ABSTRACT

Title of Document:	THE ROLE OF TASK-ESSENTIALNESS AND EXPLICIT INFORMATION IN PROCESSING INSTRUCTION
	Goretti Prieto Botana, Doctor of Philosophy, 2013
Directed By:	Dr. Robert DeKeyser, Professor, Second Language Acquisition

Numerous studies in the SLA literature document the advantages of Processing Instruction over traditional instruction methods (e.g., VanPatten & Oikkenon, 1996; Benati, 2001; Wong, 2004; Fernández, 2008). A common finding of many of those studies is that as long as learners are exposed to structured input, explicit information offers no additional benefit in the learning process. Structured input is defined as practice that requires learners to attend to the grammatical item (Lee & VanPatten 1995), pushing them to process the targeted form (VanPatten 2000) so that a connection between form and meaning is made (VanPatten 2004). Effectively, what such conditions spell out is that structured input practice should be task-essential, i.e., that "the task cannot be successfully performed unless the structure is used" (Loschky & Bley-Vroman 1993). Indeed, virtually all research studies implementing structured input use task-essential activities in both treatment and exit tests. In light of these facts, questions arise as to a) the extent to which the benefits attributed to Processing Instruction originate in task-essentialness and b) whether explicit information might only be superfluous if practice fulfills that condition. The present study sought to address these issues through an empirical study with a pre-, post- and delayed posttest design. The variables +/- explicit information and +/- task-essential were combined to form four experimental conditions. A control group was added to obtain a baseline. One hundred and thirty learners of Spanish randomly assigned to each group completed a treatment on the ser/estar copula distinction and object-verb-subject structures. Findings suggest a) that structured input needs to be task-essential in order for its benefits to obtain, b) that in the absence of task-essential practice explicit information becomes necessary and c) that benefits obtained from explicit information are both more consistent and more durable than those obtained from task-essential practice.

THE ROLE OF TASK-ESSENTIALNESS AND EXPLICIT INFORMATION IN PROCESSING INSTRUCTION

.

By

Goretti Prieto Botana

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
2013

Advisory Committee:
Dr. Robert DeKeyser, Chair
Dr. Steven Ross
Dr. Catherine Doughty
Dr. Kira Gor
Dr. Patricia Alexander

Acknowledgements

I would like to extend my most sincere gratitude to my advisor, Dr. Robert DeKeyser, for his unwavering support and dependability throughout this process. His guidance and mentorship were invaluable for the completion, not only of this dissertation but also my doctoral studies. Furthermore, I would like to thank the members of my committee: Dr. Catherine Doughty, Dr. Kira Gor, and Dr. Patricia Alexander, for their insightful comments, and, especially Dr. Steven Ross, for patiently teaching me all the useful statistics I know.

My thanks also go to the Dr. Manel Lacorte and the Department of Spanish and Portuguese at the University of Maryland, whose generosity and flexibility made the data collection for this study a swift procedure. Thank you, also, to María Gómez Martín, Luis Charry, Sebastián Bartis, Lina Morales, Lisa Warren and Evelyn Canabal Torres for graciously allowing me to invade their classrooms.

In addition, I would like to thank my peers, in particular, David Libber, Suzanne Freynik, Àngels Llanes Baró, Jason Bartles, Carolina Gómez-Montoya, Amy Karp, Sofia Calzada, Rocio Gordon, Norman González, Elena Becerril and Chris Lewis. The laughs, levity and support I got from you restored my sanity more than a few times.

Finally, my deepest gratitude to my family. To my husband, Chris, for his understanding, his optimism, his constant encouragement and, above all, his contagious strive to always rise up to a challenge; and to my parents and sister, whose faith in me knows no end.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Since the early 1990's a considerable body of literature has been generated on the issue of input processing (IP) and Processing Instruction (PI). IP is concerned with how subsets of the input learners are exposed to are processed in working memory. PI, in turn, is a pedagogical technique for grammar instruction which is based on the insights of IP research and whose aim is to identify and alter processing strategies that result in inadequate processing of input. In its original form, this approach comprises three components: the first is explicit information (EI), which provides learners with the grammatical properties of a linguistic form as well as the rules that govern it. Driven by its IP foundations, the second component articulates the inadequate processing tendencies that learners often instinctively apply upon interpreting a given form, and it provides coaching on how to adequately parse it. The third and final component is structured input (SI), which consists of practice devised to promote the mapping of a form to its correct meaning, thus leading the learner to repeated instances of proper processing.

Starting with a number of studies by Bill VanPatten, numerous research endeavors in recent years have attempted to determine whether the three components of PI do indeed successfully adjust learner default processing strategies more effectively than other frequently implemented types of instruction. Indeed, a number of studies in the literature document the advantages of processing instruction (PI) over traditional instruction (TI) methods, as shown by experiments with various target languages and a number of different structures (e.g., VanPatten & Cadierno, 1993;

VanPatten & Oikkenon, 1996; Benati, 2001; Wong, 2004). One of the earliest examples and perhaps the most groundbreaking was VanPatten and Cadierno (1993), a study on the Spanish object pronouns clitics that featured three groups: one receiving PI, one TI and one control group receiving no treatment. Results from this seminal study revealed that the PI group performed significantly better than TI and control groups in an interpretation posttest, which suggested that PI is a method superior to the ones habitually employed in classrooms. These results, however, were criticized and questioned due to differences in the explicit information that the treatment groups received; namely, while the group receiving PI was alerted to the invalid processing strategies, the TI group was only provided with metalinguistic explanations thus opening up the possibility that the advantage recorded for the PI group may be due to the nature of the explicit information received, rather than the type of practice. Largely triggered by this criticism, a follow-up study was conducted (VanPatten and Oikkenon, 1996), which attempted to tease apart the impact of EI and whether benefits of PI would obtain in its absence. With a group receiving EI only, and two other treatment groups which were exposed to PI (EI and SI) and SI only, VanPatten and Oikkenon (1996) found the latter two groups did not perform significantly differently from one another, but both showed significant improvement with respect to the EI only group. This suggested that EI was not necessary for the benefits of PI to obtain; or in other words, it suggested that of the different components of PI, SI sufficed to match the benefits of TI, with the other two (EI and processing strategies) being expendable[1]. Sparked by these findings, several research

[1] According to the definition in DeKeyser (1995), explicit instructional treatments include those where rules are provided (in which case the instruction is explicit deductive) as well as cases where learners

endeavors in the last decade have further studied the role of EI in L2 acquisition and found no effect (e.g., Benati, 2004; Sanz & Morgan-Short, 2004).

Across this literature SI is primarily defined as practice that has been purposefully prepared to highlight a certain grammatical feature or put it in a privileged position (VanPatten 1993; Lee & VanPatten1995). Crucially, SI activities require "that the learner attend to the grammatical item" (Lee & VanPatten 1995), pushing them to process the targeted form (VanPatten 2000). As stated in VanPatten (2004), in this context processing the form should be understood to mean that a form-meaning connection has been made. In addition, a second defining quality of SI activities is that they are strictly receptive, and do not require any language production on the part of the learner.

Interestingly, and although not spelled out in so many words in most of the PI literature, what such conditions appear to imply is that processing the targeted form should be crucial for the completion of the task. In other words, based on the definitions and operationalizations found in the literature, SI activities should be task-essential (TE) such that "the task cannot be successfully performed unless the structure is used" (Loschky & Bley-Vroman, 1993, p. 132). Although Sanz (2004) and Sanz & Morgan-Short (2004) explicitly mention that SI needs to be TE, the great majority of authors, including VanPatten himself, simply make mention of the fact

---

are asked to focus on a form and try to discern the rules themselves (explicit inductive learning). In this view, SI could very well constitute in and off itself an explicit treatment and it would be unwarranted to argue that results from VanPatten and Oikkenon (1996) suggest EI is not necessary. In this paper all mentions of EI refer to explicit deductive metalinguistic information. No claim is being made in terms of the means (implicit or explicit learning) in which learners obtained their gains. The focus of this paper revolves around what it is that may make PI effective and whether metalinguistic information may, indeed, become necessary under certain circumstances.

that this type of practice should result in learners making a form-meaning connection. Perhaps unsurprisingly, nearly all research studies implementing SI have used TE activities in both treatments and posttests. In fact, one could say that, in virtually all instances, the activities these authors label as SI are nothing but TE activities in that no other manipulation is present beyond the ones required for the form to become a crucial element. In addition to this, findings from recent research suggest that PI practice in the form of affective activities, in which the targeted form is exhibited, but not essential for task completion, do not appear to result in learning, regardless of whether alone or combined with form-focused activities (Marsden and Chen, 2011). Considering these findings in the literature, two questions arise: The first pertains to benefits that can be derived from SI practice. To what extent do the benefits attributed to SI originate in task-essentialness? The second question pertains to the role of EI. Especially in view of findings suggesting that EI is unnecessary if SI is present, is it possible that such results only obtain when SI is indeed TE? In the following section we will review the key concepts in PI in more detail before proceeding to the overview of the relevant literature.

1.2. Processing Instruction

Motivated by the prevalence of production-oriented practice in second and foreign language classrooms, VanPatten (1993) proposed a new approach to grammar teaching termed Processing Instruction (PI). This approach was explicit in nature and focused on input processing, advocating for displacing output practice from the central position it enjoyed at the time. In several articles, VanPatten put forth a model of acquisition consisting of three sets of processes (see Figure 1):

	I		II		III	
Input	→	Intake	→	Developing System	→	Output

Figure 1: Processes of Acquisition (VanPatten & Cadierno 1993, p. 226)

The first set of processes (I) involves using elements of the input for intake, where the latter is understood as the subset of input that is processed in working memory and becomes eligible for incorporation into the developing system (VanPatten, 1995, 2004). The second set of processes (II) entails the operations responsible for accommodation of new data and restructuring of the developing system. Provided that linguistic data are successfully incorporated into the long-term store, in this phase, intake will be integrated into previously existing data, possibly causing changes to the interlanguage (IL) structure (VanPatten 1993; 1995). Finally, the third set of processes (III) revolves around the mechanisms that provide access to the linguistic data that have been incorporated into the system by way of set II processes.

According to VanPatten, “the acquisition of formal features of language begins with input and input processing” (VanPatten 1996, p. 55). Following this premise, he claimed that the output-driven instruction commonly implemented in the US was neither “informed by general second language acquisition theory” (VanPatten 1996, p. 55) nor psycholinguistically motivated. More precisely, VanPatten noted that, while SLA literature highlighted the importance of input for acquisition to take place, this was obviated in classrooms, where practice was solely concerned with learners producing the targeted form.

"Note the contradiction between traditional grammar practice and our model of acquisition. The development of a system is input dependent; it happens when learners receive and process meaning-bearing input. Traditional grammar practice, on the other hand, is exclusively output oriented. That is learners get an explanation and then are led through output-practice (…) [the drills] require learners to produce the language. Under this traditional scenario, how is the developing system provided with relevant input data that is both comprehensible and meaningful? Because it focuses on output, traditional grammar instruction engages those processes involved in accessing a developing system rather than those involved in forming the system." (Lee and VanPatten 1995, p. 94)

Unless practice draws on "exposure to language samples in the input" (VanPatten 1996, p. 59), VanPatten argued, it can never lead to incorporation and restructuring of the learner's developing system (VanPatten 1996). Thus, his contention was that, in order for IL to accommodate new forms, grammar instruction ought to focus first on those processes related to the transformation of input into intake, i.e., processes that precede the output phase. It would be during that first set of processes that the connection between grammatical form and its meaning is established, a crucial step that, as VanPatten proposed, is skipped when input is immediately followed by output practice alone. In his new approach, instead of the traditional mechanical to communicative practice sequence, VanPatten advocated for more communicatively driven and input-rich classrooms (VanPatten 1996, p. 59), where proper processing of target forms would be paramount, heeding the importance of input that is documented in the SLA literature. As might be expected, from this perspective, the way in which learners process a given string of linguistic data is the necessary starting point. In this vein, PI offers a number of claims in the form of principles that lay out what mediates input processing and which serve as the basis for devising more effective processing strategies. As of the latest revision of PI, those principles are excerpted below:

"Primacy of Content Words Principle: Learners process content words in the input before anything else.
(...)
Lexical Preference Principle: If grammatical forms express a meaning that can also be encoded lexically (i.e., that grammatical marker is redundant), then learners will not initially process those grammatical forms until they have lexical forms to which they can match them.
(…)
The Preference for Nonredundancy Principle: Learners are more likely to process nonredundant meaningful grammatical markers before they process redundant meaningful markers.
(…)
The Meaning before Nonmeaning Principle: Learners are more likely to process meaningful grammatical markers before nonmeaningful grammatical markers.
(…)
The First Noun Principle: Learners tend to process the first noun or pronoun they encounter in a sentence as the subject.
(…)
The L1 Transfer Principle: Learners begin acquisition with L1 parsing procedures.
(…)
The Event Probability Principle: Learners may rely on event probabilities, where possible, instead of the First Noun Principle to interpret sentences.
(…)
The Lexical Semantics Principle: Learners may rely on lexical semantics, where possible, instead of the First Noun Principle (or an L1 parsing procedure) to interpret sentences.
(…)
The Contextual Constraint Principle: Learners may rely less on the First Noun Principle (or L1 transfer) if preceding context constrains the possible interpretation of a clause or a sentence.
(…)
The Sentence Location Principle: Learners tend to process items in sentence initial position before those in final position and those in medial position." (VanPatten 2007)

In instances where the default processing strategy results in faulty or no form-meaning mapping, in order for correct IL development to take place, a new and adequate strategy needs to deployed. That, in fact, is the manifest goal of PI, namely, "to alter the processing strategies that learners take to the task of comprehension and to encourage them to make better form-meaning connections than they would if left to their own devices" (VanPatten 1996, p. 60). Based on these principles, PI proposes three different steps to alter learners' processing mechanisms where necessary. The first component of PI is explicit information (EI). Learners are provided the properties and rules pertaining to the target form. Importantly, in EI the relationship between the form and its meaning should be overtly explained. In addition to EI, PI should also incorporate information about processing strategies designed to steer learners away from acting according to a default processing course of action that would result in incorrect form-meaning mappings. Finally, PI features a third component, namely, Structured Input. Essentially Structured Input is practice. Specifically, this consists of activities purposefully manipulated to contain meaning-bearing input and which are deliberately designed to push learners to actively process the target form and connect it to its function. Given that PI seeks to focus on and deepen the phase of acquisition when input is processed (rather than the phase during which the internalized data are retrieved), SI activities do not require learners to produce any linguistic output. According to VanPatten (1996), while EI can be found in traditional instruction, suppliance of strategies and SI are unique to PI[2]. In addition to the principles regarding how learners process input, PI offers six guidelines (Lee and VanPatten 1995; VanPatten 1996) to inform the creation of SI activities:

1) Teach only one thing at a time: This first guideline mandates that one function and one form be the focus at any given time. An example of this would be that new Spanish verb forms should be introduced one grammatical person conjugation at a time.

---

[2] While VanPatten (1996) clearly considers EI and strategy suppliance as two separate items, these appear to be considered as two forms of EI in studies such as VanPatten and Cadierno (1993) and VanPatten and Oikkenon (1996) where upon both of these being eliminated from the treatment for some of the groups the conclusion is that EI is not needed for PI benefits to obtain. In the same way, authors attempting to replicate the two seminal PI studies also consider strategies part of EI.

[3] Based on the report in Cheng (2002), it would seem that the design in the study incorporated experimental treatments that differed with respect to the EI and the type of practice they provided. Given that such a difference proved to be a rather serious weakness in VanPatten and Cadierno (1993), it seems somewhat peculiar this detail would have been incorporated in a subsequent study. However,

2) Keep meaning in focus: For SI activities to be effective, they should require that the learners attend to meaning, avoiding practice which can be completed ignoring the neighboring context.

3) Learners must do something with the input: Resting on the premise that attention is a prerequisite for input to be internalized, this guideline calls for SI activities to demand a reaction from learners as a measure of increasing the likelihood that the input is being attended to.

4) Use both oral and written input: This guideline seeks to accommodate different learners' preferences for input in written or oral form.

5) Move from sentences to connected discourse: In order to facilitate focus on form-meaning connection of the targeted form, SI activities should move from presenting input at the sentence level first and only after that integrate it into connected discourse.

6) Keep psycholinguistic processing strategies in mind: This last guideline mandates that the way in which learners process input be kept in mind. Thus, in a sentence where the first word is not the subject, for instance, it is important to remember that learners are likely to still process the element in sentence-initial position as such.

It is this last guideline that brings task-essentialness into the picture (at the very least when default processing results in misinterpretation). After all, creating an activity in compliance with guideline 6, while pushing for the connection between a form and its meaning to be made, will necessarily result in the targeted item being made essential for the completion of the task. In this vein, then, and to recapitulate, the questions that this research wishes to pose are (a) what is the contribution that task-essentialness makes to the benefits obtained in PI; and (b) is it the case that the superfluous nature of EI previously reported in literature only obtains in the presence of task-essentialness. The importance of these questions is twofold: on the one hand evidence in support of the claim that TE is critical for PI to result in benefits would offer a more theoretically parsimonious explanation for the gains reported in previous research, as well as shed light on what aspects of this approach, as it is currently defined, must absolutely be incorporated in practice for gains to be observed. On the other hand, and perhaps most importantly, in the SLA literature PI studies are among those most directly linked to language instruction. At present, the PI studies overwhelmingly report that if and when SI is present, EI offers no additional benefit for learning. In the absence of further research, the only evidence in the PI literature on the role of EI would appear to suggest that classrooms should instead focus on providing learners with SI instead of investing time in provision of rules and other metalinguistic information.

Without proper qualification as to the circumstances under which SI can suffice, such a message dismissing EI could constitute a mischaracterization of the potential benefits EI has to offer. For this reason it would be of utmost interest to examine whether activities that focus on form-meaning connections must be implemented if EI is to be dispensed with and whether provision of EI under experimental conditions that more closely resemble those in a classroom may reveal a beneficial effect for this aspect that has thus far gone undetected. In the following section we will review prior research endeavors relevant to our present purposes.

Chapter 2: Review of the Literature

2.1. EI in PI Literature

As mentioned previously, two early articles in the PI literature are particularly germane to the discussion of its benefits and the role of EI. The first of them is VanPatten and Cadierno (1993), a study which sought to examine the effects of different types of explicit instruction in the acquisition of grammar. The paper reported on an experiment comparing traditional methods of instruction to PI. After pretesting all three groups in the study, learners in PI and TI treatment groups, received instruction on Spanish object pronouns. The treatment that PI and TI groups received differed in two ways: First, while the PI group received information on inefficient processing strategies, the TI group simply received metalinguistic information about form and placement of the object pronouns. The second difference pertained to the type of practice each of the groups engaged in. In a progression from mechanical to meaningful, the TI group (a) completed oral and written activities where the direct object in a sentence had to be replaced by its corresponding pronoun, (b) answered sentence-level and open-ended questions consisting of disconnected sentences which called for the target structure, (c) and subsequently engaged in more open-ended question-answer interactions and conversation. In the PI group, practice did not involve any linguistic output. Learners completed both picture and sentence matching tasks where correct interpretation of the target structure was crucial. After these, participants were administered affective activities that exposed them to object pronouns. Posttests revealed superior performance for PI over TI and control groups in the interpretation task and no difference between PI and TI in production, although both were significantly better than the no-instruction control group. These results, interpreted by the authors as clear indications of the superiority of PI as a grammar teaching approach, received a fair amount of attention and also a great deal of criticism. Mainly, critics pointed out that, unlike the TI group, the PI group was provided with processing strategies that specifically directed learners to beware of interpreting the first noun in the sentence as the subject. Given that the treatments also differed in the type of practice learners performed, this introduced a confound, since the different EI could have resulted in learners in the PI group monitoring their responses in the interpretation task, which would account for the superiority PI exhibited over TI.

In order to address this concern, a new study was conducted. VanPatten and Oikkenon (1996) carried out a replication of VanPatten and Cadierno (1993) aimed at isolating the effects of EI from the benefits of PI. As in VanPatten and Cadierno (1993), the target structure in Van Patten and Oikkenon (1996) was also the Spanish object pronoun system. In this case, the three groups being compared consisted of a control group receiving EI only and two PI groups. The first PI group, termed the regular PI group, as had been the case with the PI group in VanPatten and Cadierno (1993), received EI, information on processing strategies and SI. The second PI group, termed SI Only, differed from the other PI group in that participants received neither EI nor information on processing strategies. Using the same practice items as in VanPatten and Cadierno (1993), during the treatment participants in the PI groups worked first on sentence interpretation, after which they completed affective activities that exposed them to the target form. Immediately after the fourth and last treatment session, all groups were administered a posttest consisting of an interpretation and a production test. Results revealed no significant differences between the PI groups, although both groups performed significantly better than the EI Only group. VanPatten and Oikkenon (1996) took these results to indicate that the gains the Regular PI group exhibited did not originate in EI. In view of past results from VanPatten and Cadierno (1993), the suggestion was that SI suffices to obtain the benefits recorded for PI. By extension, then, these results also appeared to suggest that SI is superior to TI for teaching grammar.

Fueled by the possibility that a demonstrably more effective way to teach grammar may have been identified, the field saw a proliferation of studies seeking to replicate the findings from VanPatten and Cadierno (1993) and VanPatten and Oikkennon (1996). Among the many replications of VanPatten and Cadierno (1993) is Cheng (2002), which reported on data from a study that intended to examine whether the effectiveness of PI would extend to the Spanish ser/estar copula distinction. Arguing that, unlike OVS, this distinction is a linguistic phenomenon lacking an obvious one-to-one form-meaning mapping and of low communicative value, Cheng (2002) assigned 109 participants to two experimental conditions, PI and TI, and to a control group. The TI group received EI on the usage of ser/estar with adjectives and past participles, followed by activities sequenced to provide first mechanical and then communicative output practice. By contrast, the PI group received EI which deliberately directed them to ser and estar as crucial items to decode the full meaning of the items[3] and SI practice that "pushed the learners to get meaning from the target forms rather than using their existing strategy of assigning ser as the default during the act of comprehension" (Cheng 2002, p. 311). After screening participants by way of a pretest, learners engaged in two days of practice which were followed by an immediate posttest consisting of a sentence interpretation task, sentence production and guided composition. A delayed posttest was administered three weeks later to check the durability of instructional gains. When results were analyzed combining both items calling for ser and items calling for estar, no difference was found between the gains obtained by TI and PI groups. However, upon limiting the analyses to items calling for estar[4] Cheng (2002) found that the PI group showed significant gains over TI in interpretation; both experimental groups exhibited comparable gains with respect to production. The same effects emerged again in the delayed posttest for sentence interpretation and the sentence production tasks but not for the guided composition task[5]. Findings from this study, then, partially mirrored those in VanPatten and Cadierno (1993) to the extent that the PI group appeared superior to TI in interpretation. Unlike in the original study, however, this difference did not emerge in production, a fact attributed by Cheng to the amount of communicative practice included in the TI treatment.

---

[3] Based on the report in Cheng (2002), it would seem that the design in the study incorporated experimental treatments that differed with respect to the EI and the type of practice they provided. Given that such a difference proved to be a rather serious weakness in VanPatten and Cadierno (1993), it seems somewhat peculiar this detail would have been incorporated in a subsequent study. However, Cheng (2002) clearly states that the study reported on was meant to replicate VanPatten and Cadierno (1993), and not VanPatten and Oikkenon (1996), and thus it is possible that this detail was purposefully mirrored in order to make results comparable to those in the original study.

[4] As documented by VanPatten (1987) learners initially generalize SER to all linguistic environments, gradually acquiring contexts that call for ESTAR.

