2*
LLAMA on GovsT2	.18	.005	.004	.22	6.1*	.123	.109	7.18*	.004, 2.4*	.001, .351	.001, .329	.004, 2.7*
LLAMA on GovsT3 (115)	.008	.06	.004	.2	.214	1.7	.112	6.4*	.001, .463	.001, 1.3	.0, .33	.004, 2.5*
LLAMA on GovsT4 (115)	.06	.15	.04	.2	1.9	4.5*	1.05	6.5*	.002, 1.3	.002, 2.1*	.001, 1.02	.003, 2.5*
WM on CovsT2	.01	.16	.02	.01	.31	4.5*	.73	.35	.09, .5	.5, 2.15*	-.19, -.85	-.09, -.55

WM on CovsT3	.23	.06	.02	.0	8.2**	1.7	.42	.006	.29, 2.9**	.36, 1.3	.11, .65	-.014, -.078
WM on CovsT4	.03	.25	.06	.0	.83	8.7**	1.7	.0	.18, .91	1, 2.9**	.28, 1.3	-.006, -.02
WM on PovsT2	.01	.15	.001	.006	.004	4.7*	.042	.16	.016, .06	1, 2.18*	-.07, -.2	-.13, -.4
WM on PovsT3	.02	.15	.04	.006	.68	4.4*	1.1	.15	.19, .82	.96, 2.1*	.36, 1.05	.13, .39
WM on GovT2	.08	.12	.02	.02	2.5	3.6^	.58	.7	.32, 1.5	.69, 1.8^	.19, .76	.22, .83
WM on GovT4	.13	.16	.02	.01	4.2*	4.9*	.69	.32	.33, 2.06*	.61, 2.2*	.16, .83	.14, .57

Table A3 Simple linear regressions for SER predicting outcome measures from LLAMA and WM for each group separately

Regression Model	R^2				Regression model *F* value *p<.05, **p<.01,***p<.001				Unstandardized B *t* value *p<.05, **p<.01,***p<.001			
	Gr1	Gr2	Gr3	Gr4	Gr1	Gr2	Gr3	Gr4	Gr1	Gr2	Gr3	Gr4
LLAMA on CserT2 (113)	.04	.08	.04	.24	1.16	2.2	1.1	8.2**	.001, 1.07	.001, 1.5	.001, 1.05	.002, 2.9**
LLAMA on CserT2 (113)	.04	.08	.04	.24	1.16	2.2	1.1	8.2**	.001, 1.07	.001, 1.5	.001, 1.05	.002, 2.9**
LLAMA on CserT4 (111)	.04	.12	.05	.11	1.17	3.5^	1.5	3.2^	.001, 1.08	.002, 1.8^	.002, 1.2	.002, 1.8

LLAMA on PserT2 (114)	.14	.04	.05	.32	4.42*	1.2	1.6	12.5**	.002, 2.1*	.001, 1.1	.001, 1.2	.003, 3.5**
LLAMA on PserT3 (114)	.04	.09	.002	.16	1.3	2.6	.063	5.04*	.001, 1.1	.001, 1.6	.0, .25	.002, 2.25*
LLAMA on PserT4 (111)	.03	.25	.05	.01	1.04	8.13**	1.5	.43	.001, 1.02	.003, 2.85**	.001, 1.2	.001, .65
LLAMA on GserT2	.02	.1	.06	.15	.66	2.9^	2.02	4.6*	.001, .8	.001, 1.7^	.002, 1.4	.002, 2.15
WM on CserT2	.1	.34	.19	.01	3.5^	13.48***	6.2*	.38	.2, 1.8^	.77, 3.6***	.34, 2.5*	.08, .62
WM on CserT3	.14	.30	.27	.06	4.7*	11.14**	10.6**	1.8	.2, 2.2*	.66, 3.3**	.5, 3.26**	.19, 1.3
WM on CserT4	.44	.29	.01	.12	22.21***	10.03**	.49	3.6^	.44, 4.7***	.76, 3.16**	.16, .7	.28, 1.8^
WM on PserT2	.37	.31	.14	.005	16.72***	11.37**	4.3*	.13	.36, 4.09***	.7, 3.4**	.28, 2.06*	-.05, -.37
WM on PserT3	.009	.38	.16	.02	.25	15.24***	5.4*	.48	.04, .5	.68, 3.9***	.26, 2.3	.08, .69
WM on PserT4	.40	.16	.002	.05	18.34***	4.6*	.047	1.5	.34, 4.3***	.62, 2.2*	-.04, -.21	.21, 1.2
WM on GserT2	.12	.18	.007	.002	3.8^	5.8	.209	.06	.2, 1.9^	.45, 2.4	-.08, -.45	-.03, -.24