[5] Results indicated that in the guided composition task the control group improved over time, resulting in no significant differences between either of the groups at the time of the delayed posttest. Cheng accounts for this by adducing that the number of tokens for ESTAR in this task was very low.

Various studies, however, did replicate the results of VanPatten and Oikkenon (1996). Benati (2004), for instance, conducted a small scale study aimed at examining whether prior results that "showed no effect for EI in the treatments administered" had been "a fluke" (Benati 2004, p. 207) or whether, on the contrary, they could be expected to generalize. In order to test whether SI activities gave evidence of being sufficient beyond Spanish object (O) verb (V) subject (S) structures, Benati (2004) chose a different linguistic target, namely, the Italian future tense. Using a pretest-posttest design, he randomly assigned participants to PI, SI only and EI only groups. The groups were provided with practice items devoid of time adverbs, thus promoting form-meaning mapping between future tense verb endings and their function. Aside from constituting a change in TL, the variation in target structure also represented a change in the processing preference that learners would rely on (from First Noun Principle to Lexical Preference Principle), thus permitting Benati to examine whether the PI seemed to succeed in altering different processing strategies in the absence of EI. After treatment, exit tasks in the form of written comprehension and production yielded results parallel to those of VanPatten and Cadierno (1993): PI and SI groups outperformed the EI only group, although they exhibited no difference between one another. In addition, the gains recorded were maintained over time by learners in both groups, as shown by the results from the delayed posttest administered a month after treatment. From such findings, Benati concluded that in line with results from VanPatten and Oikkenon (1996), the gains recorded in PI were to be attributed to the SI, EI being neither necessary nor beneficial.

Similarly, and motivated by the possibility that SI may be effective only with highly meaningful aspects of grammar, such as word order, Wong (2004) conducted a study also aimed at observing whether PI would indeed suffice to address faulty processing strategies other than the First Noun Principle. Like Benati (2004), she drew on the Lexical Preference Principle and chose a form of low communicative value, the particle *de* from French negatives, to test the robustness of VanPatten and Oikkenon's (1996) findings. After a pretest, participants were divided into three treatment groups and a control groups. Learners in the PI group read the EI and proceeded to SI activities, whereas the SI and EI groups were trained only through SI activities or metalinguistic information, respectively. Immediately after the one-day treatment, exit tasks in the form of sentence interpretation and a production task were administered to all groups. In line with previous studies, both exit tasks yielded significantly higher results for SI and PI over EI and control groups, once again lending credence to the claim that EI could be dispensed with.

With the same motivation, Farley (2004) set out to test whether VanPatten and Oikkenon's (1996) results could be replicated with a structure of greater complexity[6] such as the Spanish subjunctive. Two groups, SI and PI, received two consecutive days of treatment by way of 10 SI activities of both the referential and affective kind. In the case of the PI groups, EI on the subjunctive was read aloud prior to the treatment and never again repeated. Results from two exit tasks (which were composed of a multiple-choice task in which learners had to pick the correct main

[6] Although the claim that the subjunctive is of greater complexity is rather uncontroversial and is rationalized in Farley (2004 Chapter 7), it must be noted that there is no evidence to date to support this claim empirically.

clause and a sentence completion task) revealed that both PI and SI groups had made significant gains from the treatment. Interestingly, however, Farley's results also revealed an interaction between time and instruction, which appeared to originate from the greater gains obtained by the PI group. This suggested that although SI was sufficient for gains to occur, the benefits obtained from that type of practice were not equal to those generated by the PI treatment. Thus, on the one hand, Farley (2004) yielded results that were in line with previous studies in that SI appeared to be sufficient for benefits to ensue, lending support to prior claims advancing that EI is not necessary. On the other hand, however, findings from this study differed from those of previous research in that suppliance of EI was shown to further facilitate acquisition. In all, then, Farley's (2004) findings suggest that, although EI may not be necessary, it is still beneficial. By way of explanation, Farley proposed that the difference in the role of EI may have been caused by the greater complexity of the subjunctive. He argues that, unlike previous target structures, the subjunctive "is a form that needs to be connected to multiple submeanings" (p. 238), a circumstance under which EI can help create form-meaning mappings more quickly than SI only. Thus, Farley speculates that EI may be beneficial for language structures whose form-meaning connections are not transparent, stating that testing this hypothesis would involve computerized treatment delivery that would allow tracking response accuracy over time to determine how long the mapping had taken (Farley, 2004 p. 238).

The baton left by Farley was readily taken by Fernández (2008), who, in a recent study, set out to shed light on the EI debate by tracking the behavior of learners engaged in practice. She chose two different linguistic targets in Spanish: OVS sentences, and the subjunctive. Each posing a different type of processing problem, these targets offered the opportunity to observe how different structures may be affected by the presence or absence of EI. After reviewing the directions, learners in the PI condition had the chance to read the EI in a series of set-pace frames that appeared on their computer screens. Once they had been exposed to the EI, they proceeded to practice items which required them to match the pictures (in the case of OVS) or main clauses (for the subjunctive) to sentences they heard. Learners in the SI group underwent the same procedure but received no EI presentation. Data were analyzed to determine trials to criterion, defined as the number of trials learners needed before they could process three target items and a distracter correctly and response time. While results revealed no difference between the two treatment groups in OVS processing, in line with findings from Farley (2004), the EI group started to process the subjunctive forms significantly sooner than both the PI groups, hinting again that EI may only be expendable with certain linguistic phenomena.

The results from Fernández (2008) inspired Henry, Culman and VanPatten (2009) to conduct a replication study that set out to examine whether the effects of EI depend on the target structure or relate to a processing problem. This time with German as the TL, two different groups were randomly assigned to +EI and – EI conditions. As Fernández had done, Henry et al. (2009) administered their treatment via computer to control the trials to criterion (reaction time was not measured). The treatment, which encompassed German OVS structures and the marking of the accusative case in the definite article, consisted of a picture matching task.

Analyses showed that the +EI group was faster to start processing the OVS strings correctly, suggesting once again that EI had had a positive effect. The data showed that the +EI group took significantly fewer trials to reach criterion (Henry et al., 2009, p. 570). In other words, these findings are in line with Farley (2004) and Fernández (2008) in the sense that they depict EI as a facilitator of acquisition. However, results reported in Henry et al. (2009) are also at odds with those in Fernández (2008) in that the latter found no EI effect in OVS sentences, whereas the former did. Henry et al. (2009) advanced that the more complex explanations required in the Spanish OVS EI may have been more taxing for learners, thus making an EI effect less likely to surface.

Given that Henry et al. (2009) never intended "to make claims about the relative outcome of PI and structured input-only" (Henry et al. 2009, p. 572), the authors argue that these results cannot speak to the role of EI in PI and propose that such conclusions would demand a different experimental design. Although a third group acting as a control would have indeed strengthened the conclusions, the argument that findings from Henry et al. (2009) are not relevant to the EI debate seems somewhat incongruent. After all, the treatment consisted of "EI for the +EI group and structured input items for both groups" (Henry et. al, 2009 p. 566), and its design and procedure mirror Fernández's (2008), a study that legitimately makes claims about the role of EI in PI. Be that as it may, it remains the case that the facilitative effect found in Henry et al. (2009) prompts questions as to whether PI without EI may only be viable with a limited set of structures where figuring out EI inductively may be possible. As DeKeyser (2003) argues, it is possible that the combination of relevant practice and feedback leads learners to come up with the rules that are provided as EI to PI groups. In this case, it becomes interesting to pinpoint under what circumstances EI may become necessary. Furthermore, in the studies reviewed above, learners were given a single chance to review EI prior to commencing the treatment. This represents a departure from the kind of access that learners generally have to EI. Furthermore, a single exposure would seem to make any positive effects EI could have contingent upon memory and the degree of attention participants paid during exposure, factors that, needless to say, were never controlled for. In addition to questions relating to the role of EI, as was mentioned at the outset of this paper, all the studies reviewed above featured designs that incorporated TE practice. Thus, examining what the impact of the presence and absence of TE would mean in terms of EI expendability would also be necessary in order to test whether benefits attributed to SI may be explained more parsimoniously.

2.2. EI in other SLA literature

Contrary to the impression one may get after looking at the PI literature, a rather large number of studies from a diverse set of angles have investigated the role of EI in L2 acquisition, only to find that it did indeed have beneficial effects. Three examples of this can be found in work by Ellis (1993) and Robinson (1996, 1997), which provide evidence on this matter from the perspective of exposure to EI at the receptive level.

In a study intended to test the effects of implicit vs. explicit learning, Ellis (1993) exposed learners to the soft mutation of Welsh initial consonants under four different conditions: a condition in which learners explicitly received the rules governing the mutation (the Rule group), a condition where participants were exposed to both the rules from condition one as well as illustrative examples (Rule & Instances group), and lastly two conditions aimed at mirroring naturalistic learning which exposed learners to the mutation, but provided neither EI nor examples (the Random groups)[7]. After practice activities consisting of translating phrases from Welsh into English, learners were administered a Grammaticality Judgment Test (GJT) to gauge gains derived from the treatment. In short, results showed that both groups receiving EI performed statistically better than the Random groups at discriminating legitimate and illegitimate instances of the mutation. Interestingly, Ellis (1993) also found that the gains were more pronounced for the Rules & Instances group, a fact which would suggest that although evidence from which rules can be derived sufficed for gains to be made, EI represented an additional benefit that facilitated the learning task.

Along the same lines of comparing implicit vs. explicit learning, further evidence speaking to the beneficial effects of EI comes from studies by Robinson (1996, 1997). Robinson (1996) studied the effects of instruction (in this study operationalized as explicit rule presentation) on acquisition of simple and complex English rules[8]. For this purpose, 103 Japanese learners of English were assigned to

---

[7] The reason why two random groups were included pertained to the amount of exposure. Given that the Random groups received neither EI nor examples, the exposure to the mutation received by learners in this group would have been less to that in the Rules and Rules & Instances groups. This would have introduced an unwanted additional difference that could have confounded results. In order to observe the effect of exposure as well as the effect of method of instruction, two Random groups were included, where one was requested to continue translating phrases past the point where they had learned them and until they had matched the exposure time of the two Rule groups.

[8] The complex rule pertained to the formation of pseudo-cleft sentences of location whereas the easy rule pertained to the fact that subject-verb inversion is possible in cases when adverbials of movement or location are fronted.

four different groups. Learners in the Implicit group were instructed to memorize sentences illustrating both the easy and complex rule. The incidental group was asked to read sentences for their meaning. A third group, the rule search group, was asked to read sentences and try to identify rules that governed the phenomena they exemplified. Finally, in the explicitly instructed group learners read the rules and proceeded to the training session (during which access to the rules remained possible). After exposure to 40 sentences illustrating the simple and the complex rule, learners were tested by way of a GJT. Results showed that, for the simple rule, EI appeared to have a beneficial effect, as shown by the greater accuracy the instructed group displayed as compared to the rule-search and the incidental groups. This, however, was only partially the case for the complex rule, for which groups exhibited better accuracy only when it came to the grammatical items of the GJT. Along the lines of Ellis (1993), Robinson suggested that EI combined with "structured exposure to relevant examples" offered greater benefits than EI in isolation, as shown by the superiority of the instructed group over the implicit one (Robinson 1996, p. 47).

Work by Robinson (1997) featured a design almost identical to the one laid out above. In this case, however, the rule-search condition was replaced by an enhanced condition. In the enhanced condition, learners were exposed to examples where items critical to the rule at issue[9] appeared in a box such that they were more salient. After exposure to 55 instances of the rule, learners were administered a GJT that featured both items which learners had encountered during the treatment and

[9] The rule in Robinson (1997) pertained to the fact that "monosyllabic verb stems can occur in either double object or *to*-object sentences, whereas verbs with disyllabic stems can only occur in *to*-object sentences" (Robinson 1997, p. 230)

novel ones. While all four groups exhibited a similar performance with regard to the items they had been previously exposed to, the instructed group's performance on novel items was statistically more accurate than that of the other three groups.

It is clear from the studies reviewed above that findings suggesting that EI can have a positive effect in L2 acquisition are not lacking. To complement these, we find studies, such as those by Alanen (1995), DeKeyser (1995) and De Graaff (1997), whose results are even more relevant to the extent that these, like the PI studies reviewed in the previous section, also explored the effects of EI in production.

For instance, Alanen (1995) examined the effects of EI and enhancement on the acquisition of Finnish[10] locative suffixes and consonant changes. The design of the study included three groups, one receiving a reading passage in which the consonantal changes had been enhanced (the Enhance group), a Rule group which received EI before reading an unenhanced version of the same passage, and finally the Rule & Enhance group, in which EI was followed by the enhanced reading passage. After treatment, participants were tested by way of GJT as well as a sentence completion task. With regard to production, results clearly showed that both groups who received EI scored significantly higher on the sentence completion task, suggesting that EI indeed had a positive effect. In the same way, results from the GJT revealed that the Rule group had outperformed the Enhance group. It was also the case, however, that the Rule and Enhance group scored lower than the other two groups. After reviewing data from think-aloud protocols, Alanen argued that the

[10] The target language of the experiment was actually not real Finnish but a modified version thereof.

language input received had led this group to overgeneralizations and concluded that all in all, EI had been shown to improve learner performance.

Further evidence to the same effect comes from a study by DeKeyser (1995), which, using an artificial language, looked at the effects of EI exposure from the explicit/implicit learning perspective. This investigation set out to study which learning condition, implicit-inductive (II) or explicit-deductive (ED), would be more effective for learning categorical and prototypical rules. Learners in both conditions were exposed to pictures with their corresponding captions in the artificial language (Implexan) through 20 sessions of 25 approximately minutes each. In addition to exposure to the images and Implexan sentences, learners in the ED condition were also provided with the grammar rules that governed the linguistic phenomena illustrated in the input. The grammar rules were presented on their computer screens at the beginning of the second, third and eleventh treatment sessions. Once they had concluded the treatment sessions, learners were administered judgment and production tests which contained previously encountered and novel items. Results indicated that the group receiving EI outperformed the II group in categorical rules, a finding which the author interpreted to mean that exposing learners to EI was superior to mere exposure.

Similarly beneficial results were reported by De Graaff (1997) in a study that examined the effects of EI on the acquisition of morphological and syntactic structures of varying complexity. De Graaff (1997) featured two learning conditions, implicit and explicit, in which learners were exposed to 150 hours of treatment on eXperanto, an artificial language based on (the also artificial) Esperanto. During the

10 sessions that comprised the treatment, participants in both groups received exposure to the language and performed vocabulary and grammar activities that involved both discreet and sentence level production. This study was ultimately conceptualized to observe the effect of explicit vs. implicit learning rather than EI, and, EI provision aside, the treatments also differed on the saliency of the target features, and on feedback. To be precise, the explicit group saw the target structures highlighted during practice and received feedback that, beyond simply indicating whether the response was right or wrong, also provided a grammatical explanation. Participants' progress was tested a total of three times, (twice during the treatment and once more after treatment completion) by way of a timed and an untimed GJT, fill in the gap, and contextualized vocabulary translation tasks. Although findings were unclear with regard to the effect of the different types of instruction over differentially complex structures, results clearly revealed a superior performance by the explicit group on all tests. Thus, in line with results from Robinson (1996, 1997), Alanen (1995) and DeKeyser (1995), findings from De Graaff (1997) suggest that EI is beneficial for L2 acquisition both at the receptive and productive levels.

The facilitative effect of EI has, in the same way, been supported by findings from the attentional literature in SLA. Studies such as the ones by Rosa and O'Neill (1999), Rosa and Leow (2004), and Sanz and Morgan-Short (2004) have examined this issue from the perspective of awareness and/or the effect of EI during treatment (i.e., feedback). Importantly, except for the last of those studies, Sanz and Morgan-Short (2004), findings from this line of research also counter the results reported in the PI literature. To be precise, their conclusions have pointed to EI as an effective means of raising attention and, thereby, increasing the likelihood of learning. This was the case for the investigation by Rosa and O'Neill (1999), in which the variables of ±rule search (RS) and ±formal instruction (FI) were manipulated in an attempt to examine the effects of awareness on intake. In order to establish whether metalinguistic explanations contributed to raising awareness in the first place, Rosa and O'Neill (1999) exposed two of their four experimental groups to explicit rules on contrary-to-fact conditionals in Spanish. Participants in the two groups receiving FI (+FI +RS and +FI -RS) were allowed 5 minutes to review explicit information on the target structure. Afterwards, they proceeded to complete a puzzle task where they were given one clause from a conditional sentence and were asked to provide the remaining one by interpreting a visual cue. While performing the task, participants in the +RS condition were urged to search for a rule that governed the target task whereas those in the –RS condition received instructions to memorize the propositional content in preparation for subsequent questions. The performance of [+FI +RS] and [+FI –RS] was compared to groups who completed the same task under the same conditions but who did not receive FI, as well as a control group that received no FI and was not directed to memorize or search for rules. Results showed that groups [+FI – RS] and [+FI +RS] made statistically more gains than the [–FI – RS] group, suggesting that EI contributed positively to intake. Interestingly, this study found no significant difference between the gains made by the +FI and the [–FI +RS] group. Rosa and O'Neill (1999) account for this concluding that the task- essential nature of their tasks combined with the provision of feedback might have fulfilled the same need as EI, resulting in comparable gains.[11]

[11] Due to the lack of statistical difference between control and experimental groups, some authors such

A beneficial effect for EI was recorded again in work by Rosa and Leow (2004), a study also primarily focused on examining the effects of awareness. The experiment featured five experimental groups and a control in which the variables of grammatical explanation as well as implicit and explicit feedback were manipulated under constant conditions of task-essentialness. The two groups receiving explicit information were provided with metalinguistic information and glossed examples about contrary-to-fact conditional sentences in Spanish before being asked to complete a problem-solving task in which the main clause of a conditional sentence needed to be matched with its correct subordinate. During the practice treatment, one of these groups received explicit feedback, while the other one received implicit feedback. The remainder of the experimental conditions included two groups identical to the ones just described except for the absence of EI, and a third where participants received EI only but no feedback of any kind. In addition, a control group received neither EI nor feedback and did not perform the task under conditions of task-essentialness. Learners were asked to narrate their thoughts as they performed the practice tasks so that investigators could subsequently gauge their awareness of rules. Immediately after the practice sessions, participants were given a multiple-choice recognition task and a sentence completion task. Finally, three weeks after treatment, learners were given a delayed posttest. Once again, results suggested that the higher level of explicitness learners had been exposed to, the higher the level of awareness and capacity to verbalize rules. By the same token, Rosa and Leow (2004) found that learners who reported higher levels of awareness had statistically higher

---

as Sanz and Morgan-Short (2004) cite Rosa and O'Neill (1999) as providing evidence that EI does not have a facilitative role in L2 learning.

scores than learners whose awareness was lower. Thus, their findings align with Rosa and O'Neill's (1999) in pointing to EI as a beneficial element in L2 learning.

Finally, Sanz and Morgan-Short (2004) conducted a PI study implementing methodology reminiscent of the awareness research tradition, but whose findings align with VanPatten and Oikkenon (1996). Designed to examine the individual and combined effects of EI and explicit feedback, this computer-mediated study featured four groups. Two of them, +E +F and +E –F, received explicit instruction. As their labeling suggests, one of these two groups received feedback throughout the treatment, while the other did not. The remaining two groups, –E +F and –E –F, also included one which received feedback and one where no feedback was provided, but these groups differed from the first two in that neither group received EI. Participants were randomly assigned to one of the groups and were provided with PI-based metalinguistic explanations on their computer screens (in the case of +E –F or +E+F), after which they were given practice in the form of SI activities. Participants in the –E groups went straight to the SI practice. Once treatment had been completed, learners were administered three exit tasks, a picture matching task, a sentence completion task and a video-retelling task. Results showed that, although all groups had improved from pretest to posttest, the gains obtained were not different across conditions for any of the contrasts suggesting that when SI is incorporated neither EI nor feedback is necessary for OVS acquisition.

This study is interesting for a number of reasons. First, Sanz and Morgan-Short's (2004) definition of SI is novel in that unlike VanPatten's and that of other authors, it overtly states that activities should be task-essential. Second, the lack of difference extended to group –E –F, suggesting SI is not only capable of making up for lack of EI (a result found before), but rather that it is as effective as both EI and feedback combined. In order to account for the lack of difference across groups, Sanz and Morgan-Short (2004) propose that the TE nature of their activities is enough for learners to obtain such gains. In the same vein, the claim is remarkable in that it echoes Rosa and O'Neill's (1999) explanation for lack of difference between their experimental and control groups. With two separate studies hinting at the same issue (albeit only in a post-hoc manner), and given that all studies reporting no EI effects included SI incorporating TE activities, the question as to whether EI may be expendable only under TE conditions inevitably emerges.

Chapter 3: Research Questions and Hypotheses

Based on the above, the present study sought to answer the following research questions:

RQ1: Is TEness critical for PI to result in benefits?

RQ2: Is EI superfluous only when SI practice is TE?

RQ3: Will any benefits obtained from practice persist across time?

Our expectation was that, in the absence of both TE and EI, PI would not result in any benefits. In addition, we propose that exposure to EI would generate benefits when SI (i.e., practice) was non-TE, and that these benefits would be less pronounced than those obtained from EI followed by TE practice. Conversely, we predicted that even without exposure to EI, when SI incorporated TE, practice would result in benefits, although we expected these to be less pronounced than the benefits obtained when participants were exposed to EI only. Lastly, it was also our expectation that while, in TE conditions, EI would not be necessary for benefits to obtain, combining both EI and TE would result in greater benefits than practice with one or the other only. The predictions laid out above yielded the following formal hypotheses:

H1: In the absence of both TE and EI, PI (i.e., SI only in this case) will result in no benefits.

H2: When practice is TE, EI will not be necessary for PI benefits to obtain. SI will suffice, because it is TE.

H3: EI alone will result in greater gains than SI, when SI is not TE.

H4: EI with TE will result in greater gains than EI alone.

H5: Only benefits obtained from [+TE SI] practice will prevail across time.

Chapter 4: Design

In order to test the hypotheses mentioned above the present study manipulated the variables of EI and TE as illustrated in Figure 2 below:

Group	Treatment	
Group 1	[–EI]	[–TE SI]
Group 2	[–EI]	[+TE SI]
Group 3	[+EI]	[–TE SI]
Group 4	[+EI]	[+TE SI]
Group 5	--	--

Figure 2: Design of the study

As can be seen, the study featured five groups, the first of which was not exposed either to EI nor to practice (i.e., SI) under task-essentialness conditions (Group 1, [-EI] [-TE SI]). The second group, [-EI] [+TE SI], did not receive EI but practiced under conditions of TE EI. Group 3, [+EI] [-TE SI], was the exact opposite case, with learners receiving EI but practicing (SI) under non-TE conditions. Finally, the fourth group, [+EI] [+TE SI], performed SI activities under TE conditions and received exposure to EI, as well. A fifth group, which did not receive any treatment acted as a control. This group performed computer-delivered activities on gender agreement but was tested on OVS and ser/estar.