Bibliography

Abraham, R. (1985). Field independence-dependence and the teaching of grammar. *TESOL Quarterly*, *19*, 689-702.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of experimental psychology: Learning, memory, and cognition*, *16*(5), 883.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of experimental psychology: General*, *117*(3), 288.

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. *Attention and awareness in foreign language learning*, 259-302.

Anderson, J. R. (1993a). *Rules of the mind*: Routledge.

Anderson, J. R. (1993b). Problem solving and learning. *American Psychologist, 48*(1).

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, *29*, 313–342.

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of experimental psychology: learning, memory, and cognition*, *20*(6), 1322.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 932.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*(4), 341-380.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological review, *111*(4), 1036.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*(4), 341-380.

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal, 22*, 84–94.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136-R140.

Bates, E. (1994). Modularity, domain specificity and the development of language. *Discussions in Neuroscience*, *10*, 136-149.

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning—output instruction on the acquisition of the English past simple tense. *Language Teaching Research, 9*(1), 67-93.

Berry, D.C. (1994). Implicit and explicit learning of complex tasks. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 147-165). London: Academic Press.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes, (Ed.) *Cognition and the development of language* (pp. 279-362). New York: Wiley.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics, 27*(02), 185-209.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal, 79*(2), 179-193.

Carroll, J.B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (Ed.). *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, MA: Newbury House.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (1999). Cross-task prediction of working memory performance: Working memory capacity as source activation. In *Sixth Annual ACT-R Workshop*, Fairfax, VA.

De Graaff, R. (1997). The eXperanto experiment. *Studies in Second Language Acquisition, 19*(02), 249-276.

De Graaff, R., & Housen, R. (2009). Investigating the effects and effectiveness of L2 instruction. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp.726-755). Wiley-Blackwell.

Decoo, W. (1996). The induction-deduction opposition: Ambiguities and complexity of the didactic reality. *IRAL*, *34*(2), 95-118.

De Jong, N. (2005). Can second language grammar be learned through listening?: An experimental study. *Studies in Second Language Acquisition*, *27*(02), 205-234.

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533-568.

DeKeyser, R. M. (1995). Learning second language grammar rules. *Studies in Second Language Acquisition, 17*(03), 379-410.

DeKeyser, R. M. (1997). Beyond explicit rule learning. *Studies in Second Language Acquisition, 19*(02), 195-221.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practic- ing second language grammar. In C. D. J. Williams (Ed.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge, England: Cambridge University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*(04), 499-533.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 313-348). Oxford, UK: Blackwell.

DeKeyser, R. M. (2007a). Skill acquisition theory. *Theories in second language acquisition*, 97-113.

DeKeyser, R. M. (2007b). (Ed.). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

DeKeyser, R. M. (2012). Interactions Between Individual Differences, Treatments, and Structures in SLA. *Language Learning, 62*(s2), 189-200.

DeKeyser, R. M. (2015). Skill Acquisition Theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 94-113). New York: Routeledge.

DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning, 46*(4), 613-642.

DeKeyser, R.M., & Koeth, J. (2011). Cognitive Aptitudes for Second Language Learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 2, pp. 395-406). New York: Routledge.

DeKeyser, R. M., & Criado-Sánchez, R. (2012a). Automatization, skill acquisition, and practice in second language acquisition. . In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 323-331). Oxford, UK: Wiley-Blackwell.

DeKeyser, R. M., & Criado-Sánchez, R. (2012b). Practice in second language instruction. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 4501-4504). Oxford, UK: Wiley-Blackwell.