The expectations were that Group 1, [–EI] [–TESI], whose members would receive no EI and only non-TE practice, would obtain no significant gains, and that its results would be on par with those from the control group (Group 5). In addition, we would expect any gains obtained by Group 1, [–EI] [–TESI] to be significantly lower than those obtained by the [+TE SI] and [+EI] groups. These findings would yield support for Hypothesis 1. Given that by definition task-essentialness is not a requirement of SI, outcomes to that effect would bear directly on the claim that PI can result in benefits in the absence of EI (VanPatten and Oikkenon, 1996 and others), suggesting, instead, that not all kinds of SI are necessarily sufficient for the benefits previously observed for PI to ensue. Related to this point, we expect Group 2, [–EI] [+TESI], to show significant gains as compared to those exhibited by Group 1, [–EI] [–TESI], and the control group (Group 5), suggesting that task-essentialness is critical for PI to be effective. Such results would lend support to Hypothesis 2 as well as to claims from previous research suggesting that when practice is TE, EI is not necessary (Sanz and Morgan-Short, 2004). As for Group 3, [+EI] [–TESI], and as stated in Hypothesis 3, we expected that in the absence of TE in practice, receiving rules and other metalinguistic information throughout the treatment would result in some gains from pretest to posttest or as compared to Group 1,[–EI] [–TE SI] and the control group (Group 5), and that these would be significantly higher than those obtained by Group 2, [– EI] [+TE SI], but significantly lower than those obtained by Group 4, [+EI][+TE SI]. Finally, with regard to Group 4, [+EI] [+TE SI], the expectation is that in support of Hypothesis 4, the combination of TE practice and EI suppliance throughout the treatment will effect statistically higher gains from pretest to posttest compared to Group 2, [- EI] [+TE SI]. Previous research on the role of EI relative to the PI effectiveness has yielded results to the contrary; the present work differs crucially however, from previous studies in that EI will be administered periodically during the treatment rather than only prior to it. If Hypothesis 4 were borne out, this too would bear on previous findings suggesting that EI offers no additional benefit to PI (VanPatten and Oikkenon 1996 and others). After all, such results would indicate that if not necessary, EI constitutes an added advantage that, along with TE SI, can have a facilitative effect on the process of acquisition. A final prediction regarding the performance of the groups is that for the [+TE SI] only condition, whatever gains are recorded in the immediate posttest will endure over time and manifest themselves again in the delayed posttest.

Chapter 5: Method

5.1. Participants

Participants for the present study were NS of English enrolled in first- and second-semester Spanish courses in the Spring term of 2012[12]. This population was chosen on the assumption that learners enrolled in courses of this level would likely have been exposed to the target structures chosen for this study (direct object pronouns and copula verbs, ser and estar), prior to our treatment but would possess declarative knowledge of either phenomenon at best. This minimal incipient knowledge was deemed desirable to make learners more receptive to the treatment and in order to be able to observe gains after a single treatment session per structure. Respondents in their second semester of Spanish were recruited during the first weeks of the spring 2012 semester, so that they were virtually no different from an experienced first semester learner. Participants from first semester were approached eight[13] weeks into the semester, so that they had had some exposure to the language, were familiar enough with the input featured in the practice items, and, therefore, were receptive to the treatment. The researcher visited all first and second semester classes to announce the study and recruit volunteers to take part in situ. A total of 242 learners, aged 18 to 32, took part. Of those, 23 were eliminated from the pool for being heritage speakers of other languages or for not speaking English natively. Data of 18 participants who reported having had three or more years of instruction in other

[12] Learners were placed in their levels as per their scores in a placement test.
[13] Starting the treatment around week 8 allowed enough time until the end of the semester for its completion.

foreign languages were also withdrawn. In order to preserve as many participants as possible, any other elimination was done on a task-by-task basis. In other words, any participant who did not complete a session (i.e., the pre-, post-, or delayed posttest) or one more tasks was eliminated from the analysis pertaining to that task in particular but left in for analysis of any other task where he/she had completed all testing sessions. As this procedure yielded slightly different numbers for each analysis, the exact n size for each task is specified in the corresponding part of the results section. All sessions demanded by the study took place within class time either in the regular classroom or in a computer lab.

5.2. Target Structures: OVS structures and ser/estar

The linguistic phenomena chosen to test the hypotheses above are Object (O) Verb (V) Subject (S) sentences as well as the ser/estar distinction in Spanish. Although the canonical order in the target language is SVO, due to its rich morphology, Spanish exhibits a fairly flexible word order. Structures featuring the direct object pronoun in sentence initial position, and in which the subject appears post-verbally, are both grammatical and common in the language. It is a documented phenomenon that such structures are problematic for language learners both in L1 and L2 acquisition. Evidence from various studies indicates that both first (Bates et al., 1984; Bever, 1970) and second language learners (Gass, 1989; LoCoco, 1987) tend to interpret the first noun phrase (NP) in any sentence as the subject (VanPatten's First Noun Principle). When parsing OVS structures, the strategy of parsing the first NP as the subject results in the erroneous allocation of agent and theme roles. Thus, in a sentence such as "lo abraza María" (*Him embraces María*), learners tend to understand the sentence as "Él abraza a María" (*He embraces María*). Perhaps because they represent a departure from default processing strategies of NSs of English, OVS structures have been the target structure of choice for the vast majority of studies seeking to test the benefits of PI. In light of that fact, and in an attempt to obtain results that are interpretable within the existing research framework, the present study will also adopt OVS sentences as one of its target structures.

In addition to OVS, this study sought to examine the effects of EI and TE on the acquisition of a second linguistic phenomenon, namely, the ser/estar distinction in verb+adjective predicates. The rationale to include this second structure is that, given the preponderance of OVS structures in PI research, it is important to test whether benefits recorded for PI extend to structures that pose different processing problems. Ser/estar has been examined in one PI study, reported in Cheng (2002, 2004), which investigated the acquisition of this distinction from the pure TI vs. PI vantage point. The present study is an attempt to contribute to the rather scant evidence about the suitability of PI to address the processing problem this structure poses.

Indisputably a difficult distinction, ser/estar arguably constitutes a different processing problem from OVS. First, rather than drawing on word order or the First Noun Principle,[14] the problematic nature of ser/estar presumably originates in the fact that it requires learners to perceive a difference not present in their NL. While correct

[14] Cheng (2004) argues ser/estar draws on the Lexical Preference Principle (LPP), since learners tend to forego the verb and simply focus on the adjective that follows it. Although of the Input Processing (IP) Principles proposed by VanPatten (2007) and Lee and VanPatten (1995) the LPP would indeed somewhat capture the learnability problem posed by ser/estar, it would seem that the difficulty of this distinction originates to a far greater extent from other factors, where the LPP only exacerbates the issue. For that reason, the present paper will not attempt to couple ser/estar with a principle as was done for the First Noun Principle-OVS coupling.

OVS processing demands that learners refrain from automatically processing the first noun as the subject (thus allowing for the possibility that the theme may come first), in the case of the ser/estar distinction learners are faced with the need to separate a single element of their NL into two different representations and to learn the specific rules that govern the use of one as opposed to the other in different contexts. The task of identifying contexts that call for each of the two different copula manifestations is made particularly difficult by a number of factors. First, as already mentioned, both ser and estar map onto a single element – the verb 'to be' in English. Therefore, learners need to grasp which uses of the verb 'to be' should be mapped onto ser versus estar. In addition, although there are certain environments in which ser and estar occur in obligatory complementary distribution, the majority of predicates will allow both verbs with the choice of one versus the other resulting in semantic changes that range in subtlety. Thus, more often than not judgment of the linguistic context on a case-by-case basis becomes necessary, as application of a merely form-based rule or rote memorization of expressions that go with ser and estar are not viable options. This involves judging the nature of predicates and the adjectives contained in them in terms of dichotomous categories such as "inherent versus circumstantial" and "permanent vs. temporary". However, those distinctions only apply probabilistically, since additional contextual factors, including speakers' perception of perfectivity, are often decisive to deem a trait as inherent or permanent vs. circumstantial or temporary. This means that where object pronouns offer virtually no variation[15] in terms of the meaning they map onto, copula choice is notorious for native speaker

[15] One might argue that the phenomena known as *loísmo* and *leísmo* present in certain varieties of Spanish could constitute variation as to the meaning that direct object pronouns map on to. However, it by all accounts most unlikely that learners should be exposed to that variety at all in instructed settings.

variation (Silva-Corvalán, 1994). Thus, one might argue that whereas, in the case of object pronouns integration into IL grammars is hindered by constant connection to the wrong meaning, driven by a deeply ingrained processing strategy, the incorporation of ser/estar distinction into IL grammar is arrested by the inability to delineate the meaning that should be mapped to each of the copula verbs.

Unsurprisingly, research on the acquisition of ser and estar has shown that it is precisely those instances in which both ser and estar verbs are possible options that are acquired latest, suggesting that they are more difficult for learners to acquire (VanPatten, 1985; 1987)[16]. As it seems that it is in those predicates that the heart of the *ser/estar* problem lies, the present project investigated exclusively adjective predicates where both ser and estar could be correct.

It must be noted that two authors (Silva-Corvalán 1986, 1994; Geeslin 2000, 2001, 2002, 2003) have argued that the binary dichotomies used to analyze learner data in studies advancing ser/estar learning stages are an unsuitable unit of analysis as they fail to take into consideration contextual factors. Proponents of this view have looked at ser/estar use in terms of the contextual features they appeared surrounded by, arguing that these could provide a more valid representation of how learners approach the distinction in general and, thus, result in more accurate description of

---

[16] In an expansion of the categories presented in his 1985 study, VanPatten (1987) advanced the following learning stages for ser/estar:
Stage 1: Omission of any copula
Stage 2: Ser chosen across the board
Stage 3: Emergence of ESTAR in the progressive
Stage 4: Emergence of ESTAR with locatives
Stage 5: Emergence of ESTAR with adjectives denoting condition

Ryan and Lafford (1992), Gunterman (1992) and Briscoe (1995) reported findings in agreement with the stages above.

their IL grammars. Geeslin (2000) was the first researcher to adopt this approach in a study that included ten contextual features (see Figure 3) as potential predictors of the appearance of *estar* in a forward regression analysis.

Significant predictors of estar *by level of enrollment in Geeslin (2000)*

Variable	Level 1 (low)	Level 2	Level 3	Level 4 (high)
Adjective class	X	X	X	
Semantic transparency			X	X
Animacy				X
Susceptibility to change	X			
Dependence on experience	X	X		X
Directionality	X		X	
Dynamicity	X	X	X	X
Perfectivity	X	X	X	X
Frame of reference			X	X
Telicity			X	

Figure 3: Contextual features in Geeslin (2000)

Results from 77 high school Spanish learners of four different levels of proficiency revealed that all ten features were significant predictors at some point in the acquisitional continuum recreated in the study, but they interacted differently depending on learner ability.

Importantly, using a revised coding scheme, Geeslin (2002, 2003)[17] noted that while NS appear to base their copula choice on semantic features, (i.e., on whether the attribute was permanent or not, to be precise), advanced NNS grammars appear to be governed by pragmatic features pertaining to frame of reference, i.e., whether the referent was being compared to itself or not.

[17] Geeslin (2002) looked at the same data from Geeslin (2001), while Geeslin (2003) reported on a new data set obtained from NS of Spanish and very advanced NNS learners.

Similar results were obtained by Cheng et al. (2008), in a corpus study, which, using the coding criteria based on Geeslin (2001, 2003),[18] found that beginner Chinese learners of Spanish appeared to rely on different cues to determine when estar was needed than the more advanced ones. Namely, while beginners appeared to produce estar+adjective driven by lexical cues such as formulaic expressions in the linguistic environment, learners of higher proficiency appeared to base their judgment on pragmatic considerations. This consonance in results is especially interesting given the dramatically different language background of the learners featured in each of the two studies, which suggests that the transition in reliance from lexical to pragmatic cues may obtain cross-linguistically. A lack of NL effect is further found in Geeslin and Guijarro-Fuentes (2005) and Geeslin and Guijarro-Fuentes (2006), who, using German, French, English and Portuguese participants found that the NNS tendency to rely on pragmatic cues still obtained.

In addition to the interaction between lexical and pragmatic aspects, Cheng et al. (2008) also concluded that the type of discourse seemed to have an impact on copula choice, advancing that estar+adjective was more pervasive in descriptive essays. Driven by this finding, Collentine (2010) researched the lexico-grammatical features that co-occur with ser+adjective and estar+adjective, as well as the discursive functions they each serve. His results revealed a greater preponderance of estar+adjective in texts if greater grammatical and syntactic complexity accompanied

[18] In addition to the differences in the linguistic environments that were considered predictive, Cheng et al. (2008) also introduced a different type of regression analysis. While Geeslin (2001, 2003) use forward stepwise regression models Cheng et al. (2008) analyzed their data by way of the Poisson regression model. Collentine (2008) warns that this statistical analysis may not be appropriate to analyze the data in Cheng et al. (2008), as this type of regression assumes the occurrence of *estar*+adjective to be random and unpredictable.

by a reduced presence of ser+adjective in such text types, where only the former was consistent with native use. The analyses in that study also found that estar+adjective appears in discourse with a lower level of inflectional sophistication, with the opposite being the case for ser+adjective. Collentine argues that such a scenario, where the structure in the inflectionally more sophisticated environment appears in texts of lower complexity, and the structure in the less inflectionally sophisticated environment appears in text types that are more complex, suggests that copula choice for NNSs is also governed by the amount of processing demands available.

From the descriptions above it is clear that the variationist approach to ser/estar has identified a number of contexts that give evidence of becoming significant at different points in the acquisition process. Despite a certain consensus, however, the number and types of relevant context that appear to be relevant vary depending on the researchers' conceptualization implemented as well as factors such as learner proficiency. In addition to this, given the somewhat abstract description of each context, the relevant linguistic environments for copula choice seen overlapping and can be difficult to operationalize, at best. At a practical level, following the more fine-grained classifications implemented in these studies, coming up with large numbers of instances that unequivocally belong under a single category becomes virtually impossible. Given these circumstances, this paper chose to present ser/estar to participants primarily according to the more parsimonious inherent vs. circumstantial distinction. Thus, we considered any instances of ser/estar utterances where copula choice could be successfully decided upon following the inherent vs. circumstantial rule suitable for inclusion in the study. Following Silva-Corvalán

(1994), the EI featured in our experimental conditions pointed out that estar is used with circumstantial attributes (meaning that estar appears with an adjective to express a state rather than an inherent condition, such as in 'Juan está tranquilo', Juan feels calm). All practice tasks and tests in the present study are comprised of uses of estar that bear on these contextual features. As suggested earlier, the rationale behind the adoption of these specific contextual features over the rest is twofold. Not only are they more amenable to clear-cut encapsulation in a rule; most importantly, the contexts they describe are the ones that adapt best to graphic representation and are, therefore, a good fit for a picture-matching task format, as described in the next section.

5.3. Materials and Instruments

The present study featured materials consisting of two different elements: explicit information (EI) and practice tasks. EI was composed of two slides where the rules concerning OVS and *ser/estar* were laid out and furnished with examples (Appendix A).[19] In the case of OVS learners read about the direct object pronouns and their function as replacements for the direct object, i.e., "the part of the sentence receiving the action of the verb." Importantly, the EI pointed out that, with simple verb forms, object pronouns are always located before the verb and that it is possible and very common to find the subject dislocated to a post-verbal position. In order to further clarify this point, learners were prompted to note the difference between third person singular personal pronouns (él, ella), which refer to the agent or doer of the action expressed by the verb, and their direct object counterparts (lo, la), which refer

[19] EI for ser/estar in this study were based on Cheng (2004).

to the recipient of the action. Finally, learners were warned that due to their being NS of English, a language in which the noun in sentence-initial position is always the subject, they would be tempted to process the object pronouns lo/la at the beginning of an utterance as though they were él/ella, so they should be careful to pay attention to those elements and distinguish them appropriately.

In the case of ser/estar the EI spelled out that Spanish has two equivalents of the English verb 'to be,' namely ser and estar. Learners were informed that the former is generally used to refer to inherent, or ingrained traits, as opposed to estar, which is used for traits that are caused by a circumstance. In terms of the strategy, learners were urged to make a point to notice the verb, as it is the verb form that conveys the information as to whether a trait is inherent or not, while the adjective itself remains the same. For both structures, the metalinguistic information was complemented with examples and a brief paragraph interpreting them. In addition, in an attempt to try and ensure that participants would pay attention to the content, a slide containing a question about the information just read was included after each EI slide (Appendix B). These questions tested the understanding of very basic facts conveyed in the EI and were followed by feedback as to whether the correct response option had been chosen. Participants were privy to the fact that their understanding of the EI would be tested. In order to maintain treatment on both structures as equal as possible, a conscious effort was made to keep the EI for OVS and ser/estar roughly at the same length, and distribute the information across slides in an equivalent manner.

As for practice, the second component of the materials in this study, it consisted of picture matching and sentence interpretation activities for each of the target structures. Two versions of each type of task were created. One version incorporated task-essentialness (Appendix C) while the other one did not (see Appendix D). In the non-TE items, learners were exposed to the object pronouns and they interpreted them (whether tacitly or not) as the theme of the sentence, as the two responses always depicted them as such. The task in non-TE items required focusing on the number of the object pronoun, or the semantics of the verb. Thus, although the nature of the task promoted processing of the object pronoun, failure to link the target form to its meaning did not result in incorrect response choice; in other words, the target form was not essential for successful task completion. On the contrary, in the TE condition the task required learners to choose from response options that presented the subject and object in the prompt sentence correctly as the agent and theme (respectively) and incorrectly as the theme and agent (respectively). In the TE items, then, learners were required to decide whether the object pronoun in sentence initial position represented the doer or recipient of the action in order to successfully pick a response option. Thus, the target form was essential for successful task completion. In order to prevent items in the OVS treatment from becoming mechanical drills (stimulus-response pairs), eight items that featured SVO order were also included in the practice session (Appendix E). This was not deemed necessary in the case of *ser/estar*, as alternation occurred throughout that treatment by virtue of having two verbs. In addition to the experimental practice described above, an additional picture-matching test targeting gender agreement was created for participants in the control group. This task was purposefully designed to target a grammar point irrelevant to the present study and helped keep the procedure for all learners homogeneous, thereby making the treatment as similar as possible across groups, except for the critical features of each condition.

The EI slides and the picture-matching task in both their versions were combined to create the conditions [+/-TE] and [+/-EI] described above. As was mentioned previously, [+EI] conditions in the present study encompassed exposure to EI not once, but rather a total of five times throughout the course of the session. To be precise, participants were asked to review the EI right before starting the treatment, then again after item 1, (i.e., they had the chance to check their comprehension of the metalinguistic information), and three times more after target items 4, 8 and 16, as represented in Figure 3 below:

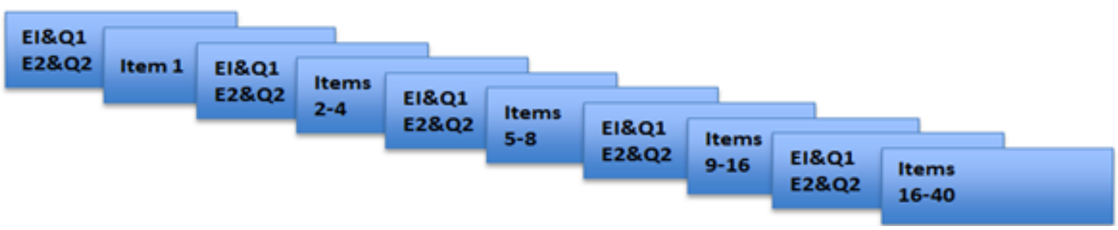


Figure 4: Sequence of EI and Practice Items in [+EI] conditions.

All efforts were made to use cognates and/or high frequency lexical items to prevent vocabulary comprehension issues from hindering the establishment of form-meaning connections that the practice intended to trigger. In addition, a gloss was provided for all verbs, in OVS items, and adjectives in ser/estar items, to ensure that vocabulary problems did not interfere with learner performance and learning. In both cases, the treatment presented the 20 sentence-matching task items first and only then the 20 sentence interpretation task items.

A task corresponding to each of the outcome measures (a sentence-matching task, a sentence interpretation task and production task) was administered to participants in the three testing sessions, namely pretest, immediate posttest and delayed posttest, as Figure 5 below illustrates:

Structure	Pretest	Treatment	Immediate Posttest	Delayed Posttest
OVS	Production Task	Picture Matching & Sentence Interpretation Tasks	Production Task	Production Task
	Picture Matching Task		Picture Matching Task	Picture Matching Task
	Sentence Interpretation Task		Sentence Interpretation Task	Sentence Interpretation Task
ser/estar	Production Task	Picture Matching & Sentence Interpretation Tasks	Production Task	Production Task
	Picture Matching Task		Picture Matching Task	Picture Matching Task
	Sentence Interpretation Task		Sentence Interpretation Task	Sentence Interpretation Task

Figure 5: Description of tasks per target form and experimental session.

In an attempt to control for learning at testing time as much as possible, a decision was made to start with the production task, continue with the picture matching task, and, finally, administer the sentence interpretation task, in that order. The assumption was that the opposite sequence, namely interpretation preceding production, might make it easier for learners to derive the governing rule, possibly resulting in production scores not being entirely a product of the treatment.

The purpose of the pretest was to measure participants' proficiency with regard to the target structure of the experiment prior to the treatment. The immediate posttest measured gains resulting from the treatment, and the delayed posttest assessed to what extent such gains were still present after a longer period of time (in this case three weeks). All items used in the materials for this study were piloted with NSs of Spanish to ensure that learner responses could be safely attributed to knowledge of the linguistic item and not the transparency of the prompts chosen. Items included in the pretest and immediate posttest were piloted by 14 NSs of Spanish. All items included in the delayed posttest were piloted by 11 NSs of Spanish, none of whom piloted the pretest and immediate posttest items. The two-stage piloting with different pools of NSs simply reflected to the timing of item-creation and NS availability. Whereas it became immediately obvious upon reviewing pilot scores that judgment of all OVS items was virtually 100% consistent across the board, native judgments on *ser/estar* presented some variation, as expected. Given the number of *ser/estar* items needed for this study (a total of 180 distinct *ser/estar* items, a third of which needed to be amenable to graphic representation) it was not possible to adopt an inclusion criterion of 100% agreement in NS judgment. Instead, only items where more than one NS disagreed on which copula was needed were excluded from the pool of items. In other words, a threshold of 92% agreement was set for item inclusion. Below we provide a more detailed description of each of the materials and instruments mentioned above.

Pretest

As mentioned above, three tasks, a production task, a picture matching and a sentence interpretation task, were administered in order to gauge learners' knowledge of the target structures prior to the treatment. In the case of picture matching and sentence interpretation tasks, the format and specifications were identical to those of the task-essential activities used during the treatment (see Appendix B)[20]. Unlike the treatment, however, the picture matching and sentence interpretation tasks in the pretest (as well as the production task, for that matter) were all paper-based. The motivation not to adopt a computer-based delivery was one of convenience, as the paper format allowed us to rapidly administer testing instruments to intact classes.[21] Like the pretest, post-treatment exits tasks were also paper-based.

The picture-matching task comprised a total of 40 target items. Of those, 20 gauged learners' knowledge of OVS sentences, and the remaining 20 addressed their proficiency with the ser/estar contrast. Each item featured a sentence accompanied by two pictures. Participants were required to decide which of the two pictures corresponded to the sentence. Over the course of the pretest, participants saw ten sentences consisting of a conjugated form of the verb ser and an adjective and another

---

[20] In fact, given that no feedback was provided, the sentence interpretation and sentence-matching items used in the pretest were recycled and used during the treatment session. Once again, this was motivated by the large ser/estar items needed for the study. No other items were recycled throughout the study.