DeKeyser, R. M., & Botana, G. P. (2015). The effectiveness of processing instruction in L2 grammar acquisition: A narrative review. *Applied Linguistics, 36*(3), 290-305.

Domangue, T. J., Mathews, R. C., Sun, R., Roussel, L. G., & Guidry, C. E. (2004). Effects of model-based and memory-based processing on speed and accuracy of grammar string generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(5), 1002.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah: Lawrence Erlbaum.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell.

Doughty, C. J. (2003) Instructed SLA: Constraints, compensation, and enhancement. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 206-257). New York: Blackwell.

Doughty, C., & Williams, J. (1998). Pedagogical choices in focus on form. *Focus on form in classroom second language acquisition*, 197-261.

Doughty, C.J., Campbell, S.G., Mislevy, M.A., Bunting, M.F., Bowles, A.R., & Koeth, J.T. (2010). *Predicting near-native ability: The factor structure and reliability of Hi-LAB.* Paper presented at the Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices.

El-Banna & Ibrahim, A. (1985). *Deductive versus inductive teaching of grammar: An experimental investigation.*

Ellis, N. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European journal of cognitive psychology, 5*(3), 289-318.

Ellis, N.C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143-188.

Ellis, N.C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(2), 305-352.

Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*(1), 83-107.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*(2), 285-301.

Erlam, R. (2003). The effects of deductive and inductive instruction on the acquisition of direct object pronouns in French as a second language. *The Modern Language Journal*, *87*(2), 242-260.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147-171.

Farley, A. (2004a). The relative effects of processing instruction and meaning-based output instruction. In B. VanPatten (Eds.), *Processing instruction: Theory, research, and commentary* (pp. 143-168). Mahwah, NJ: Lawrence Erlbaum Associates.

Farley, A. (2004b). Processing instruction and the Spanish subjunctive: Is explicit information needed? In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 227–239), Mahwah, NJ: Erlbaum.

Fernández, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition, 30*, 277-305.

Fitts, P. M. (1964). Perceptual-motor skill learning. *Categories of human learning*, 47, 381-391.

French, L. M. & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*, 463–87.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, 34, 445 – 474.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages*. Amsterdam: John Benjamins.

Granena, G. (2013). Reexamining the robustness of aptitude in second language acquisition. In G. Granena and M. H. Long (Eds.). *Sensitive periods, language aptitude, and ultimate L2 attainment*. Amsterdam: Benjamins.

Granena, G. (2013b). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, 63(4), 665-703.

Haight, C. E., Herron, C., & Cole, S. P. (2007). The effects of deductive and guided inductive instructional approaches on the learning of grammar in the elementary foreign language college classroom. *Foreign Language Annals*, *40*(2), 288-310.

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited. *Studies in Second Language Acquisition*, *32*(03), 465-491.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly, 40*(1), 109-131.

Hendriks, P. (2013). *Asymmetries between Language Production and Comprehension*. Springer. Retrieved from http://link.springer.com/content/pdf/10.1007/978-94-007-6901-4.pdf.

Henry, N., Culman, H., & Vanpatten, B. (2009). More on the effects of explicit information in instructed SLA . A partial replication and a response to Fernández (2008). *Studies in Second Language Acquisition*, 31, 559– 575.

Herron, C., & Tomasello, M. (1992). Acquiring grammatical structures by guided induction. *The French Review*, 65(5), 708-718.

Huitema, B. (2011). *The Analysis of Covariance and Alternatives*. New York, NY: John Wiley & Sons.

Hulstijn, J. H., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge?: a research proposal. *Aila Review, 11*, 97-112.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society*, 102, 199–226.

Juffs, A. & Harrington, M.W. (2011). Aspects of working memory in L2 Learning. *Language Teaching: Reviews and Studies*, 42 (2), 137-166.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological review*, *99*, 122-149.

Kachinske, I., DeKeyser, Robert M., & Lidz, J. (2013). *Reexamining the power of implicit learning*. Qualifying Paper. University of Maryland.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, 116, 321–340.

Kim, J.-e. (2013). *The optimal conditions for form-focused instruction: Method, target complexity and types of knowledge*. Dissertation. Georgetown University.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*, 14(1), 22-37.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11(2), 261-271.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Oxford University Press.