[21] Administering the testing instrument in paper-based format allowed us to be less disruptive in the classes where data were collected. In addition to this, the language lab is in high demand; as a result the number of times that each class can have access to it is limited and fixed in accordance to the needs of all language courses. Conducting the five meetings that participation in this study required in the lab would have monopolized all lab hours from the courses involved in the research. For these and other reasons, it was deemed preferable to have all our outcome measures be paper-based, such that data collection for pretest and delayed posttest could be carried out in the regular classroom, while also keeping testing format as similar as possible from one session to another.

ten consisting of estar and an adjective. OVS items were formed by a third person singular object pronoun, a verb and a subject (in the third person as well). Half of the object pronouns were feminine, and the other half were masculine. The sentences described a two-participant event involving an agent and a theme. All OVS events featured in the testing instruments were neutral, i.e., there was no higher probability for one of the participants to be in a particular role. This was done in an attempt to prevent event probability issues from unduly biasing learner interpretation. In order to avoid ambiguity, theme and agent were always of different gender. As was the case with ser/estar, each sentence was accompanied by two pictures which participants were asked to choose from. All items in the pretest were presented in random order. Following VanPatten (1995), object pronouns and forms of both *ser* and *estar* were limited to the third person singular. In order to comply with SI practice recommendations, the critical item was placed in sentence-initial position to make it as salient as possible.

Following VanPatten and Cadierno (1993) and VanPatten and Oikkenon (1996), in addition to the target items, 40 SVO sentence-picture pairings that did not include the target structure were used as distracters (see Appendix E). Their goal was to conceal the purpose of the study, so that participant responses would remain as unbiased as possible. In an attempt to make distracters effective, it was decided they should address gender agreement, as this is often perceived as a learnability problem by instructors and learners alike and, therefore, represented a credible potential target. In order to follow the format of the target items exactly, in the picture matching task distracters provided a prompt adjective, which was always a color term. The adjective was inflected in the masculine or feminine and followed by two pictures, featuring a masculine and a feminine item of the color expressed by the prompt. In order not to make the correct response too obvious, about a third of the objects in the pictures were nouns ending in –e, whose gender is not immediately obvious from their morphology. As with the critical items, participants were asked to match one of the pictures to the adjective in the stimulus sentence.

In the sentence interpretation task, distribution of target items was identical to the one featured in the picture matching task. In this case, however, learners were asked to match a sentence containing the target item to one of two interpretations. The interpretations were in English to ensure proper comprehension and avoid a second language effect. Items for OVS offered two response options: one that presented the clitic as the agent of the sentence and one that featured the clitic as the theme. Items for ser/estar feature response options where the adjective in the sentence is presented both as an inherent and as a circumstantial condition (see Appendix C).

In order to maintain coherence with the picture-matching task, gender-agreement distracters were included. The format, in this case, included a prompt sentence consisting of a noun phrase, a verb and an object. Importantly, the noun phrase was sometimes formed by singular demonstrative pronouns (este/esta) and a noun ending in –e or –ista, such that the morphological endings were not the better known masculine –o and feminine –a.[22] Participants were asked to choose whether

[22] 20 of the 40 distracters in the sentence interpretation task used in the pretest featured a slightly different format. They were formed by noun phrases consisting of determiners and nouns, followed by a verb and a prepositional phrase with *para*. Learners had to choose between two response options, one where the subject was the agent and one where the subject was wrongly interpreted as the head of the

the sentences referred to a man or to a woman. Distracters purposefully avoided using linguistic items targeted in this study and favored structures that were both frequent and resulted in successful interpretation when default processing strategies were applied. The information needed to successfully choose the correct picture in distracter items bore on the processing of a morpheme for gender information, rather than on overcoming the default instinct to process the first word in a sentence as the subject or on the need to develop two distinct representations corresponding to one item in the NL.

Although, in compliance with PI requirements, our learners never had to produce either of the target forms during the treatment, this study included a production task in its testing instruments (see Appendix F). A production task was included in order to test whether the benefits of PI extended beyond perception and whether they did so equally under TE and non-TE conditions. This task partially mirrored VanPatten and Cadierno's (1993), thus essentially taking the form of a fill-in-the-gap where either the feminine of masculine object pronoun (in the case of OVS) or the correct verb (ser or estar) needed to be supplied. To be precise, production items featured a single picture, which was complemented by a prompt sentence consisting of a gap followed by a verb, in the case of OVS, or an adjective, in the case of ser/estar. In order to prevent learners from providing answers that would be difficult to interpret, each item in the production task was preceded by a line prompting participants to fill in the gap with one of two items. Whereas OVS items listed lo/la as the options, ser/estar items provided a choice between es and está.

---

prepositional phrase. It was deemed more appropriate to resort exclusively to distracters that featured demonstrative pronouns in subsequent testing.

Although providing two options made this task as close to a multiple-choice task as a production task can be, it was still informative in a way that neither picture matching nor sentence interpretation tasks could (a claim that was supported by our results). Namely, both the picture matching and the sentence interpretations tasks in the study essentially fed learners the complete range of possible meanings attached to the linguistic forms targeted in the study, thereby restricting the number of interpretations possible and allowing for elimination to figure into the answering process. The production task, on the other hand, required learners to know what interpretation the images wished to convey and then pick an option accordingly. For that to be possible, based on the picture provided, learners had to be able to make a given interpretation on their own in order to be able to decide systematically. In addition, in OVS items, upon glancing at the prompt image, it was presumably entirely clear to our participants who they were trying to convey as the theme and whom as an agent, whether they could grasp that the two options provided forced them to focus exclusively on the theme or not. In the case of ser/estar, it is reasonable to assume that learners faced the problem of deciding whether the picture represented an inherent or a circumstantial condition (or something entirely different), regardless of whether they knew which of the two options to match to each of the interpretations.

As was the case with the picture-matching task, the production task included a total of 40 target items, with 20 addressing OVS and the remaining 20 addressing ser/estar. In the same way, the production task also included 40 gender-agreement based distracters. No feedback was given during the pretest.

Given that no feedback was provided during the pretest, both ser/estar and OVS items were reused during the treatment sessions. This was motivated by the large numbers of items required by the study and the difficulty in creating ser/estar items that were amenable to graphic representation. Despite the absence of feedback during the pretest, recycling the materials from the ser/estar task increased the chances for a practice effect to influence our data. In this respect, comparison with data obtained from the control group was crucial, as any advantage originating in the nature of the materials themselves would have benefited the results of the non-treatment group as well. The presence of the control group allowed us to obtain a clearer impression of what gains could be safely attributed to the treatment.

The rationale to maintain the sentence-matching task, rather than simply going with the sentence interpretation task was comparability. On the one hand, given that prior PI literature targeting OVS implemented sentence-matching tasks in both practice and testing tasks, it was ideal to incorporate that same format in the present project. By the same logic, in order to compare the effects of the treatment over structures that constitute different processing problems it is desirable to have identically formatted outcome measures for both our target structures. Data obtained in the pretest were used as a covariate when analyzing results.

Posttest

Immediately after the treatment session for each of the target structures learners were administered a posttest in the form of paper-based sentence-matching, sentence-interpretation, and production posttests. All tasks were identical in format to those in the pretest, and therefore included target items as well as distracters. However, all items in the posttest (immediate and delayed) were novel items that did not appear either in the pretests or in the treatment.

Delayed Posttest

A delayed posttest was administered three weeks after each of the treatment sessions to observe the durability of the gains. The format and number of items in the delayed posttest was identical to those of the immediate posttest and both target items as well as distracters were included.

5.4. Procedure

Participants attended five different sessions to complete the treatment of the present study. In an initial meeting they were given a background questionnaire (Appendix F), provided with all the necessary documents to record their consent, and subsequently given the pretest. During the second meeting, participants were randomly assigned to one of the five conditions and engaged in the treatment of one of the two structures; they completed the computer-delivered sentence-matching task and sentence interpretation tasks, and read the EI (if assigned to a [+EI] condition). The third meeting was identical to the second one with the only difference being the linguistic structure targeted in the treatment. Second-semester students covered OVS in the second session of the experiment and ser/estar in the third, whereas first-semester participants completed treatment sessions in the opposite order. The rationale for the reversed order originated in the ser/estar pretest data, which in the case of second-semester learners revealed a strong tendency to a ceiling effect in two out of the three tasks they completed. To increase our chances of obtaining meaningful data, then, upon administering the treatment to first-semester students, it was decided to go with ser/estar first, thereby reducing the amount of exposure to this distinction somewhat.

Treatment

Each task consisted of 40 items different from those featured in the testing instruments; these items were presented through the Respondus StudyMate software. For each item, participants saw an OVS sentence followed by two pictures (in the case of the picture matching task) or two sentences (for the sentence interpretation) marked as A and B. In the critical items of [+TE SI] conditions, as mentioned earlier, this sentence consisted of a third-person object pronoun followed by a verb and finally the subject (See Appendix B). One of the two response options illustrated the theme of the sentence as the recipient of the action expressed by the verb and the subject as the agent. Crucially, the other response option illustrated the opposite scenario: it portrayed the theme of the sentence as the agent of the action expressed by the verb and the subject of the sentence as the recipient of the action. That being the case, learners in the [+TE SI] conditions needed to process the object pronouns (placed in the initial position in the sentence) and connect that form to its corresponding meaning of recipient (as opposed to agent). For [-TE SI] conditions, learners saw the exact same sentence but the pictures did not illustrate a reversal of the roles of the agent and theme of the target sentence (See Appendix D). In order to respond accurately to each item, learners in all conditions needed to read the target sentences. In all cases the target object pronoun was in the most salient, sentence-initial, position. Accordingly, only for those in the [+TE SI] conditions was assigning the correct meaning to the target item crucial for successful completion of the task.

Participants were asked to use their mouse to click on the picture or interpretation that corresponded to the sentence at the top of their screen. After they picked a response, a little box appeared informing them whether their choice was correct, but no explanation as to why was provided, in order not to confound the effects of feedback with EI, especially in the non-EI conditions. Participants were able to proceed to the next item by clicking on a button in the left hand corner of the screen; they could not backtrack.

Unlike in previous PI experiments, where EI exposure took place exclusively prior to engaging in practice, the [+EI] conditions of this study exposed learners to EI periodically during the treatment. To be precise, learners in [+EI] conditions commenced each treatment session by going over the EI slides (see Appendix A). After this initial exposure, EI slides appeared on the screen after item 2 and again after items 6, 8 and 14. Thus, learners had the chance to access the EI a total of 5 times. The rationale to include EI periodically is twofold. For one thing, it is that is hardly representative of the classroom environment to expose learners to EI once. Moreover, by exposing learners to EI only at the beginning of the treatment, the variable of EI could very well have been modulated by memory. In addition, it was decided to incorporate the EI within the practice items, as opposed to making it available on paper for learners to check as needed, in order to exert control over the number of times EI was read. In this respect, in order to ensure that participants actually did read the EI, a brief question was included after the last EI slide before proceeding to the next practice item. This was a yes/no question about the content of the rule or about the information in one of the examples provided in the EI (see Appendix B). Immediately after concluding the treatment sessions, learners received the immediate posttest on OVS.

The procedure in the third session was identical to the one in the second session with the only difference being that the 40-item sentence-matching task focused on ser/estar. Finally, a fourth and fifth session in which learners completed the delayed posttest for each of the tasks concluded the experiment. The time elapsed between each of the treatment sessions (3 weeks) was kept constant across first- and second-semester students.

5.5. Data Analysis

Results from this study were analyzed in terms of accuracy. The expectation was that the most effective treatment would result in the fewest errors, both in the interpretation and production of OVS sentences, and similarly for sentences with ser/estar.

After submitting the scores to Rasch analysis[23], posttest and delayed posttest scores for each condition of OVS and ser/estar were submitted to an ANCOVA, with pretest scores as the covariate. Scores for each task were submitted to a separate analysis, for the posttest and delayed posttest scores. Using pretest scores as a

[23] Rash analysis provides index measures that allow for identification of participants whose performance patterns suggest random guessing. Participants who exhibited misfitting indices in the pretest were excluded from the study. No further weeding was carried out.

covariate provided a more accurate measurement of the effect of the treatments, as it factored out the effects of previous knowledge for each of the groups.

If TE were critical for PI benefits to obtain, we would expect the scores from Groups 2, [-EI] [+TE SI] and 4, [+EI] [+TE SI] in the ANCOVA to be statistically higher compared to those obtained by the other two groups. In the same way, support for the claim that when practice is TE, EI is not necessary for PI benefits to obtain would require this analysis to yield scores for both target tasks that are statistically higher for Group 3, [-EI] [+TE SI], than for Group 1, [-EI] [-TE SI] and the control. If EI were found to be more beneficial than PI alone we would expect the ANCOVA to show that scores for Group 4, [+EI] [+TE SI] are statistically higher than those of the remaining three groups. In addition, in order for the prediction that gains from treatment will be maintained across time to obtain, scores which were statistically higher in the immediate posttest as compared to the pretest should once again be statistically higher in the delayed posttest. Finally, if benefits from SI practice extend to production (in spite of the practice not involving any production), we would expect the statistical gains recorded in the interpretation task to manifest themselves in the production task as well. To conclude, given that *ser/estar* and OVS represent different processing problems for learners, there is a chance that the benefits recorded for each of the structures are different. Given the parallel formats used in the treatments, significantly different gains for one of the structures would suggest that the treatment in question is differentially effective for a certain type of processing problem, or alternatively that one of the structures appears to be more difficult than the other.

Chapter 6: Results

6.1. OVS Production – Immediate Posttest

Table 1 displays the number of participants in each of the five groups in the study who completed all stages of the OVS production. Although participants were assigned to their respective groups randomly, as Table 1 reveals, Groups 1, [-EI][-TE SI], and 2, [-EI][+TE SI], were the largest, with 35 and 30 subjects respectively, followed by Groups 3, [+EI][-TE SI], and 4, [+EI][+TE SI], both of which had 24 participants, and finally the control group, which gathered 19 participants. In the case of the OVS production immediate posttest, the total n for OVS production amounted to 132 participants. (The n size for all the analyses in the results section is always in the neighborhood of 130 participants, and the number of participants in each group is always similar to the one below.)

Table 1: Unadjusted means OVS production ANCOVA (Immediate Posttest)

Group	Mean	Std. Deviation	N
1 [-EI][-TE SI]	53.41886	28.033523	35
2 [-EI][+TE SI]	55.76667	32.284238	30
3 [+EI][-TE SI]	81.20292	23.351950	24
4 [+EI][+TE SI]	80.62542	29.409894	24
5 (Control)	42.58684	30.566742	19
Total	62.39159	32.013652	132

Before adjustment for prior knowledge, both of the groups who received EI, Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI], emerged as the highest scoring groups with 81.20 and 80.62 logits respectively. With considerably lower means, Groups 2 [-

EI][+TE SI] and 1 [-EI][-TE SI] followed, in that order, with means of 55.76 and 53.41 respectively. Finally, Group 5 scored an average of 42.58, the lowest mean of all five. Figure 6 illustrates the information graphically.

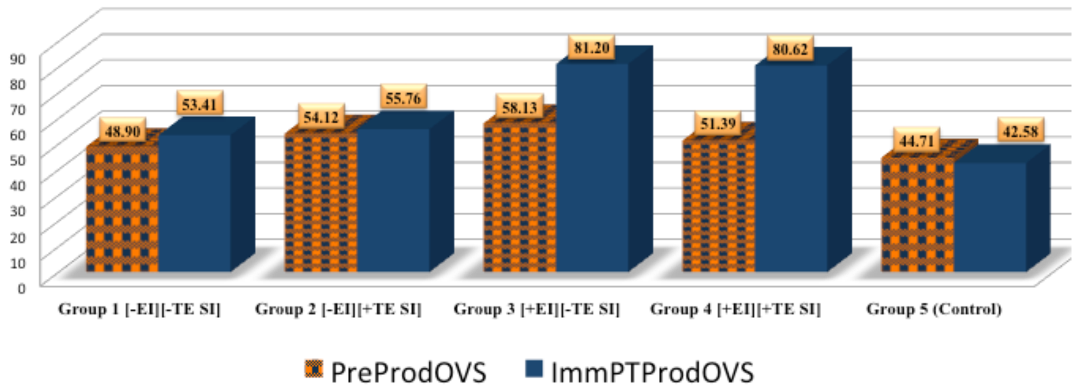


Figure 6: Mean ability scores for OVS production (Pretest and Immediate Posttest)

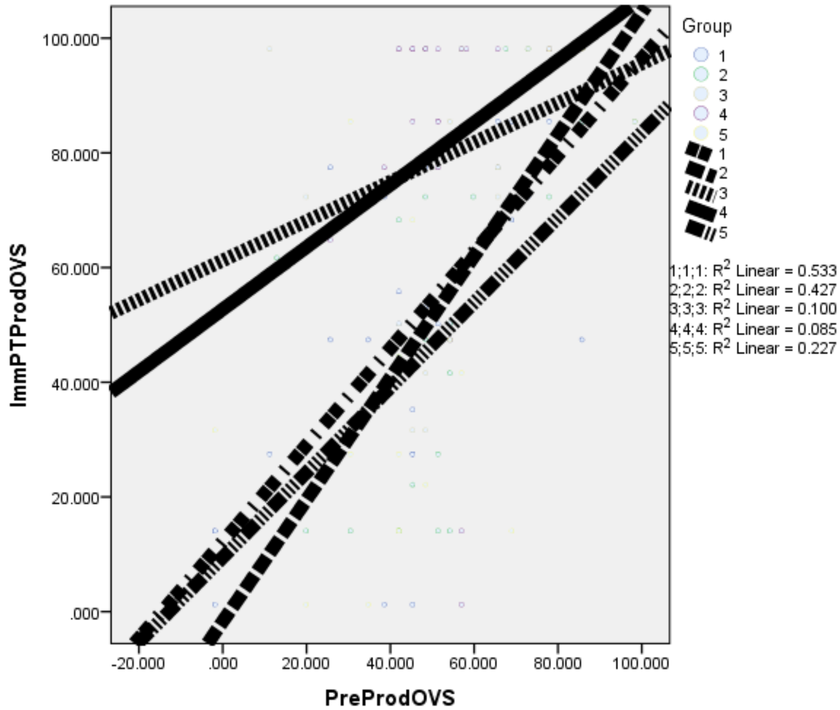
A glance at the graph quickly reveals that, with the exception of the control group, all groups scored higher after treatment than before it. In addition, the graph makes it immediately obvious that the increase in scores was far steeper for [+EI] groups, i.e., Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI]. Figure 6 also reveals that learner knowledge at pretest was slightly different across groups, ranging from 44.71 logits in the control group, to 58.13 in Group 3 [+EI][-TE SI].

Levene’s test of equality of error variances showed that the assumption of homogeneity of variance was met ($F(4,127)=1.2$; p=.314).

Table 2: ANCOVA Analysis for Group effect of unadjusted means of OVS production (Immediate Posttest)

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Hypothesis	10949.418	1	10949.418	5.143	.082	.548
	Error	9020.459	4.237	2129.062[a]			
Group	Hypothesis	9213.641	4	2303.410	3.887	.005	.113
	Error	72293.817	122	592.572[b]			
PreProdOVS	Hypothesis	22691.097	1	22691.097	38.293	.000	.239
	Error	72293.817	122	592.572[b]			
Group * PreProdOVS	Hypothesis	3296.222	4	824.055	1.391	.241	.044
	Error	72293.817	122	592.572[b]			

The ANCOVA reveals that the scores of the groups were significantly different on the immediate posttest, F (4, 122) = 3.88; $p < .01$. In addition, Table 2 shows that there is a significant relationship between the covariate (pretest scores) and the dependent variable, F (1, 122) = 38.29; $p < .01$. Effect sizes for both group and prior knowledge were small according to Cohen's (1998) criteria, .11 and .24 respectively. Finally, the analysis showed no significant interaction between pretest score and treatment group, which suggested that all groups were comparably affected by prior knowledge and that the outcomes of the ANCOVA were valid. In spite of the lack of significance, however, and upon plotting the interaction (Figure 7), it becomes clear that prior knowledge did not affect all groups in exactly the same way. While direction of the slopes for all groups exhibits a similar trend (i.e., an upward one) after treatment, regardless of prior knowledge, the regression lines are far from parallel. This would seem to suggest that prior knowledge did boost post-treatment results somewhat differently for participants depending on how familiar they were with OVS in the first place, albeit not to a significant extent.



Group 1 [-EI][-TE SI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 7: Regression slopes for OVS production (Immediate Posttest)

Once the influence of prior knowledge was factored out, the means for each of the groups changed only slightly. As Table 3 shows, Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI] remained at the top after adjustment, although the means were lower and it was Group 4 [+EI][+TE SI] that scored highest (80.78 logits) of all, with Group 3 [+EI][-TE SI] (78.96 logits) in second position. Thus, the superiority in performance exhibited by [+EI] groups in the non-adjusted means obtained after factoring out previous knowledge as well, suggesting that the difference in scores between these groups and the rest originates in the different treatments received, and not in the pre-treatment differences. Group 1 [-EI][-TE SI] ranked third highest with a mean score of 55.07, closely followed by Group 2 [-EI][+TE SI] with 53.10 logits. Finally, the mean for the control group (Group 5) was slightly higher after adjustment

(47.69), although it still remained the lowest one in the five groups.

Table 3: Adjusted means for OVS production ANCOVA (Immediate Posttest)

Group	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TESI]	55.707[a]	4.141	47.509	63.905
2 [-EI][+TESI]	53.107[a]	4.481	44.237	61.977
3 [+EI][-TESI]	78.966[a]	5.201	68.670	89.261
4 [+EI][+TESI]	80.748[a]	4.969	70.911	90.586
5 (Control)	47.699[a]	5.937	35.946	59.451

The analysis of the adjusted means (Table 4) revealed a main effect for Group, $F\ (4, 122) = 9.19$; $p < .001$.

Table 4: Analysis for Group effect of adjusted means of OVS production ANCOVA (Immediate Posttest)

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Contrast	21800.313	4	5450.078	9.197	.000	.232
Error	72293.817	122	592.572			

A Bonferroni correction (Table 5) revealed that this main effect was due to the differences between the [–EI] Groups (Groups 1 [-EI][-TE SI] and 2 [-EI][+TE SI]), and [+EI] Groups (Groups 3 [+EI][TE SI] and 4 [+EI][TE SI]), and that there was no difference between the groups within the [–EI] and [+EI] conditions. In view of the lack of statistical difference between Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI], the difference between Group 1 [-EI][-TE SI] and Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI] would appear to suggest that the improvement in performance is accounted for by the presence of [+EI]. This is further confirmed by the significant difference recorded between Group 2 [-EI][+TE SI] and Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI].

Table 5: Pairwise comparison of OVS production ANCOVA (Immediate Posttest)

(I) Group	**(J) Group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval for Difference**	
					Lower Bound	Upper Bound
1 [-EI][-TESI]	2[-EI][+TE SI]	2.600	6.101	1.000	-14.843	20.044
	3[+EI][-TE SI]	-23.259*	6.648	.007	-42.266	-4.252
	4[+EI][+TE SI]	-25.041*	6.469	.002	-43.535	-6.547
	5(Control)	8.008	7.238	1.000	-12.686	28.703
2 [-EI][+TESI]	1[-EI][-TE SI]	-2.600	6.101	1.000	-20.044	14.843
	3[+EI][-TE SI]	-25.859*	6.865	.003	-45.485	-6.232
	4[+EI][+TE SI]	-27.641*	6.691	.001	-46.772	-8.511
	5(Control)	5.408	7.438	1.000	-15.857	26.673
3 [+EI][-TESI]	1[-EI][-TE SI]	23.259*	6.648	.007	4.252	42.266
	2[-EI][+TE SI]	25.859*	6.865	.003	6.232	45.485
	4[+EI][+TE SI]	-1.782	7.193	1.000	-22.348	18.783
	5(Control)	31.267*	7.893	.001	8.702	53.832
4[+EI][+TESI]	1[-EI][-TE SI]	25.041*	6.469	.002	6.547	43.535
	2[-EI][+TE SI]	27.641*	6.691	.001	8.511	46.772
	3[+EI][-TE SI]	1.782	7.193	1.000	-18.783	22.348
	5(Control)	33.050*	7.742	.000	10.915	55.184
5 (Control)	1[-EI][-TE SI]	-8.008	7.238	1.000	-28.703	12.686
	2[-EI][+TE SI]	-5.408	7.438	1.000	-26.673	15.857
	3[+EI][-TE SI]	-31.267*	7.893	.001	-53.832	-8.702
	4[+EI][+TE SI]	-33.050*	7.742	.000	-55.184	-10.915

The fact that Group 4 [+EI][+TE SI] performed statistically better than Group 2 [-EI][+TE SI], even though both groups underwent [+TE SI] treatments suggests that it was EI that made the critical difference. Lack of difference between Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI] suggests that given the presence of EI, TE did not offer additional benefits (or at least that such benefits were not perceptible in a short treatment such as the one implemented in the present study). In the same way, the lack of statistical difference in immediate posttest production between Group 5 (Control) and Groups 1 [-EI][-TE SI] and 2 [-EI][+TE SI] suggests that in the absence of EI, receptive practice, whether of TE nature or not, cannot be expected to result in substantial gains.

6.2. OVS Production – Delayed Posttest

The present section will examine whether any of the differences recorded in the immediate posttest were still present three weeks after treatment. For this purpose, scores from the delayed posttest were submitted to a One-Way ANCOVA with pretest scores as a covariate. Table 6 below displays the unadjusted means, where we see a trend similar to the one observed in the previous analysis, i.e. [+EI] groups, (Groups 3 and 4), appear at the top of the ranking with means of around 79 logits of ability, followed by both Groups 1 and 2 ([-EI] groups) at 58.40 and 69.11 logits respectively, and the control group in last position with 38.50 logits.

Table 6: Unadjusted means OVS production ANCOVA (Delayed Posttest)

Group	**Mean**	**Std. Deviation**	**N**
1 [-EI][-TESI]	58.40943	31.716567	35
2 [-EI][+TESI]	69.11867	26.151642	30
3 [+EI][-TESI]	79.71958	22.891119	24
4 [+EI][+TESI]	79.55292	20.266423	24
5 (Control)	39.01842	38.503441	19
Total	65.77106	31.170398	132

Worthy of note in Table 6 is the mean score of 69.12 logits obtained by Group 2 [-EI][+TE SI]. Based on the results from the ANCOVA on the immediate posttest, Group 2 [-EI][+TE SI] obtained 55.76 logits immediately after treatment (which turned into 53.10 after adjustment), a score visibly lower than the one obtained in the delayed posttest. Given the score of 69.12 in the delayed posttest, it would appear that the three weeks that elapsed between immediate and delayed posttest not only did not have any adverse effect on the gains obtained from the treatment (as one might expect) but it actually benefited participants in this group. This, of course, is by all accounts counterintuitive although perhaps not entirely unreasonable if we consider the TE nature of our testing instruments and the treatment received by Group 2 [-EI][+TE SI]. It is seems reasonable that in the absence of any other source from where to derive the rules governing OVS sentences, participants in Group 2 [-EI][+TE SI] should have consciously or unconsciously engaged in an elimination process (conscious or unconsciously) as the only way to associate a form to its meaning and perform successfully. Devoting all attentional resources to this task during the TE treatment might have resulted in at least some participants in Groups 2 [-EI][+TE SI] becoming better test-takers within the TE format that characterized all of our instruments. We argue that this ability to perform better within the TE format increased gradually with the number of exposures to TE practice, despite the time elapsed between them.

Results from Levene's Test suggest that technically, the variation of scores in across groups in the delayed posttest was equal, $F(4, 127)= 2.33$; = .059. The obtained p value, however, is marginally significant. This indicates that there may be a violation of homogeneity of variance, and therefore that our data may not fit the model used for analysis perfectly, which calls for a degree of caution when interpreting the results obtained from it, as the likelihood for Type I and II errors is greater.

In all, as the lines in illustrate, it appears that time resulted in loss of some of the gains obtained from the treatments.

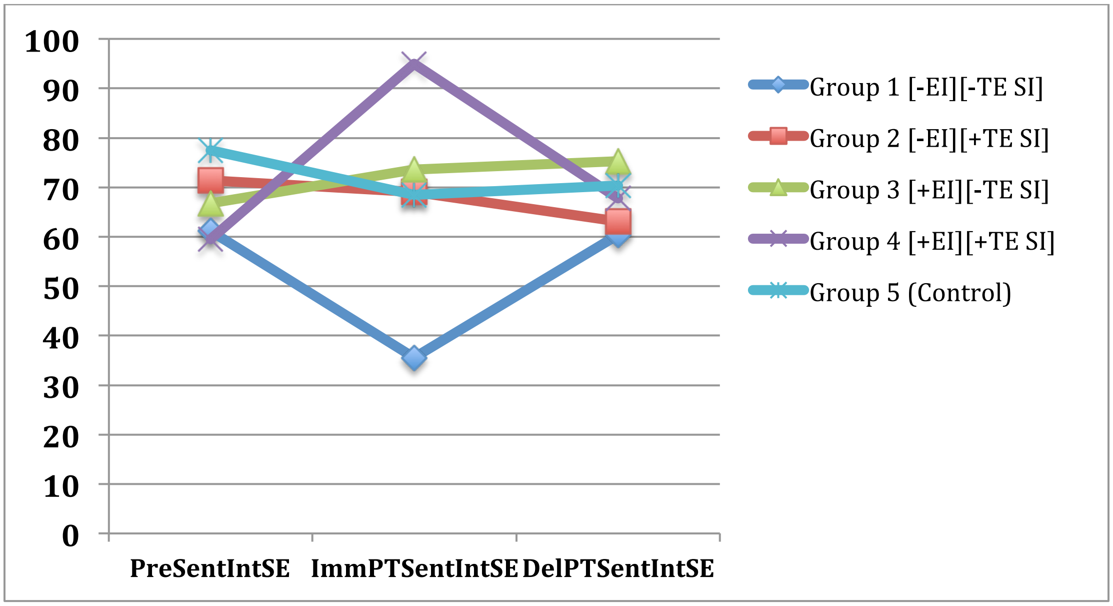


Figure 8: Development over time in OVS production.

That being said, it is also obvious from Figures 9 and 10 that scores did not regress to what they were at the pretreatment stage. In addition, non-adjusted means seem to suggest that treatments including EI generate greater gains than treatments that do not and that these are still perceptible three weeks after treatment.

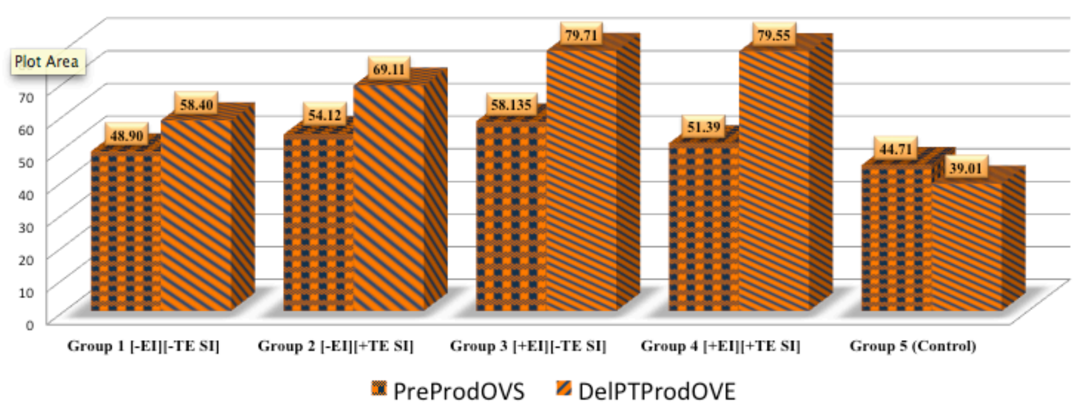


Figure 9: Mean ability scores for production OVS (Pretest and Delayed Posttest)

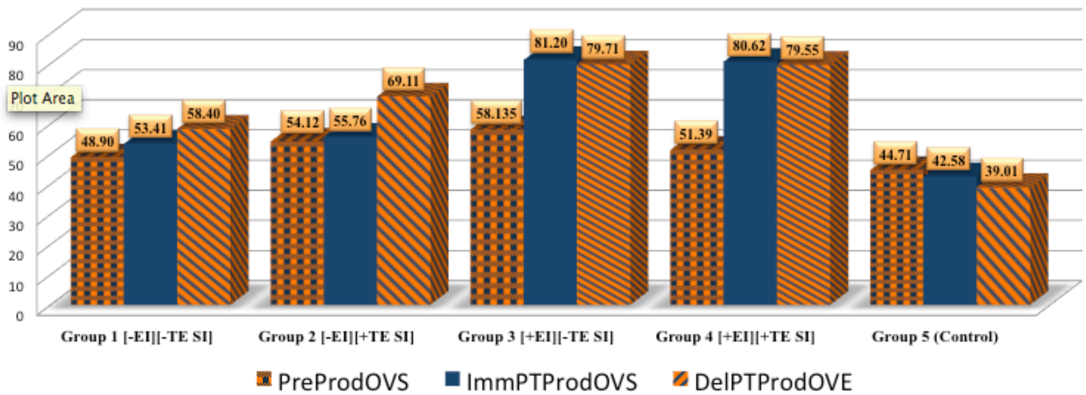


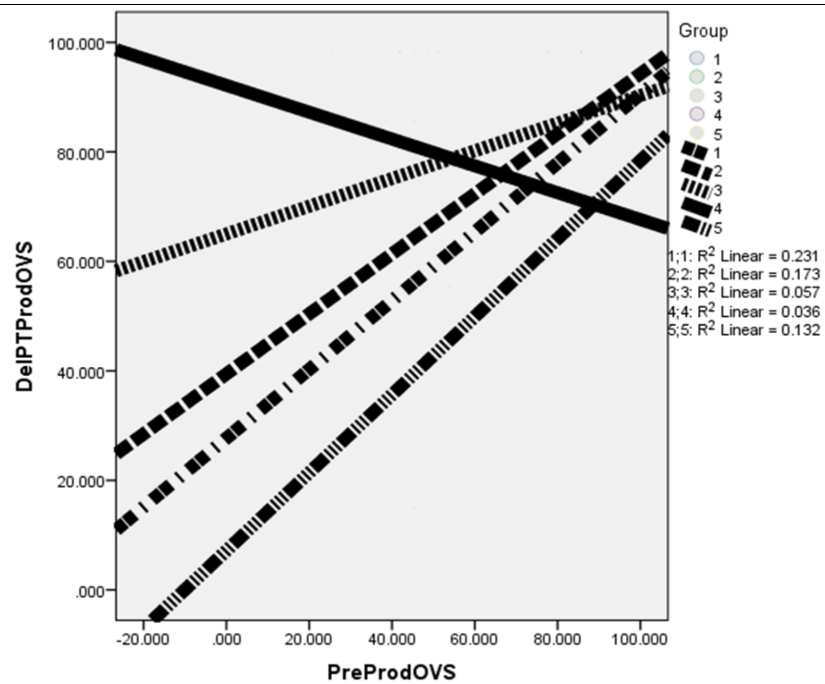
Figure 10: Mean ability scores for Production OVS (Pretest Immediate Posttest and Delayed Posttest)

Indeed, the analysis of the non-adjusted means (Table 7) revealed a main effect for group, F (1, 122) = 4.00; $p < .001$, which would appear to indicate that the differences between groups above mentioned were significant. Given that the present means are not adjusted for previous knowledge, however, it is not possible at this point to conclude that differences in gains originated in the different treatments received. This is especially so in view of the fact that in addition to the main effect for Group, a main effect for prior knowledge was found, F (1,122)= 9.251; $p < .01$. Such an effect suggested that prior knowledge had a significant influence over the scores obtained in the delayed posttest, which called for analysis of the adjusted means. Finally, no interaction between prior knowledge and Group was detected, F (1,122) = 1.555, $p = .191$, which suggests that the effect of time on the gains obtained did not differ between groups regardless of participants' prior knowledge.

Table 7: ANCOVA analysis of unadjusted means of OVS production (Delayed Posttest)

Source		**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Intercept	Hypothesis	32693.880	1	32693.880	12.487	.022
	Error	11074.073	4.230	2618.222[a]		
Group	Hypothesis	11340.173	4	2835.043	4.007	.004
	Error	86307.016	122	707.435[b]		
PreProdOVS	Hypothesis	6544.647	1	6544.647	9.251	.003
	Error	86307.016	122	707.435[b]		
Group * PreProdOVS	Hypothesis	4399.469	4	1099.867	1.555	.191
	Error	86307.016	122	707.435[b]		

Finally, while no interaction emerged between Group and prior knowledge with the unadjusted means, Figure 11 reveals that a clear difference in the direction of the slopes at the delayed posttest stage. It would appear that in Group 4[+EI][+TE SI] the effect of time resulted in a different directionality of scores as compared to the rest of the groups (as shown by the upward trend displayed by all other groups). This would seem to suggest that participants receiving a [+EI] and [+TE SI] treatment are the most sensitive to time. In addition, the slopes for groups 1[-EI][-TE SI], 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 5 (Control) are different, suggesting the different benefits obtained by participants after treatment as a function of their prior knowledge is still perceptible at the posttest stage.



Group 1 [-EI][-TESI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 11: Regression slopes for OVS production (Delayed Posttest)

With the exception of Group 4, in which case absence of treatment and the three-week time period resulted in prior knowledge negatively affecting scores, the extent to which performance was dependent on prior knowledge was comparable from immediate to delayed posttest.

Once the effect of prior knowledge was factored out, the means for the delayed posttest were as shown in Table 8.

Table 8: Adjusted means for OVS production ANCOVA (Delayed Posttest)

Group	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TESI]	60.115[a]	4.525	51.158	69.072
2 [-EI][+TESI]	67.745[a]	4.896	58.053	77.437
3 [+EI][-TESI]	78.071[a]	5.683	66.822	89.320
4 [+EI][+TESI]	79.498[a]	5.430	68.749	90.246
5 (Control)	43.935[a]	6.487	31.094	56.776

As it becomes immediately obvious, Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI], with 78.07 and 79.49 logits respectively, remained almost unchanged after the adjustment for prior knowledge, suggesting that the gains they appeared to maintain were unaffected by their knowledge of OVS structures prior to treatment. Group 2 [-EI][+TE], with a mean score of 67.74, remained in third position but displays a slightly lower index of ability as compared to the raw means of the delayed posttest, suggesting that participants in this group started out with greater knowledge of the structure and therefore their learning curve became less steep once that was factored out. On the contrary, the mean of Group 1 [-EI][-TE SI] became higher after adjustment, going from 58.40 to 60.11, suggesting that participants in Group 1 saw their gains in the delayed-posttest diminished by their inexperience with OVS. Once prior knowledge had been factored out, the same phenomenon that emerged with Group 2 [-EI][+TE SI] when reporting on the non-adjusted means, surfaced for Group 1 [-EI][-TE SI] also. Namely, it would appear that three weeks after treatment participants in Group 1 [-EI][-TE SI] performed better than immediately after receiving the treatment, as the mean in the delayed posttest was higher by about 5 logits. As we mentioned before, this may owe to the fact that TE practice (in this case not in the treatment but in the instruments) generated some knowledge by itself, as it provided an environment conducive to rule deduction. Such gains would naturally become more easily noticeable in Groups 1[-EI][-TE SI] and 2[-EI][+TE SI], where the treatments themselves did not result in large improvements, than in Groups 3[+EI][-TE SI] and 4[+EI][+TE SI], where the effects of EI may have made the benefits offered by the instruments redundant. One might argue that advancing TE as the trigger for improvement in the posttest would mean that Group 5 should have exhibited an improvement in the posttest, which was not the case (the adjusted mean in the immediate posttest was 47.69, while the adjusted posttest mean was 43.93). The dip in performance, however, may be explained in this case by the fact that in the absence of any practice at all related to OVS the TE nature of the instruments did not deliver. In addition, it is conceivable that participants in Group 5 may have experienced a loss of interest over time in performing well in the experiment.

The difference perceptible in the adjusted means was significant for Group, $F(4, 122) = 6.00$; $p > .001$, as shown in Table 9. Given that differences in prior knowledge could not have affected these scores, this main effect would suggest that the different treatments received by Groups 1-5 resulted in gains that were significantly different from one another.

Table 9: ANCOVA analysis of adjusted means of OVS production (Delayed Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	17001.632	4	4250.408	6.008	.000	.165
Error	86307.016	122	707.435			

A subsequent Bonferroni post-hoc analysis (Table 10) revealed that these differences were situated between Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4[+EI][+TE SI] on the one hand, and Group 5 (Control) on the other.

Table 10: Pairwise comparison of OVS production ANCOVA (Delayed Posttest)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TE SI]	-7.630	6.666	1.000	-26.689	11.429
	3[+EI][-TE SI]	-17.956	7.264	.148	-38.723	2.812
	4[+EI][+TESI]	-19.383	7.068	.070	-39.590	.824
	5(Control)	16.180	7.909	.429	-6.432	38.791
2 [-EI][+TESI]	1[-EI][-TE SI]	7.630	6.666	1.000	-11.429	26.689
	3[+EI][-TE SI]	-10.326	7.501	1.000	-31.770	11.119
	4[+EI][+TESI]	-11.752	7.311	1.000	-32.655	9.150
	5(Control)	23.810*	8.127	.040	.575	47.045
3 [+EI][-TESI]	1[-EI][-TE SI]	17.956	7.264	.148	-2.812	38.723
	2[-EI][+TE SI]	10.326	7.501	1.000	-11.119	31.770
	4[+EI][+TESI]	-1.427	7.860	1.000	-23.897	21.044
	5(Control)	34.136*	8.624	.001	9.480	58.791
4 [+EI][+TESI]	1[-EI][-TE SI]	19.383	7.068	.070	-.824	39.590
	2[-EI][+TE SI]	11.752	7.311	1.000	-9.150	32.655
	3[+EI][-TE SI]	1.427	7.860	1.000	-21.044	23.897
	5(Control)	35.562*	8.459	.001	11.377	59.747
5 (Control)	1[-EI][-TE SI]	-16.180	7.909	.429	-38.791	6.432
	2[-EI][+TE SI]	-23.810*	8.127	.040	-47.045	-.575
	3[+EI][-TE SI]	-34.136*	8.624	.001	-58.791	-9.480
	4[+EI][+TESI]	-35.562*	8.459	.001	-59.747	-11.377

This would suggest that in general, after three weeks, gains derived from the treatment had waned to the point that the significant superiority of [+EI] groups, Groups 3 and 4, over [-EI] groups, Groups 1 and 2, recorded immediately after treatment was no longer perceptible. Thus, production gains at the delayed posttest stage were comparable for Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4[+EI][+TE SI]. That said, the significant differences revealed by the delayed posttest pair-wise comparison do suggest that gains were maintained at least to some extent, as the higher scores exhibited by all experimental groups with the exception of Group 1[-EI][-TE SI] were still significant as compared to Group 5 (Control).

6.3. OVS Picture Matching – Immediate Posttest

In the present section we look at the OVS scores for the picture- matching task. As we did with OVS production data, first we submit pretest and immediate posttest scores to an ANCOVA analysis to determine the effects of the different treatments. In order to establish the durability of the gains ensued, a second ANCOVA was conducted with the delayed posttest scores as a dependent variable.

OVS picture matching data for the pretest presented a very pronounced positive skewness[24] resulting from participants' systematic interpretation of OVS sentences as though they were canonically-ordered SVO utterances. In addition, plotting of the regression slopes suggested there may be a significant interaction between immediate posttest scores and prior knowledge, which would constitute a violation of homoscedasticity. The combination of both of those violations has been

[24] Prior to treatment, learners consistently applied default processing strategies based on the canonical SVO English word order. As a result the distribution of pretest for interpretation tasks was bottom-heavy and extremely non-normal. Logarithmic transformation of the data was undertaken in an attempt to be able to submit scores to the more powerful parametric analyses, but the resulting distributions remained both visually and numerically non-normal.

reported in literature to result in either an over- or underestimation of the nominal level of significance (Olejnik and Algina, 1984; Seaman et al. 1985). For those reasons, and despite the obvious cost in terms of power, it was deemed appropriate to adopt a non-parametric ANCOVA model to analyze the present data. Traditionally, in cases of non-normalcy, Rank Analysis of Covariance (McSweeney and Porter, 1971; Connover and Inman, 1982) is recommended as the most straightforward and robust option (Huitema, 1980; Olejnik and Algina, 1984; Seaman et al. 1985). This procedure involves the transformation of dependent and independent variables to ranks and then using the resulting scores in a conventional parametric ANCOVA analysis (Seaman et al., 1985; Ho, 2006; Huitema, 2011). Results from the analysis using that procedure are presented in Table 11. The total number of participants for this task was 130, each group being populated by a similar number of participants as in the production task. The non-adjusted means reveal a pattern somewhat parallel to the one found in production in that [+EI] groups emerge as the highest scoring ones. In the case of picture matching, it is Group 3 [+EI][-TE SI] with a mean of 84.31 logits that ranks exhibits the highest scores, followed by Group 4 [+EI][+TE SI] with a score of 83.40. Group 2 [-EI][+TE SI], with 76.66 logits, remains in third position. Its mean score, however, is almost as high as those for the [+EI] groups, and much higher than that of the fourth-ranking Group 1 [-EI][-TE SI], which obtained a mean score of 51.12 logits. The lowest score was obtained by the control group, whose mean ability measure was 32.53.

Table 11: Unadjusted means OVS picture matching
Rank ANCOVA (Immediate Posttest)

Group	Mean	Std. Deviation	N
1 [-EI][-TESI]	51.12857	37.281982	35
2 [-EI][+TESI]	76.66667	27.471092	30
3 [+EI][-TESI]	84.31250	27.616084	24
4 [+EI][+TESI]	83.40476	33.003643	21
5 (Control)	32.52500	32.721502	20
Total	65.50000	37.016129	130

Graphic representation of the immediate posttest means and the pretest means (Figure 12) reveals a pattern not dissimilar to the one found for OVS production, with the exception of the higher score for Group 2 [-EI][+TE]. Knowledge at pretest across groups appeared to be rather high for all groups, as well as rather uniform, ranging from 61.40 in Group 2 [-EI][+TE SI] to 70.25 in Group [+EI][-TE SI]. Interestingly, whereas Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] saw an increase in their ability after treatment, Groups 1 [-EI][-TE SI] and 5 (Control) exhibit lower ability measures. The reliability index of this task was relatively high (.76), and no misfitting participants were found, which would suggest that the decrease in ability did not owe to a loss of engagement on their part. In view of the treatments Groups 1 [-EI][-TE SI] and 5 (Control) received, the decrease in ability begs the question, then, whether certain treatments that do not result in learning, may actually have a negative effect.