Lado, B., Bowden, H. W., Stafford, C. A., & Sanz, C. (2014). A fine-grained analysis of the effects of negative evidence with and without metalinguistic information in language development. *Language Teaching Research,* 18(3), 320-344.

Lane, S. M., Mathews, R. C., Sallas, B., Prattini, R., & Sun, R. (2008). Facilitative interactions of model-and experiencebased processes: Implications for type and flexibility of representation. *Memory & cognition*, *36*(1), 157-169.

Larsen-Freeman, Diane. (2009). Teaching and Testing Grammar. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp.518-543). Wiley-Blackwell.

Leow, R. P. (1998a). The effects of amount and type of exposure on adult learners' L2 development in SLA. *The Modern Language Journal*, 82, 49-68.

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal*, 97(3), 634-654.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., . . . Doughty, C. J. (2013). Hi-LAB: A New Measure of Aptitude for High-Level Language Proficiency. *Language Learning, 63*(3), 530-566.

Lightbown, P. M. (1998). The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114-138). Cambridge: Cambridge University Press.

Lightbown, P. M. (2000). Anniversary article. Classroom SLA research and second language teaching. *Applied linguistics*, *21*(4), 431-462.

Lightbown, P., & Spada, N. (1994). An innovative program for primary ESL students in Quebec. *TESOL Quarterly*, 28(3), 563-579.

Logan, G. D. (1988). Toward an instance theory of automatication. *Psychological Review*, 95, 492-528.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign language research in cross-cultural perspective, 2*, 39-52.

Long, M. H. (1998). Focus on form Theory, research, and practice Michael H. Long Peter Robinson. *Focus on form in classroom second language acquisition, 15*.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). New York: Cambridge University Press.

Lovett, M. C., Reder, L. M., & Lebiere, C. (1997). Modeling individual differences in a digit working memory task. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 460-465).

Lyster, R., & Sato, M. (2013). Skill acquisition theory and the role of practice in L2 development. In M. García Mayo, J. Gutierrez-Mangado & M. Martínez Adrián (Eds.), *Multiple perspectives on second language acquisition*. Amsterdam: Benjamins.

Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15(3), 263-287.

Lyster, R. (2004). Research on form-focused instruction in immersion classrooms: Implications for theory and practice. *Journal of French Language Studies*, 14, 321–341.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: Benjamins.

Mackey, A., Adams, R., Stafford, C., & Winke, P. (2010). Exploring the relationship between modified output and working memory capacity. *Language Learning*, 60(3), 501-533.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. *Cognition and second language instruction*, 69-90.

Martin, K. I., & Ellis, N. C. (2012). The Roles of Phonological STM and Working Memory in L2 Grammar and Vocabulary Learning. *Studies in Second Language Acquisition*, 34 (3), 379-413.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchard-Fields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6).

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11*(2), 113-128.

Meara, P. (2005). *Llama language aptitude tests: The manual*. Swansea: Lognostics.

Michas, I. C., & Berry, D. C. (1994). Implicit and explicit processes in a second-language learning task. *European journal of cognitive psychology*, *6*(4), 357-381.

Mitchell, R. (2000). Anniversary article. Applied linguistics and evidence-based classroom practice: the case of foreign language grammar pedagogy. *Applied Linguistics*, 21(3), 281-303.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, 60(1), 154-193.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, 17(01), 56-72.

Nagata, N. (1997). An experimental comparison of deductive and inductive feedback generated by a simple parser. *System*, 25(4), 515-534.

Neves, D. M. & Anderson, J. R. (198l). Knowledge Compilation: Mechanisms for the Automatization of Cognitive Skills. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition*, (pp. 57-84). Hillsdale, NJ: Erlbaum.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528.

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, *51*, 157–213.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill, Inc.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B., (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics*, *27*, 377–402.

Oxford, R. L. (1990). *Language-learning strategies: What every teacher should know*. New York: Newbury House/Harper & Row.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461-472.

Presson, N., MacWhinney, B., & Tokowicz, N. (2014). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*, 35(04), 709-737.

Prieto Botana, G. (2013). T*he role of task-essentialness and explicit information in processing instruction.* Unpublished PhD Dissertation. University of Maryland, College Park.

Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language learning*, *47*(2), 289-325.

Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, *12*(1), 61-82.

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

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118*(3), 219.

Reber, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford: Clarendon.

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5).

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 585.

Rehling, J., Demiral, B., Lebiere, C., Lovett, M., & Reder, L. M. (2003). Modeling individual difference factors in a complex task environment. In F. Detje, D. Doerner, & H. Schaub (Eds.), In *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 287–288). Universitats-Verlag Bamberg.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning 62* (1), 93-132.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 18*(01), 27-67.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*(02), 223-247.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73.

Rodgers, D. M. (2011). The automatization of verbal morphology in instructed second language acquisition. *IRAL-International Review of Applied Linguistics in Language Teaching, 49*(4), 295-319.

Rosa, R., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition,* 21, 511-556.

Rose, K. R., & Ng, C. K. (2001). Inductive and deductive approaches to teaching compliments and compliment responses. In Rose, K. R. and Kasper, G. (Eds.), *Pragmatics in language teaching*, (pp. 145-170). New York: Cambridge University Press.

Sallas, B., Mathews, R. C., Lane, S. M., & Sun, R. (2007). Developing rich and quickly accessed knowledge of an artificial grammar. *Memory & cognition, 35*(8), 2118-2133.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*, 35–78.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*(1), 206-226.

Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. *Consciousness in second language learning, 11*, 237-326.

Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N.Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-209).

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32): Cambridge University Press.

Seliger, H. W. (1975). Inductive method and deductive method in language teaching: A re-examination. *IRAL-International Review of Applied Linguistics in Language Teaching*, *13*(1-4), 1-18.

Shaffer, C. (1989). A comparison of inductive and deductive approaches to teaching foreign languages. *Modern Language Journal*, *73*(4), 395-403.

Seger, C. A. (1994). Implicit learning. *Psychological bulletin, 115*(2).

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly, 41*(2), 255-283.

Shintani, N. (2011). Task-based language teaching versus traditional production-based instruction: Do they result in different classroom processes? *TESOL*, 6, 97-120.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-Based Versus Production-Based Grammar Instruction: A Meta-Analysis of Comparative Studies. *Language Learning*, *63,* 296-329.

Skehan, P. (1990). The relationship between native and foreign language learning ability: Educational and linguistic factors. In H. Dechert (Ed.), *Current trends in European second language acquisition research* (pp.83–106). Clevedon: Multilingual Matters.

Skehan, P. (1998). *A cognitive approach to learning language*. Oxford: Oxford University Press.

Skehan, P. (2012). Language aptitude. In S. M. Gass & A. Mackey (Eds.). *The Routledge handbook of second language acquisition* (pp. 381-395). New York: Routledge.

Snow, R. E. (1991). Aptitude-treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of consulting and clinical psychology*, *59*(2), 205.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching* 30, 73-87.

Spada, N. (2007). Communicative language teaching. In J. Cummins & C. Davis (Eds), *International handbook of English language teaching* (pp. 271-288). Norwell, MA: Springer.

Spada, N. & Lightbown, P. M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15*, 205-224.

Spada, N., & Tomita, Y. (2010). Interactions Between Type of Instruction and Type of Language Feature: A Meta-Analysis. *Language Learning*, *60*(2), 263-308. doi: 10.1111/j.1467-9922.2010.00562.x

Stafford, C. A., Bowden, H. W., & Sanz, C. (2012). Optimizing language instruction: Matters of explicitness, practice, and cue learning. *Language Learning, 62*(3), 741-768.

Stanley, W. B., Mathews, R. C., Buss, R. R., & Kotler-Cope, S. (1989). Insight without awareness: On the interaction of verbalization, instruction and practice in a simulated process control task. *The Quarterly Journal of Experimental Psychology*, *41*(3), 553-577.

Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. Sun (Ed.). *Cognition and multi-agent interaction* (pp. 79-99). New York: Cambridge University Press.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*(2), 203-244.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychological review*, *112*(1).