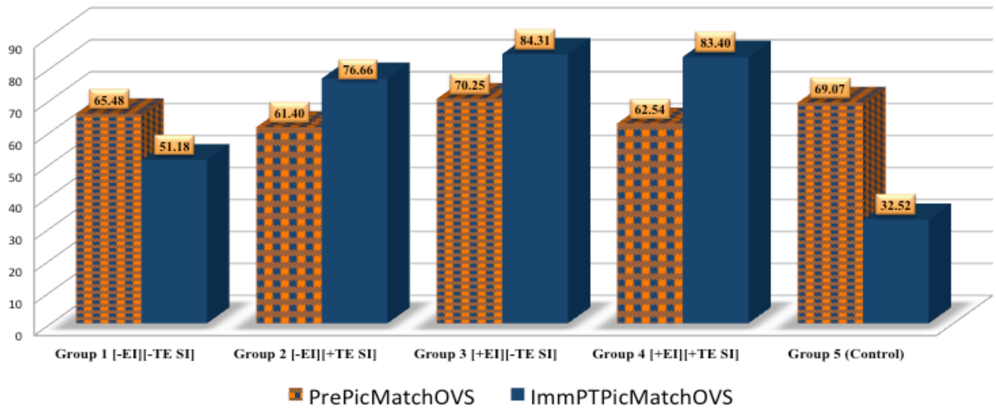


Figure 12: Mean ability scores for OVS picture matching (Pretest and Immediate Posttest)

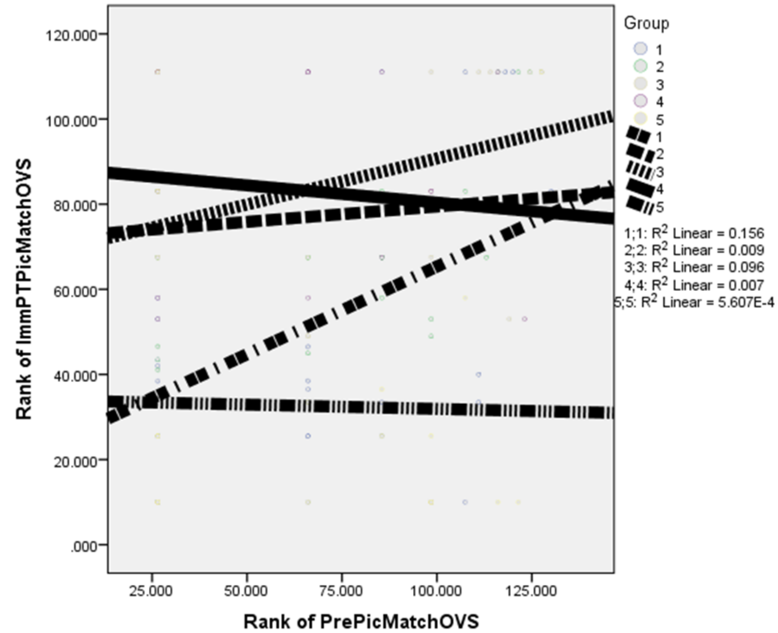
Levene's Test was not significant, $F(4, 125) = .778$; $p = .541$, which indicates the variance of scores was evenly distributed across groups and allows us to safely interpret the results form the inferential analysis.

As Table 12 shows, comparison of non-adjusted means yielded a main effect for group, $F\ (4, 120) = 4.480$; $p < .01$, suggesting group gains differed as a result of the different treatments. In addition, the analysis revealed that the covariate's effect on the dependent variable was not significant, $F\ (1, 120) = 2.28$, $p = .133$, which indicated that prior knowledge did not influence scores obtained in the immediate posttest. Finally, although numerically the interaction between Group and Pre-test Score is clearly not significant $F\ (4, 120) = 1.34$; $p = .257$, graphic representation of the regression slopes (Figure 12) reveals a less straightforward picture.

Table 12: Results of analysis of unadjusted means for OVS picture matching Rank ANCOVA (Immediate Posttest)

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	57358.796[a]	9	6373.200	6.405	.000	.325
Intercept	93899.282	1	93899.282	94.374	.000	.440
Group	17831.597	4	4457.899	4.480	.002	.130
RPrePicM	2273.797	1	2273.797	2.285	.133	.019
Group * RPrePicM	5360.323	4	1340.081	1.347	.257	.043
Error	119396.204	120	994.968			
Total	734487.500	130				
Corrected Total	176755.000	129				

Although it is clear from Figure 13 that the regression lines for immediate posttest scores plotted against pretest scores is not parallel for the five groups, all slopes, with the exception of those for Groups 4 [+EI][+TE SI] and 5 (Control) are similar in their orientation, indicating that the prior knowledge affected scores for participants in those groups positively.



Group 1 [-EI][-TE SI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 13: Regression slopes for OVS picture matching (Pretest and Immediate Posttest)

Groups 4 [+EI][+TE SI] and 5 (Control), however, display a virtually flat line, with a light downward trend, which would seemingly indicate prior knowledge had no effect, or even a negative one for at least some participants in those conditions. In the case of Group 4 [+EI][+TE SI], it is possible that the information in the treatment rendered any prior knowledge largely unnecessary. In addition, despite common upward orientation, it is clear that prior knowledge played a more positive role for Groups 1 [-EI][-TE SI] and 3[+EI][-TE SI] than for Group 2 [-EI][+TE SI].

Table 13: Adjusted means for picture matching OVS Rank of ANCOVA (Immediate Posttest)

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 [-EI][-TESI]	51.134[a]	5.332	40.578	61.691
2 [-EI][+TESI]	76.963[a]	5.797	65.485	88.440
3 [+EI][-TESI]	83.286[a]	6.487	70.442	96.130
4 [+EI][+TESI]	83.165[a]	6.912	69.480	96.850
5 (Control)	32.597[a]	7.085	18.569	46.624

After the effects of previous knowledge were controlled for, means changed very slightly and the overall ordering of groups remained very much the same; see Table 13, where Groups 3 and 4[+EI] are the highest scoring with means of 83 logits, followed by Group 2 [-EI][+TE SI] at 76.96, and finally Groups 1[-EI][-TE SI] and 5 (Control), with 51.13 and 32.58 logits respectively.

Table 14: Results for analysis of adjusted means for picture matching OVS Rank ANCOVA (Immediate Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	46545.255	4	11636.314	11.695	.000	.280
Error	119396.204	120	994.968			

The main effect for group still emerged after pretest effects had been factored out, as shown by the significant value, F (4, 120) = 11.69, $p < .01$, in Table 14.

Table 15: Pairwise comparison of OVS picture matching Rank ANCOVA (Immediate Posttest)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TE SI]	-25.828*	7.876	.014	-48.353	-3.304
	3[+EI][-TE SI]	-32.152*	8.397	.002	-56.166	-8.137
	4[+EI][+TE SI]	-32.031*	8.729	.004	-56.996	-7.066
	5(Control)	18.538	8.867	.387	-6.821	43.896
2[-EI][+TESI]	1[-EI][-TE SI]	25.828*	7.876	.014	3.304	48.353
	3[+EI][-TE SI]	-6.323	8.700	1.000	-31.203	18.557
	4[+EI][+TE SI]	-6.203	9.021	1.000	-32.002	19.596
	5(Control)	44.366*	9.154	.000	18.186	70.546
3[+EI][-TESI]	1[-EI][-TE SI]	32.152*	8.397	.002	8.137	56.166
	2[-EI][+TE SI]	6.323	8.700	1.000	-18.557	31.203
	4[+EI][+TE SI]	.121	9.479	1.000	-26.989	27.230
	5(Control)	50.689*	9.606	.000	23.217	78.162
4[+EI][+TESI]	1[-EI][-TE SI]	32.031*	8.729	.004	7.066	56.996
	2[-EI][+TE SI]	6.203	9.021	1.000	-19.596	32.002
	3[+EI][-TE SI]	-.121	9.479	1.000	-27.230	26.989
	5(Control)	50.569*	9.898	.000	22.261	78.876
5 (Control)	1[-EI][-TE SI]	-18.538	8.867	.387	-43.896	6.821
	2[-EI][+TE SI]	-44.366*	9.154	.000	-70.546	-18.186
	3[+EI][-TE SI]	-50.689*	9.606	.000	-78.162	-23.217
	4[+EI][+TE SI]	-50.569*	9.898	.000	-78.876	-22.261

Given the raw and adjusted means, unsurprisingly, a post-hoc test (Table 15) located the difference in means between Group 1 [-EI][-TE SI] and Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] as well as between Group 5 (Control) and Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI]. These differences are suggestive of the superiority of treatments that include EI or TE. Lack of difference between Groups 2 [-EI] [+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] once again suggest that combining EI and TE does not appear to offer any additional benefit over exposing learners to one of them only.

6.4. OVS Picture Matching – Delayed Posttest

As Table 16 shows, the ordering of the groups in the delayed posttest of OVS picture matching, according to the unadjusted means, is the same as for the immediate posttest. Thus, the [+EI] groups score highest: Group 4 [+EI][+TE SI] with 87.97 logits and Group 3 [+EI][-TE SI] with 78.91 logits. Group 2 [-EI][+TE SI] comes in third with 73.13 logits and finally, in the last two positions we find Groups 1 [-EI][-TE SI] and 5 (Control), in that order, with visibly lower means, (i.e., 50.52 and 40.55 logits respectively).

Table 16: Unadjusted means for picture matching OVS Rank ANCOVA (Delayed Posttest)

Descriptive Statistics			
Dependent Variable: Rank of DelPTPicMatchOVS			
Group	Mean	Std. Deviation	N
1 [-EI][-TE SI]	50.52857	36.980906	35
2 [-EI][+TE SI]	73.13333	34.824915	30
3 [+EI][-TE SI]	78.91667	33.753797	24
4 [+EI][+TE SI]	87.97619	28.181322	21
5 (Control)	40.55000	29.879539	20
Total	65.50000	37.114061	130

Figures 14 and 16 make it evident that despite the three weeks elapsed, gains obtained for picture matching were preserved for three of the four experimental groups.

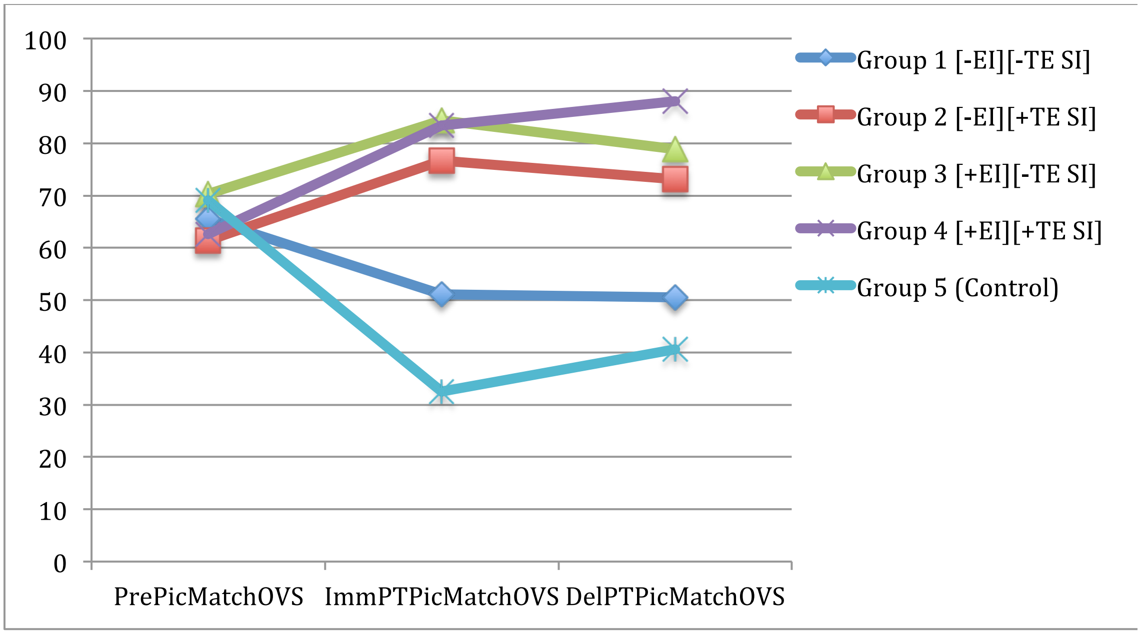


Figure 14: Development over time for OVS picture matching.

In fact, Figure 15 reveals that delayed test scores for Group 4 [+EI][+TE SI] are descriptively even a bit higher immediate than scores for the immediate posttest (83.47 as compared to 87.96).

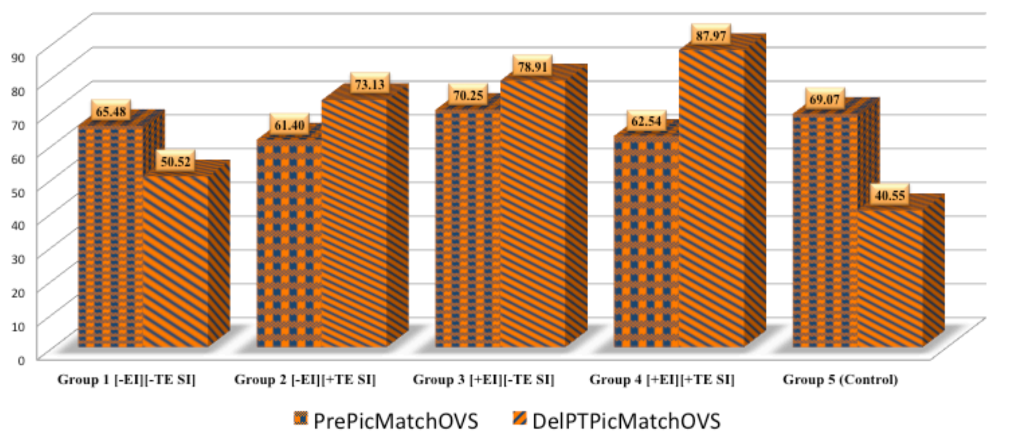


Figure 15: Mean ability scores for OVS picture matching (Pretest and Delayed Posttest)

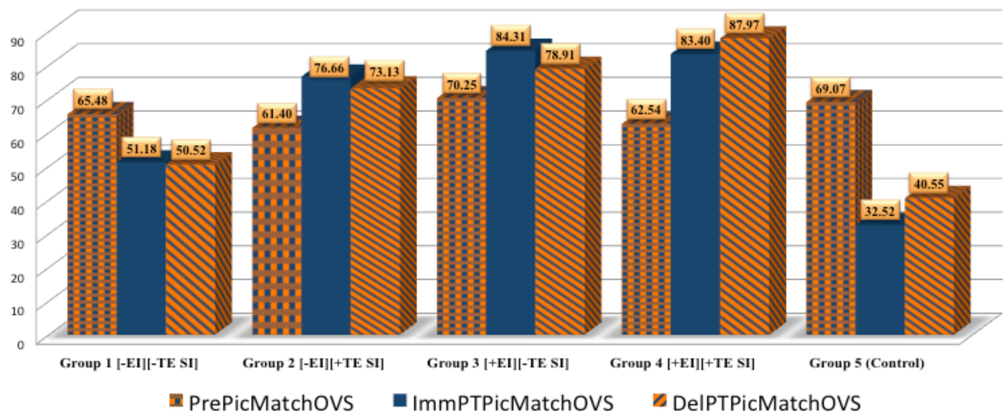


Figure 16: Mean ability scores for OVS picture matching (Pretest and Immediate and Delayed Posttest)

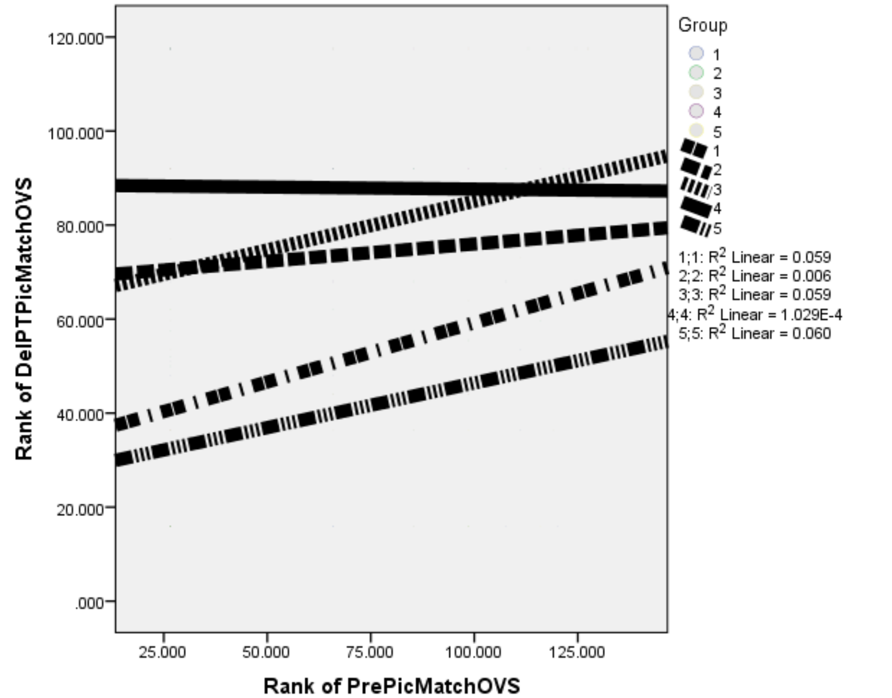
Levene's test was not significant, $F(4, 125) = 1.7$; $p = .154$; the groups have approximately the same variance with respect to the dependent variable (i.e., the requirement of homogeneity of variance has been met).

Results from the analyses of unadjusted means (Table 17) reveal an effect for Group, $F\ (4,120) = 3.05$; $p < .05$, suggesting that even after three weeks but before factoring out prior knowledge there appear to be differences between groups. As was the case in the comparison between pretest and immediate posttest scores for picture matching, no significant effect was found for prior knowledge, and no interaction appeared to be present between the grouping variable and knowledge of the structure prior to treatment.

Table 17: Rank ANCOVA analysis of unadjusted means for OVS (Delayed Posttest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	42498.413[a]	9	4722.046	4.191	.000	.239
Intercept	90713.961	1	90713.961	80.519	.000	.402
Group	13784.754	4	3446.189	3.059	.019	.093
RPrePicM	3226.192	1	3226.192	2.864	.093	.023
Group * RPrePicM	1375.354	4	343.839	.305	.874	.010
Error	135193.087	120	1126.609			
Total	735424.000	130				
Corrected Total	177691.500	129				

Once again, looking at the slopes of the different groups (Figure 17) we see that there is a common upward trend for all the groups, this time with the exception of Group 4 [+EI] [+TE SI]. This suggests that prior knowledge had a similar effect for the retention of gains for all groups, except for Group 4 [+EI] [+TE SI], in which case, again, a flat line indicates prior knowledge had no effect on retention. The graph makes it clear that prior knowledge also interacted differently with the [-EI][+TE SI] treatment received by Group 2 [-EI][+TE SI], as the upward slope is far less perceptible. This suggests that the TE treatment combined with the previous knowledge of participants in Group 2 [-EI] [+TE SI] over time resulted in more modest gain retention than those of groups whose participants had similar pretest scores but were assigned to a different condition.



Group 1 [-EI][-TE SI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 17: Regression slopes for OVS picture matching (Pretest and Delayed Posttest)

Excepting Groups 1 [-EI][-TE SI] and 4 [+EI][+TE SI], the regression lines for OVS picture matching were comparable at the immediate and delayed testing stages. For Group 1[-EI][-TE SI], prior knowledge in combination with the treatment received resulted in a more pronounced boost in performance at the immediate posttest than at the delayed posttest stage. It seems reasonable to believe that in the absence of further treatment, the effect of the proficiency level of certain learners in Group 1 [-EI][-TE SI] may have subsided as a result of the three weeks that elapsed between testing times. Along the same lines but with opposite effect, Group 4 [+EI][+TE SI], exhibited a flatter slope at the delayed posttest stage. This may indicate that the confusion that certain participants had experienced at the immediate posttest diminished over time, causing the level of prior knowledge that interacted with the treatment to play virtually no role.

Mean scores for the delayed posttest were almost identical after controlling for prior knowledge (Table 18), a fact that in and of itself is telling. Groups remain in the exact same order as reported above, namely, Group 4 [+EI][+TE SI], Group 3 [+EI][-TE SI], Group 2 [-EI][+TE SI], Group1 [-EI][-TE SI] and Group 5 (Control), in that order.

Table 18: Adjusted means for picture matching OVS Rank ANCOVA (Delayed Posttest)

Group	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TESI]	50.532[a]	5.674	39.299	61.765
2 [-EI][+TESI]	73.436[a]	6.168	61.223	85.649
3 [+EI][-TESI]	77.933[a]	6.903	64.266	91.600
4 [+EI][+TESI]	87.951[a]	7.355	73.388	102.513
5 (Control)	39.871[a]	7.539	24.944	54.798

After analyzing the differences in adjusted means, an effect for Group was still found (Table 19), suggesting that at least some of the groups did indeed maintain the gains observed from the treatment, F (4, 120) = 8.18; $p < .001$, regardless of their familiarity with OVS interpretation prior to treatment.

Table 19: Rank ANCOVA analysis of adjusted means for OVS picture matching (Delayed Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	36867.771	4	9216.943	8.181	.000	.214
Error	135193.087	120	1126.609			

Subsequent post-hoc analyses (Table 20) located that difference between Groups 1[-EI][-TE SI] and Groups 3[+EI][-TE SI] and 4[+EI][+TE SI]. Further differences were found between Group 5 (Control) and Groups 2[-EI][+TE SI], 3[+EI][-TE SI] and 4[+EI][+TE SI]. In the Delayed Posttest, then, the difference between Groups 1[-EI][-TE SI] and 2 [-EI][+TE SI] that was present at the immediate posttest stage disappears. This is not surprising, as the mean logit of Group 2 [-EI][+TE SI] was close to that of Group 1[-EI][-TE SI]. The overall trend, then, is the same here as for the immediate posttest in that the pairwise comparison confirms the superiority of [+EI] with respect to the durability of gains.

Table 20: Pairwise comparison of OVS picture matching Rank ANCOVA (Delayed Posttest)

(I) Group	**(J) Group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval for Difference**	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TE SI]	-22.904	8.381	.072	-46.872	1.064
	3[+EI][-TE SI]	-27.401*	8.935	.027	-52.954	-1.847
	4[+EI][+TE SI]	-37.419*	9.289	.001	-63.984	-10.853
	5(Control)	10.661	9.435	1.000	-16.323	37.645
2[-EI][+TESI]	1[-EI][-TE SI]	22.904	8.381	.072	-1.064	46.872
	3[+EI][-TE SI]	-4.497	9.257	1.000	-30.972	21.978
	4[+EI][+TE SI]	-14.515	9.599	1.000	-41.967	12.938
	5(Control)	33.565*	9.741	.008	5.707	61.423
3[+EI][-TESI]	1[-EI][-TE SI]	27.401*	8.935	.027	1.847	52.954
	2[-EI][+TE SI]	4.497	9.257	1.000	-21.978	30.972
	4[+EI][+TE SI]	-10.018	10.087	1.000	-38.865	18.830
	5(Control)	38.062*	10.222	.003	8.829	67.295
4[+EI][+TESI]	1[-EI][-TE SI]	37.419*	9.289	.001	10.853	63.984
	2[-EI][+TE SI]	14.515	9.599	1.000	-12.938	41.967
	3[+EI][-TE SI]	10.018	10.087	1.000	-18.830	38.865
	5(Control)	48.080*	10.533	.000	17.958	78.201
5(Control)	1[-EI][-TE SI]	-10.661	9.435	1.000	-37.645	16.323
	2[-EI][+TE SI]	-33.565*	9.741	.008	-61.423	-5.707
	3[+EI][-TE SI]	-38.062*	10.222	.003	-67.295	-8.829
	4[+EI][+TE SI]	-48.080*	10.533	.000	-78.201	-17.958

Recent publications (Hettmansperger and McKean 2011; Huitema, 2011) warn that fitting non-normal data to Rank ANCOVA can be misleading in a number of respects, suggesting instead that a Wilcoxon-estimation-based Robust General Linear Model ANCOVA be used with data presenting non-normalcy. In view of these developments, the present data were submitted to McKean's Robust ANCOVA. Results concurred with the ones reported above in that a Group effect, was found both for the immediate, F (4, 4) = 20.02; $p < .001$, and delayed posttests F (4, 4) = 8.75; p

< .001, as shown in Tables 21 and 22. At the time of writing this paper, the software to perform these analyses was being finalized. These analyses are a courtesy of Joseph McKean and Bradley Huitema, at the University of Western Michigan, who kindly provided enough results to determine a group effect was present. Note that these analyses are also robust to heteroscedasticity, which was somewhat of a concern after plotting the group slopes graphically in the immediate posttest data.