Sun, R., Zhang, X., Slusarz, P., & Mathews, R. (2007). The interaction of implicit learning, explicit hypothesis testing learning and implicit-to-explicit knowledge extraction. *Neural networks, 20*(1), 34-47.

Taatgen, N. A. (2001). A model of individual differences in learning air traffic control. In *Proceedings of the fourth international conference on cognitive modeling* (pp. 211-216). Mahwah, NJ: Lawrence Erlbaum Associates.

Taatgen, N. A., & Wallach, D. (2002). Whether skill acquisition is rule or instance based is determined by the structure of the task. Cognitive Science Quarterly, 2(2), 163-204.

Takimoto, M. (2005). *Effects of deductive and inductive instruction on Japanese learners' pragmatic competence*. Unpublished doctoral dissertation, Temple University Graduate Board. Japan: Tokyo.

Tanenhaus, M. K., Carroll, J. M., & Bever, T. G. (1976). Sentence-picture verification models as theories of sentence comprehension: A critique of Carpenter and Just. *Psychological Review, 83*, 310-317.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, *33*(1), 67-81.

Vafaee, P. & DeKeyser, R., (2014). *Re-examining the Role of Explicit Information in Processing Instruction: Bridging a Methodological Gap*. (Unpublished Qualifying Paper), University of Maryland, College Park.

VanPatten, B. (1996). *Input processing and grammar instruction*. Norwood, NJ: Ablex.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, *52*(4), 755–803.

VanPatten, B. (2004). (Ed.). *Processing instruction: Theory, research, and commentary*. Mahwah, NJ: Erlbaum.

VanPatten, B., Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*, 225-225.

VanPatten, B., & Oikkenon, S. (1996). Explanation vs. structured input in processing instruction. *Studies in Second Language Acquisition, 18*, 495–510.

VanLehn, K. (1996). Cognitive skill acquisition. In J. Spence, J. Darly, & D. J. Foss (Eds.), *Annual review of psychology* (Vol. 47, pp. 513–539). Palo Alto, CA: Annual Reviews.

Vatz, K., Tare, M., Jackson, S. & Doughty, C. (to appear, 2013a).. In Granena, G. & Long, M.H. (Eds.) *Sensitive periods, language aptitude, and ultimate L2 attainment. Aptitude-treatment interaction studies in second language acquisition: Findings and methodology* (pp. 273-92). Amsterdam: John Benjamin's Publishing.

Wang, L. Y. (2002). *Effects of inductive and deductive approach on EFL learning collocation patterns by using Concordancers.* Unpublished master's thesis, National Yunlin University of Science and Technology, Taiwan.

Watari, Y. (2014, August). *What does "explicit" mean? A methodological consideration in explicit grammar teaching research.* Paper presented at AILA World Congress 2014, Brisbane, Australia.

Webb, S. (2005). Receptive and productive vocabulary learning. *Studies in Second Language Acquisition, 27*, 33-52.

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. *Individual differences and universals in language learning aptitude*, 119-154.

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12*, 416-432.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*(01), 1-48.

Williams, J. N. (2009). Implicit learning in second language acquisition. In W.C. Ritchie & T. K. Bhatia (Eds). *The new handbook of second language acquisition* (pp. 319-353). Bingley: Emerald Press.

Williams, J. N., & Lovatt, P. J. (2003). Phonological Memory and Rule Learning. *Language Learning*, *53*(1), 67-121.

Willingham, D. B., & Goedert-Eschmann, K. (1989). The Relation Between Implicit and Explicit Learning: Evidence for Parallel Development. *Psychological Science, 10*(6), 531-534. doi: 10.1111/1467-9280.00201

Wong, W. (2004a). The nature of processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 33-63). Mahwah, NJ: Lawrence Erlbaum Associates.

Wong, W. (2004b). Processing instruction in French: The roles of explicit information and structured input. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary (pp. 187-205). Mahwal, NJ: Erlbaum.*

Xia, Z. H. (2005). The effectiveness of different grammar instructions. *CELEA Journal Bimonthly*, *28*(1), 23-28.

Yang, L. R. & Givon, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keck second language learning project. *Studies in Second Language Acquisition*, *19*, 173-193.