Table 21: Results of Robust ANCOVA for OVS picture matching (Immediate Posttest)

	DF	**RD**	**MRD**	**F**	**p-value**
Groups	4	800.7379	200.1845	20.02429	0.0001
Homog Slopes	4	122.0108	30.5027	3.05116	0.0196

Table 22: Results of Robust ANCOVA for OVS picture matching (Delayed Posttest)

	DF	**RD**	**MRD**	**F**	**p-value**
Groups	4	446.51304	111.62826	8.756198	0.0001
Homog Slopes	4	64.21171	16.05293	1.259203	0.2899

Perhaps worthy of note is the fact that the p values observed in the Robust ANCOVA are drastically smaller than the ones observed by the Rank ANCOVA, which may indicate that the Rank ANCOVA is a more conservative analysis than the Robust ANCOVA.

6.5. OVS Sentence Interpretation – Immediate Posttest

The present section will present the results from the last OVS task, namely, the sentence interpretation task. As was the case with the OVS picture matching task, data from the sentence interpretation task presented a very pronounced positive skew that did not result in a normal distribution even after LN transformation. Thus, in this section we report on the results from a Rank ANCOVA, an analysis which we complement with McKean's Robust ANCOVA mentioned in the previous section.

Initially, results from the OVS sentence interpretation Rank ANCOVA aligned perfectly with results from both production and picture matching. In Table 23 we see that based on unadjusted means, the ordering of groups in the OVS sentence interpretation immediate posttest is the same we saw in the two previous tasks. In other words, both [+EI] groups emerge as the ones having obtained most gains (Group 3 [+EI][-TE SI] at 82.56 logits and Group 4 [+EI][+TE SI] at 81.38), followed by Group 2 [-EI][+TE SI] with 80.38 and finally by Groups 1 [-EI][-TE SI] and Group 5 (Control), both of which exhibit by far the lowest means (49.54 and 33.16, respectively).

Table 23: Unadjusted means for OVS sentence interpretation task Rank ANCOVA (Immediate Posttest)

Group	**Mean**	**Std. Deviation**	**N**
1 [-EI][-TE SI]	49.54545	34.094621	33
2 [-EI][+TE SI]	80.38333	22.441373	30
3 [+EI][-TE SI]	82.56250	17.042984	24
4 [+EI][+TE SI]	81.38636	22.741263	22
5 (Control)	33.16667	28.836753	21
Total	65.50000	32.411418	130

Figure 18 below makes it clear that knowledge at pretest was rather uniform across groups, as shown by the rather small range in scores at pretest (6.91 logits). Prior to treatment Group 5 (Control) had the lowest knowledge of OVS sentence interpretation at 62.09, and Group 4 [+EI][+TE SI] exhibited the highest level of knowledge at 68.90 logits.

Interestingly, and as was the case in OVS picture matching, although there is clear indication of gains from pretest to immediate posttest in sentence interpretation overall, Groups 1 [-EI] [-TE SI] and 5 (Control) appear to score lower after treatment than prior to it. Although these two groups were the ones to receive the weakest (if any) treatment, their performance in the immediate posttest still begs an explanation. As shown in the Procedures section, the sentence interpretation task was the last one of the exit measures to be administered. As such, one might simply argue participants in those groups where the nature of the treatment did not allow for principled decisions lost motivation and/or focus towards the end of the session. However, the reliability of the OVS sentence interpretation task was .88, and no misfitting persons were found, which suggests that although certain participants may not have monitored their responses with the same rigor as they did in previous tasks, their answers were still not erratic.

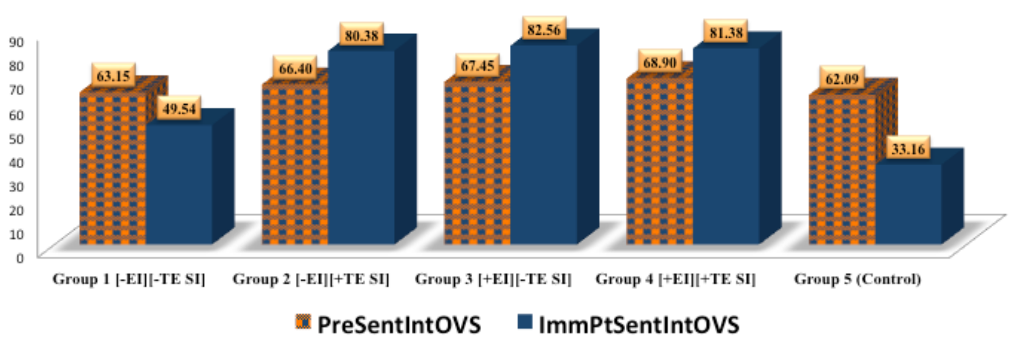
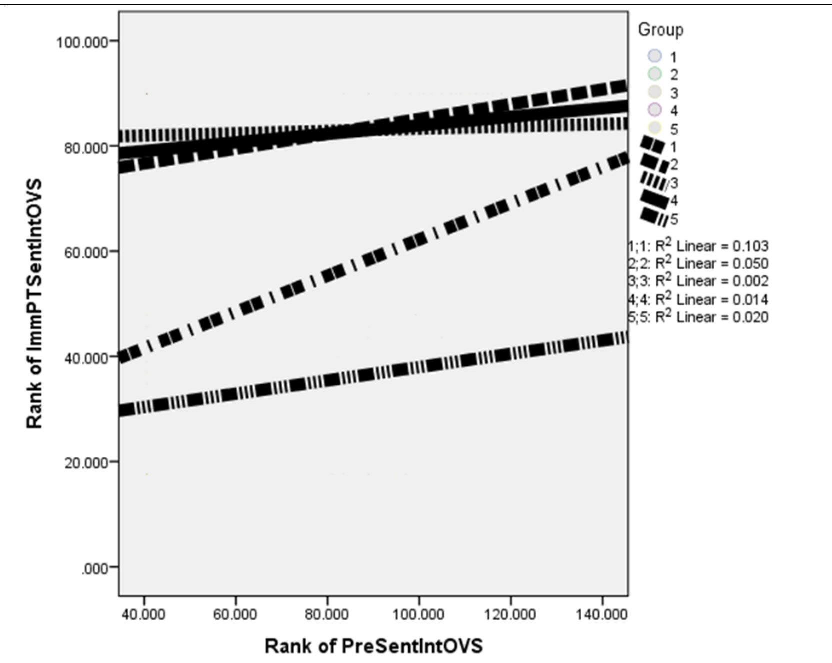


Figure 18: Mean ability scores for OVS sentence interpretation (Pretest and Immediate Posttest)

Analyses of unadjusted means (Table 24) revealed a main effect for Group, F (4, 120) = 5.61; $p < .001$. The influence of the covariate on the dependent variable was significant also, F (4, 120) = 4.03; $p < .05$, suggesting that in this case prior knowledge did have an impact over scores immediately after treatment. Finally, no interaction was found between Groups and prior knowledge, F (4, 120) = .658; p = .622, which indicates that the effect of prior knowledge was comparable across groups. This can be graphically confirmed by looking at Figure 19, where despite the somewhat different slopes, the orientation of the lines is common for all groups, with the exception of Group 4 [+EI][+TE SI], which showed virtually no prior knowledge effect, again, presumably as a result of receiving both EI and TE practice.



Group 1 [-EI][-TESI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 19: Regression slopes for OVS sentence interpretation

Unfortunately, Levene's test was highly significant, $F(4, 125) = 6.10$; $p < .001$, which suggests the variance of scores across groups was not equal and that the assumption of homogeneity of variance was not met. Therefore, the results of our analysis should be interpreted with the utmost caution.

Table 24: Rank ANCOVA analysis of unadjusted means for OVS sentence interpretation task (Immediate Posttest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	54595.346[a]	9	6066.150	8.996	.000	.403
Intercept	78477.443	1	78477.443	116.379	.000	.492
Group	15154.594	4	3788.649	5.618	.000	.158
RPreSent	2721.232	1	2721.232	4.035	.047	.033
Group * RPreSent	1774.759	4	443.690	.658	.622	.021
Error	80919.154	120	674.326			
Total	693247.000	130				
Corrected Total	135514.500	129				

The means obtained by each of the groups remained virtually identical after the effects of previous knowledge were controlled for (Table 25). Although, descriptively speaking, the ordering of groups did change slightly (Group 3 [+EI][-TE SI] was the highest scoring one and not Group 4 [+EI][+TE SI]), once again the [+EI] groups emerged at the top and Groups 2 [-EI][+TE SI] and 1[-EI][-TE SI] took third and 4th position respectively, followed by Group 5 (Control) in final position.

Table 25: Adjusted means for OVS sentence interpretation Rank ANCOVA (Immediate Posttest)

Group	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 [-EI][-TESI]	50.354[a]	4.533	41.379	59.329
2 [-EI][+TESI]	80.257[a]	4.743	70.867	89.647
3 [+EI][-TESI]	82.520[a]	5.310	72.007	93.033
4 [+EI][+TESI]	81.109[a]	5.567	70.086	92.131
5 (Control)	33.597[a]	5.700	22.312	44.883

Given the lack of change in the means, unsurprisingly, the Group effect prevailed, F (4,120) = 17.57; $p < .001$, suggesting that there are differences across groups and these were caused by the different treatments received (Table 26).

Table 26: Rank ANCOVA analysis of adjusted means for sentence interpretation (Immediate Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	47402.736	4	11850.684	17.574	.000	.369
Error	80919.154	120	674.326			

Results from the Bonferroni post-hoc test (Table 27) revealed that the differences were between Group 1[-EI][-TE SI], and Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI], as well as between Group 5 (Control) and Groups 2 [-EI][+TE SI], 3[+EI][-TE SI] and 4 [+EI][+TE SI].

Table 27: Pairwise comparison for OVS sentence interpretation task Rank ANCOVA (Immediate Posttest)

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TESI]	-29.903*	6.561	.000	-48.665	-11.140
	3[+EI][-TESI]	-32.166*	6.982	.000	-52.132	-12.199
	4[+EI][+TESI]	-30.755*	7.179	.000	-51.287	-10.223
	5(Control)	16.757	7.283	.231	-4.071	37.584
2[-EI][+TESI]	1[-EI][-TE SI]	29.903*	6.561	.000	11.140	48.665
	3[+EI][-TESI]	-2.263	7.119	1.000	-22.624	18.098
	4[+EI][+TESI]	-.852	7.313	1.000	-21.768	20.063
	5(Control)	46.659*	7.415	.000	25.453	67.865
3[+EI][-TESI]	1[-EI][-TESI]	32.166*	6.982	.000	12.199	52.132
	2[-EI][+TESI]	2.263	7.119	1.000	-18.098	22.624
	4[+EI][+TESI]	1.411	7.693	1.000	-20.591	23.413
	5(Control)	48.922*	7.790	.000	26.644	71.201
4[+EI][+TESI]	1[-EI][-TESI]	30.755*	7.179	.000	10.223	51.287
	2[-EI][+TESI]	.852	7.313	1.000	-20.063	21.768
	3[+EI][-TESI]	-1.411	7.693	1.000	-23.413	20.591
	5(Control)	47.512*	7.968	.000	24.725	70.298
5(Control)	1[-EI][-TESI]	-16.757	7.283	.231	-37.584	4.071
	2[-EI][+TESI]	-46.659*	7.415	.000	-67.865	-25.453
	3[+EI][-TESI]	-48.922*	7.790	.000	-71.201	-26.644
	4[+EI][+TESI]	-47.512*	7.968	.000	-70.298	-24.725

These results are largely in consonance with previous group differences and, once again, suggest that EI is the element that generates most gains. Unlike in previous analyses, however, we see a difference between Group 1 [-EI][-TE SI] and 2 [-EI][+TE SI] here, which suggests that TE did generate significant gains. Lack of difference between Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] further suggest that the gains obtained from TE alone, EI alone, or the combination of both, are not statistically different.

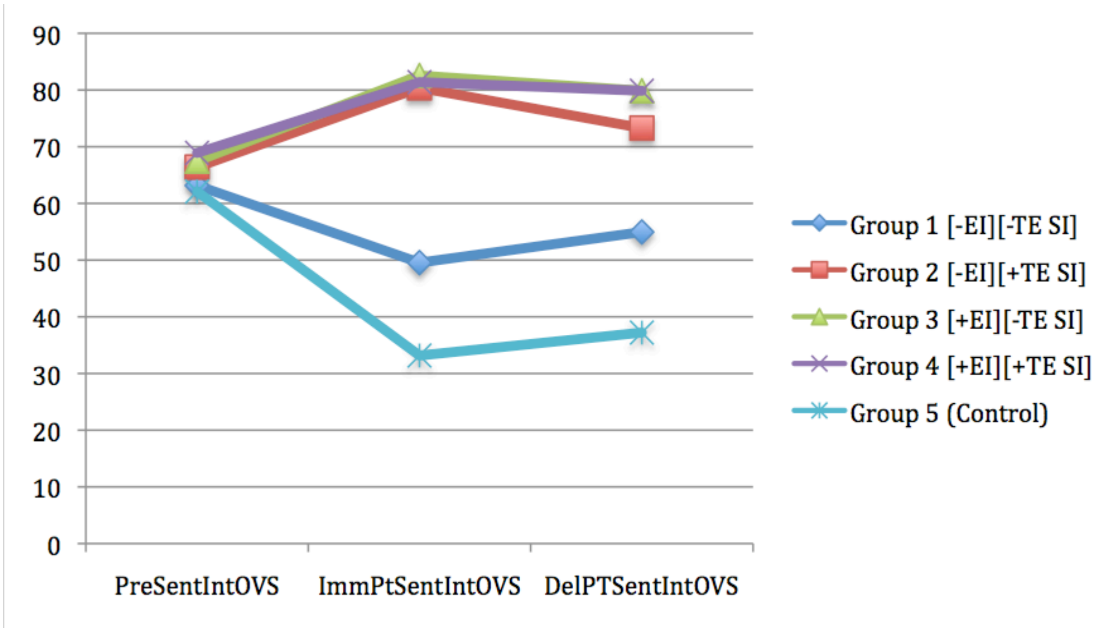
6.6. OVS Sentence Interpretation – Delayed Posttest

To finish with the OVS results, this section will report on a Rank ANCOVA comparing means obtained by the different groups in the sentence interpretation delayed posttest. As Table 28 below shows, results from the posttest are consonant with immediate posttest results, and largely with results from the previous two OVS tasks. Three weeks after treatment, unadjusted means revealed the same picture as the immediate posttest. Groups 3 [+EI] [-TE SI] and 4 [+EI][+TE SI], i.e., the [+EI] groups, emerge as the highest scoring conditions with means just slightly lower than before (81.85 and 79.86 respectively). These are followed by Group 2 [-EI][+TE SI], which at 73. 09 exhibits a mean seven logits below its immediate posttest counterpart. In fourth position we find Group 1 [-EI] [-TE SI], at 54.95 logits, and finally the Control Group, (Group 5) with 37.21 logits.

Table 28: Unadjusted means for OVS sentence interpretation Rank ANCOVA (Delayed Posttest)

Group	Mean	Std. Deviation	N
1 [-EI][-TESI]	54.95455	36.662388	33
2 [-EI][+TESI]	73.28333	31.584229	30
3 [+EI][-TESI]	81.85417	29.562196	24
4 [+EI][+TESI]	79.86364	28.971325	22
5 (Control)	37.21429	28.451086	21
Total	65.50000	35.119287	130

As Figure 20 below shows, three of the four experimental groups appear to have retained at least some of the gains derived from the treatment.



Figures 20: Development over time for OVS Sentence interpretation.

As can be clearly seen in Figure 21, Groups 1[-EI] [-TE SI] and 5 (Control), again exhibit lower post-treatment means here than on the pretest. We infer, as we did before, that this owes to the fact that those groups were particularly lacking in terms of rules on which to base their decisions and ended up feeling confused as a result of the treatment.

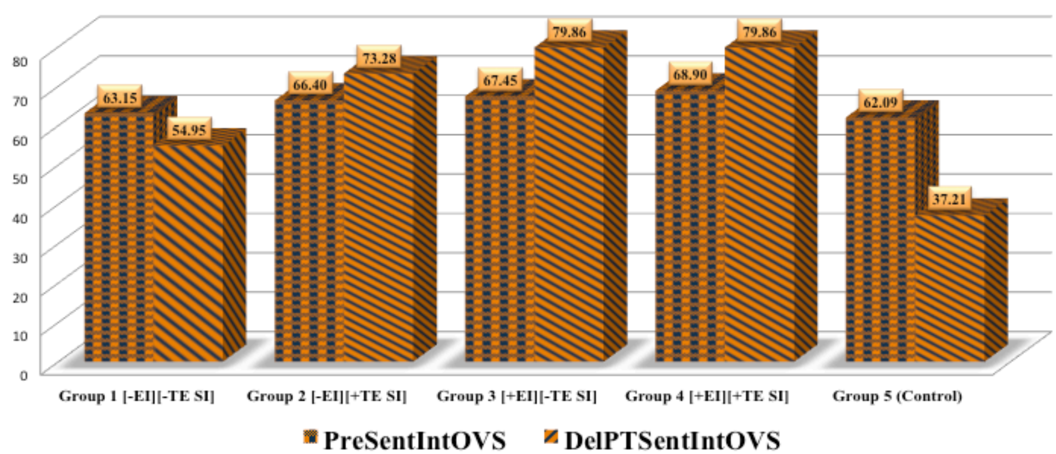


Figure 21: Mean ability scores for OVS sentence interpretation (Pretest and Delayed Posttest)

In addition, the bar graph (Figure 22) showing performance of all groups in all three testing sessions reveals that, as could be expected, all groups perform slightly below their immediate posttest mean in the delayed posttest.

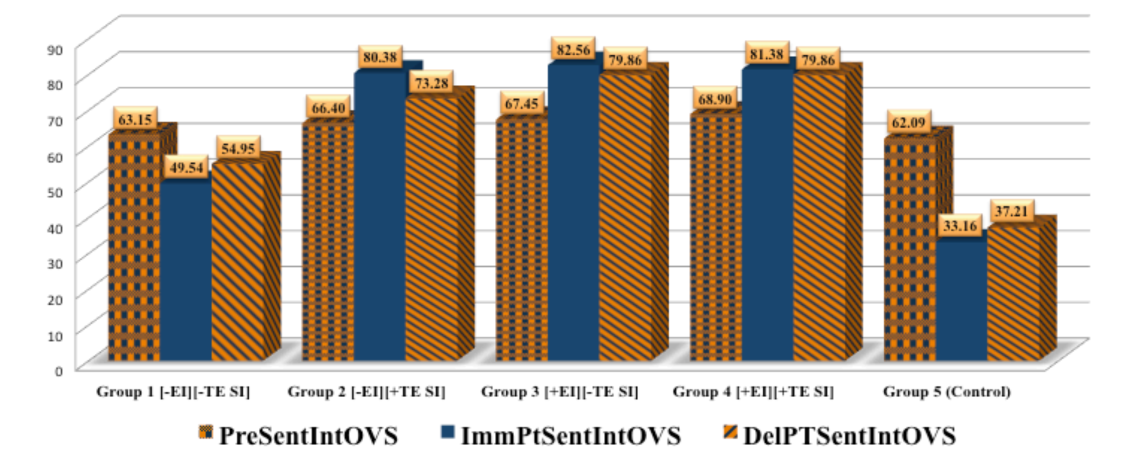


Figure 22: Mean scores for OVS sentence interpretation (Pretest, Immediate and Delayed Posttest)

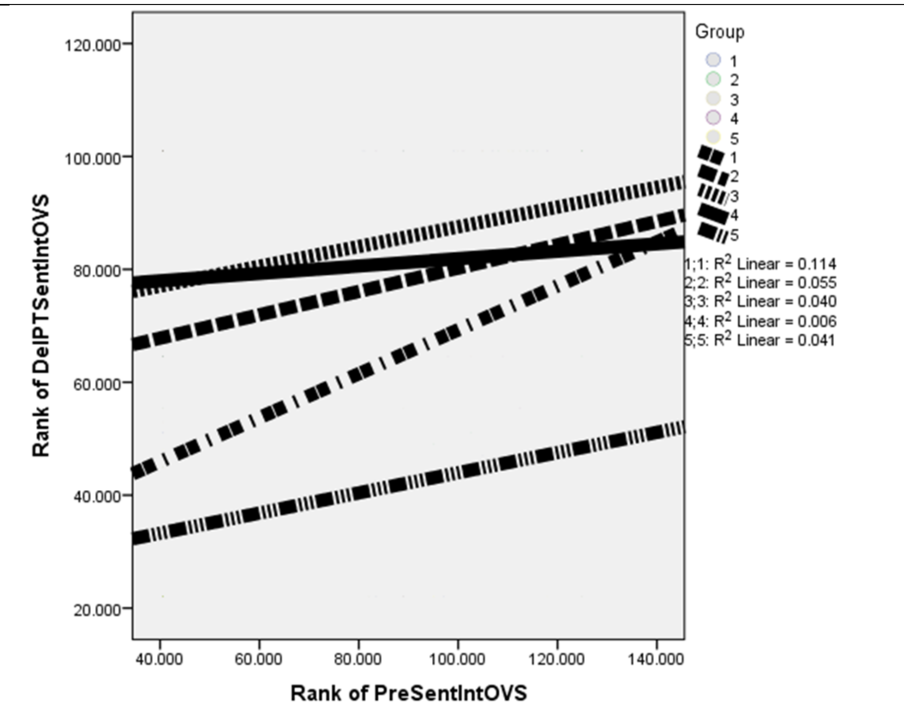
Levene's test was not significant, $F(4, 125) = 1.29$; $p = .274$, which indicates that the error variance was equal across groups.

After analyzing unadjusted means, a Group effect was found, $F\ (4, 120) = 4.67$; $p < .001$, indicating that some of the gains maintained resulted in significant differences across groups, which originated in the different treatments (Table 29). In addition, the covariate, prior knowledge, was found to be significantly related to the dependent variable as shown by the $F\ (1,120) = 5.59$; $p < .05$. Finally, no interaction was found between the grouping variable and prior knowledge. $F\ (4,120) = 3.98$; $p = .809$.

Table 29: Rank ANCOVA analysis of unadjusted means for OVS sentence interpretation (Delayed Posttest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	41291.941[a]	9	4587.993	4.673	.000	.260
Intercept	67924.594	1	67924.594	69.186	.000	.366
Group	9934.071	4	2483.518	2.530	.044	.078
RPreSent	5490.025	1	5490.025	5.592	.020	.045
Group * RPreSent	1564.419	4	391.105	.398	.809	.013
Error	117812.059	120	981.767			
Total	716836.500	130				
Corrected Total	159104.000	129				

A similar result was found when plotting the slopes of each group, as shown in Figure 20. With the exception of Group 4 [+EI][+TE SI], which continues to display a rather flat line, the groups are approximately parallel, suggesting that all treatment gains were maintained across all groups in a comparable manner, regardless of prior knowledge.



Group 1 [-EI][-TESI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 23: Regression slopes for OVS sentence interpretation (Pretest and Delayed Posttest)

Except for Group 3 [+EI] [-TE SI], regression slopes for all groups were similar from immediate to delayed posttest. Group 3 [+EI] [-TE SI] presents a greater boost in performance at the delayed posttest than at the immediate posttest stage. It would appear from that difference that the immediacy of treatment may have reduced the effect of initial proficiency in the earlier testing stage. When the immediate effects of treatment start wearing out, at least certain learners in Group 3 [+EI] [-TE SI] saw their performance stand out as a result of their prior knowledge.

Much like we saw in the immediate posttest, the difference between unadjusted and adjusted means in the sentence interpretation delayed posttest is negligible and the ordering of the groups remains exactly the same (Table 30).

Table 30: Adjusted means for OVS sentence interpretation Rank ANCOVA (Delayed Posttest)

Group	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TE SI]	55.869[a]	5.470	45.039	66.698
2 [-EI][+TE SI]	73.097[a]	5.723	61.767	84.427
3 [+EI][-TE SI]	81.511[a]	6.407	68.826	94.197
4 [+EI][+TE SI]	79.640[a]	6.717	66.340	92.940
5 (Control)	37.822[a]	6.878	24.205	51.440

These differences in means yielded a significant group effect, F (4,120) = 7.92; $p < .001$, as shown by Table 31. Thus, it would appear that the gains in sentence interpretation were maintained for three weeks, at least in some of the groups.

Table 31: Analysis of adjusted means for OVS sentence interpretation (Delayed Posttest)

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Contrast	31149.686	4	7787.421	7.932	.000	.209
Error	117812.059	120	981.767			

Unsurprisingly, the pairwise comparison (Table 32) revealed that although the difference between Group 1 [-EI][-TE SI] and 2 [-EI][+TE SI] recorded in the immediate posttest had disappeared, the rest of the significant pairwise contrasts had been maintained. That is to say, a significant difference was found between Group 1 [-EI][-TE SI] and Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI], and between Group 5 (Control) and Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI]. It seems that the smaller gains generally found for Group 2 [-EI][+TE SI] are the most likely to regress to the point of non-significance, whereas the greater gains obtained by [+EI] groups are more durable.

Table 32: Pairwise comparison for OVS sentence interpretation Rank ANCOVA (Delayed Posttest)

(I) Group	**(J) Group**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Confidence Interval for Difference**	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TE SI]	-17.229	7.916	.315	-39.868	5.410
	3[+EI][-TE SI]	-25.643*	8.424	.029	-49.735	-1.551
	4[+EI][+TE SI]	-23.772	8.663	.070	-48.546	1.002
	5(Control)	18.046	8.788	.422	-7.085	43.177
2[-EI][+TESI]	1[-EI][-TE SI]	17.229	7.916	.315	-5.410	39.868
	3[+EI][-TE SI]	-8.414	8.590	1.000	-32.982	16.153
	4[+EI][+TE SI]	-6.543	8.824	1.000	-31.780	18.694
	5(Control)	35.275*	8.947	.001	9.687	60.862
3[+EI][-TESI]	1[-EI][-TE SI]	25.643*	8.424	.029	1.551	49.735
	2[-EI][+TE SI]	8.414	8.590	1.000	-16.153	32.982
	4[+EI][+TE SI]	1.871	9.283	1.000	-24.677	28.419
	5(Control)	43.689*	9.400	.000	16.807	70.570
4[+EI][+TESI]	1[-EI][-TE SI]	23.772	8.663	.070	-1.002	48.546
	2[-EI][+TE SI]	6.543	8.824	1.000	-18.694	31.780
	3[+EI][-TE SI]	-1.871	9.283	1.000	-28.419	24.677
	5(Control)	41.818*	9.614	.000	14.324	69.312
5(Control)	1[-EI][-TE SI]	-18.046	8.788	.422	-43.177	7.085
	2[-EI][+TE SI]	-35.275*	8.947	.001	-60.862	-9.687
	3[+EI][-TE SI]	-43.689*	9.400	.000	-70.570	-16.807
	4[+EI][+TE SI]	-41.818*	9.614	.000	-69.312	-14.324

Results from a Robust ANCOVA, an alternative non-parametric analysis robust both to non-normalcy and heteroscedasticity, concurred with the results above (see Tables 33 and 34) for both testing times, as shown by the significant Group effects.

Table 33: Results of Robust ANCOVA for OVS sentence interpretation (Immediate Posttest)

	DF	RD	MRD	F	p-value
Groups	4	1434.1957	385.5489	90.73669	0
Homog	4	472.8871	118.2218	29.91796	0

Table 34: Results of Robust ANCOVA for OVS sentence interpretation (Delayed Posttest)

	DF	RD	MRD	F	p-value
Groups	4	485.15940	121.28985	11.946572	0.0001
Homog	4	94.68721	23.67180	2.331579	0.05977

6.7. Ser/Estar Production – Immediate Posttest

We now turn to the results for the second linguistic target this paper focused on, i.e., the ser/estar copula distinction. In Table 35 we see the unadjusted means for ser/estar production immediately after treatment. A glance at the different group means quickly reveals that the ser/estar treatments appear to have a similar effect on the participants as the ones for OVS. As we saw for some cases of OVS, Group 2 [-EI][+TE SI] obtains scores similar to these for Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI]. Thus, we see Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] with means of 74.40, 73.62 and 73.62 respectively as the highest scoring groups with almost identical means. Groups 1 [-EI][-TE SI] and 5 (Control) at 59 and 58.80, exhibit clearly lower means to occupy the same positions (third and fourth) as they did in all of the OVS tasks as well.

Table 35: Unadjusted means for ser/estar production ANCOVA (Immediate Posttest)

production SE	Mean	Std. Deviation	N
1 [-EI][-TE SI]	59.00000	12.151344	30
2 [-EI][+TE SI]	74.40000	14.580573	30
3 [+EI][-TE SI]	73.42308	13.559272	26
4 [+EI][+TE SI]	73.62500	9.595345	24
5 (Control)	58.80952	11.487467	21
Total	68.03817	14.346484	131

Graphic presentation of group performance in the production pretest and the immediate posttest (Figure 24) immediately reveals that, based on unadjusted means, all groups experienced improvement after treatment. Worthy of note in this respect is that the initial level of knowledge is very similar, with all groups exhibiting ability scores between 52 and 55 logits.

Perhaps predictably, improvement was strongest for the groups with explicit information and/or task-essential practice, i.e., Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI]).

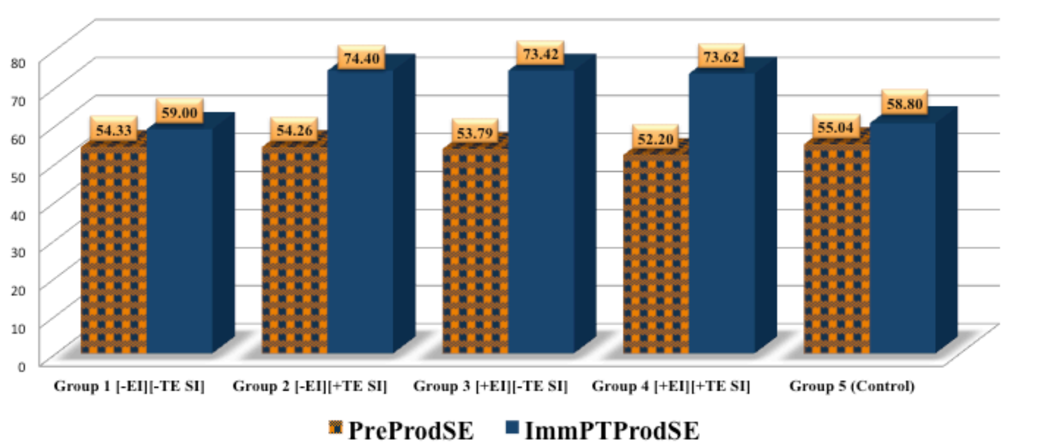


Figure 24: Mean ability scores for ser/estar production (Pretest and Immediate Posttest)

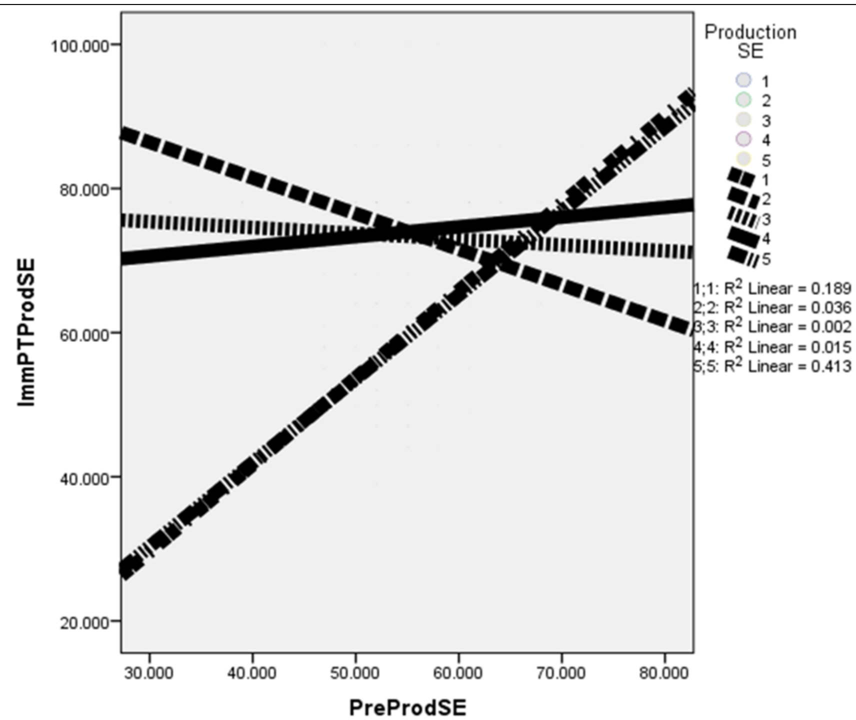
Levene's test, $F(4, 126) = 1.39$; $p = .239$, indicates equal error variance across groups; the assumption of homogeneity of variance is met.

Inferential analysis of the unadjusted (Table 36) scores yielded a significant main effect for Group, $F(4,121) = 4.49$; $p < .01$, as well as a significant main effect for the covariate, prior knowledge, $F(1,121) = 4.67$; $p < .05$. Thus, prior knowledge appeared to have a significant effect on immediate posttest scores and prior to controlling for that effect, the mean differences across groups were significant. In addition, an interaction between prior knowledge and Group was found, $F(4,121) = 3.18$; $p < .05$.

Table 36: ANCOVA analysis of unadjusted means for ser/estar production (Immediate Posttest)

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	9119.029[a]	9	1013.225	6.951	.000	.341
Intercept	3387.993	1	3387.993	23.243	.000	.161
Group	2619.622	4	654.906	4.493	.002	.129
PreProdSE	681.791	1	681.791	4.677	.033	.037
Group * PreProdSE	1854.095	4	463.524	3.180	.016	.095
Error	17637.781	121	145.767			
Total	633181.000	131				
Corrected Total	26756.809	130				

A plot of the group slopes for mean ability scores (Figure 25) indeed reveals that they are not equal in their orientation. Group 2 [-EI][+TE SI] exhibits a downward trend whereas Groups 1 [-EI] [-TE SI] and 5 (Control) exhibited an identical upward trend. Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI] exhibited a rather flat line, thus evidencing no effect for prior knowledge. These patterns indicate that something in the treatments upset the knowledge that participants in Groups 2 [-EI][+TE SI] gave evidence of having at pretest. The opposite was true for the rest of the groups, 1 [-EI] [-TE SI] and 5 (Control), i.e., something in the treatment activated with the prior knowledge of the participants such that better posttest scores ensued. This interaction constitutes a violation of the assumption of homogeneity of regression slopes, which means that the outcomes from the present ANCOVA should be interpreted with uttermost caution, as the relationship between the covariate and dependent variable is different across groups, which may have resulted in over or underestimation of the observed *p* value.



Group 1 [-EI][-TE SI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 25: Regression slopes for ser/estar production
(Pretest and Immediate Posttest)

Controlling for prior knowledge had virtually no effect over the means, as shown by the adjusted scores in Table 37, and as such the ordering of groups remained Group 2 [-EI][+TE SI], Group 4[+EI][+TE SI], Group 3 [+EI][-TE SI], Group 1 [-EI][-TE SI] and Group 5 (Control), in that order.

Table 37: Adjusted means for ser/estar production ANCOVA (Immediate Posttest)

production SE	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 [-EI][-TESI]	58.542[a]	2.213	54.161	62.923
2 [-EI][+TESI]	74.555[a]	2.208	70.184	78.926
3 [+EI][-TESI]	73.455[a]	2.371	68.760	78.149
4 [+EI][+TESI]	73.861[a]	2.516	68.880	78.842
5 (Control)	58.202[a]	2.644	52.968	63.437

Bearing the heteroscedastic nature of the data in mind, it is uncertain how safely we can interpret the main effect for Group yielded by the analysis (Figure 23, Table 38), F (4,121) = 12.88; $p < .001$.

Table 38: ANCOVA analysis of adjusted means for ser/estar production (Immediate Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	7510.383	4	1877.596	12.881	.000	.299
Error	17637.781	121	145.767			

A Bonferroni correction (Table 39) revealed that those differences were located between Group 1 [-EI][-TE SI] and Groups 2[-EI][+TE SI], 3[+EI][-TE SI] and 4[+EI][+TE SI], as well as between Group 5 (Control) and Groups 2 [-EI][+TE SI], 3[+EI][-TE SI] and 4[+EI][+TE SI]. These results represent a familiar pattern, as we saw very similar pairwise contrasts for OVS. Thus, the immediate posttest data for ser/estar production seem to confirm that both [+TE SI] alone and [+EI] alone treatments result in significant gains, that combining [+TE SI] and [+EI] does not seem to result in greater gains than exposure to one of them alone, and that in the absence of TE, SI does not appear to generate any gains.

Table 39: Pairwise comparison for ser/estar production ANCOVA (Immediate Posttest)

(I) production SE	(J) production SE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TE SI]	-16.013*	3.126	.000	-24.951	-7.074
	3[+EI][-TE SI]	-14.913*	3.243	.000	-24.187	-5.638
	4[+EI][+TE SI]	-15.319*	3.351	.000	-24.899	-5.738
	5(Control)	.340	3.448	1.000	-9.519	10.199
2[-EI][+TESI]	1[-EI][-TE SI]	16.013*	3.126	.000	7.074	24.951
	3[+EI][-TE SI]	1.100	3.240	1.000	-8.165	10.364
	4[+EI][+TE SI]	.694	3.347	1.000	-8.877	10.265
	5(Control)	16.352*	3.445	.000	6.503	26.202
3[+EI][-TESI]	1[-EI][-TE SI]	14.913*	3.243	.000	5.638	24.187
	2[-EI][+TE SI]	-1.100	3.240	1.000	-10.364	8.165
	4[+EI][+TE SI]	-.406	3.457	1.000	-10.292	9.479
	5(Control)	15.253*	3.552	.000	5.097	25.408
4[+EI][+TESI]	1[-EI][-TE SI]	15.319*	3.351	.000	5.738	24.899
	2[-EI][+TE SI]	-.694	3.347	1.000	-10.265	8.877
	3[+EI][-TE SI]	.406	3.457	1.000	-9.479	10.292
	5(Control)	15.659*	3.650	.000	5.223	26.095
5(Control)	1[-EI][-TE SI]	-.340	3.448	1.000	-10.199	9.519
	2[-EI][+TE SI]	-16.352*	3.445	.000	-26.202	-6.503
	3[+EI][-TE SI]	-15.253*	3.552	.000	-25.408	-5.097
	4[+EI][+TE SI]	-15.659*	3.650	.000	-26.095	-5.223

6.8. Ser/Estar Production – Delayed Posttest

This section will examine the durability of the gains reported in the previous section. The table of unadjusted means (Table 40) below displays means that are lower than before for the groups that scored highest on the ser/estar production immediate posttest 2 [-EI][+TE SI], 3[+EI][-TE SI] and 4[+EI][+TE SI].

Table 40: Unadjusted means for ser/estar production ANCOVA (Delayed posttest)

production SE	**Mean**	**Std. Deviation**	**N**
1 [-EI][-TE SI]	64.73333	16.933813	30
2 [-EI][+TE SI]	68.70000	18.221151	30
3 [+EI][-TE SI]	71.23077	12.919157	26
4 [+EI][+TE SI]	69.12500	14.918073	24
5 (Control)	68.86000	14.401389	21
Total	68.25954	15.692030	131

Although clearly some of the gains were retained, a quick glance at Figure 26 reveals that losses, as one might expect, also occurred.

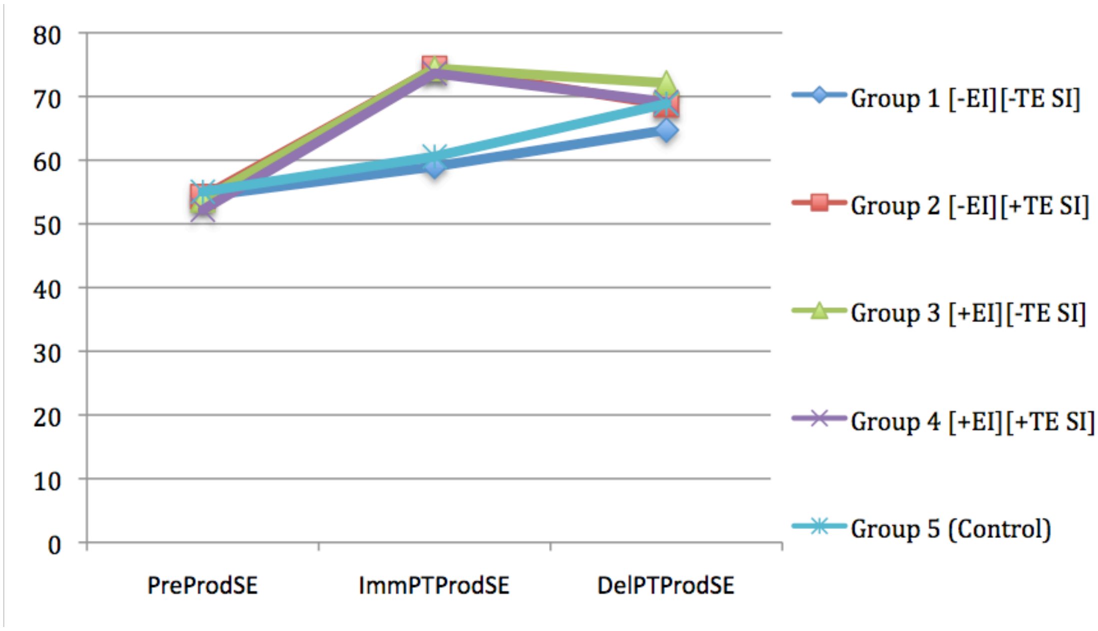


Figure 26: Development over time for ser/estar production.

From Figure 27 it is clear that Group 2 [-EI][+TE SI] lost the highest amount of logits, going from 74 logits to 68.70. Exhibiting a larger retention rate, Group 3 [+EI][-TE SI] went from 73.42 to 71.23 (thereby becoming the highest scoring group in the delayed posttest) and finally Group 4 [+EI][+TE SI] exhibited a 69.12 as opposed to the 73.62 in the immediate posttest.

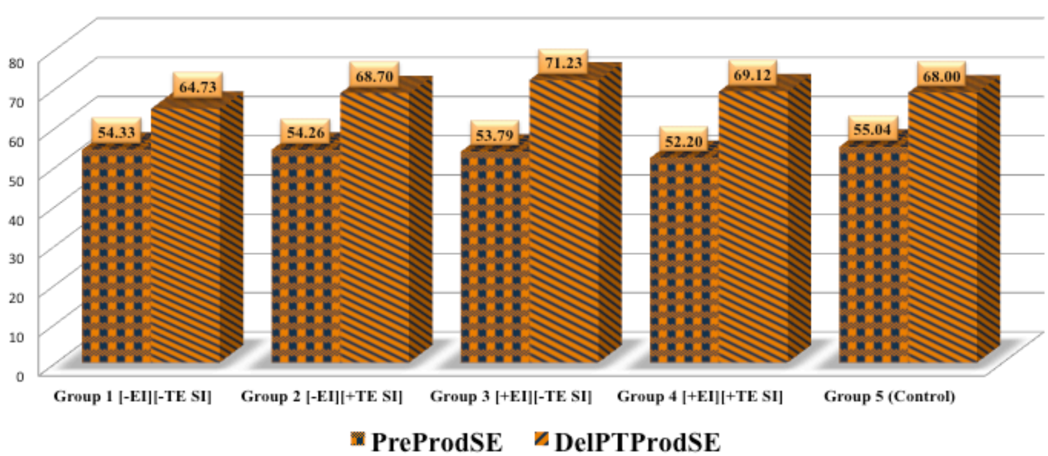


Figure 27: Mean ability scores for ser/estar production (Pretest and Delayed Posttest)

Given the time elapsed between immediate and delayed posttest, these losses are not surprising. The same cannot be said, however, about the change in performance exhibited by the remaining two groups, Group 1 [-EI][-TE SI] and Group 5 (Control). As Figure 28 shows, Group 1 [-EI][-TE SI] and Group 5 (Control) experience a pronounced improvement at the delayed posttest stage (while still scoring lower than the other three groups, however).

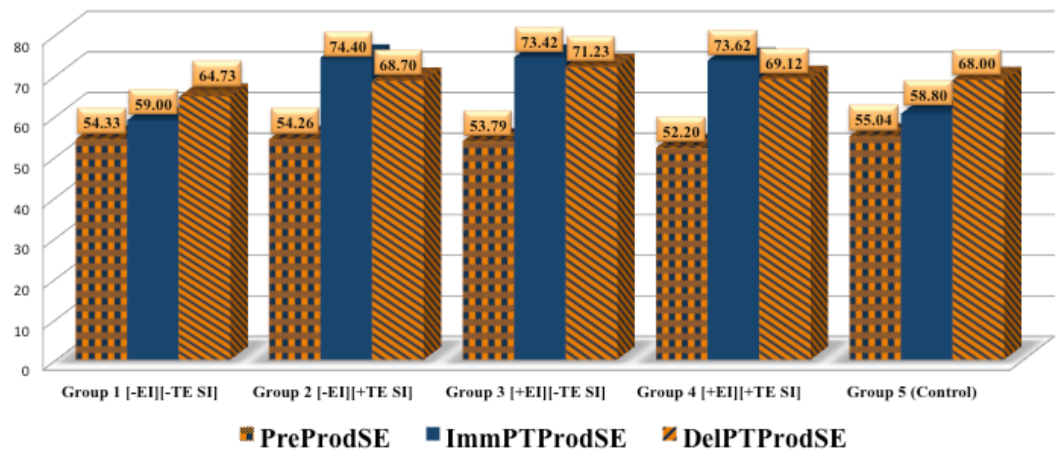


Figure 28: Mean ability scores for ser/estar production (Pretest, Immediate and Delayed Posttest)

In the absence of treatment between immediate and delayed posttests, this improvement could only originate from learning at testing or from a difference in the difficulty of the testing instruments. The latter would be troubling, as an easier delayed posttest could result in faulty interpretation of the results observed for Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI], i.e. we may interpret the means in those groups as retained gains when they could have simply come from the greater ease of the instrument. To rule out this option we looked at the mean, median and mode logit of difficulty of the items in the pre-, post- and delayed posttests.

Table 41: Mean, median and mode logits of difficulty for ser/estar production (Pretest, Immediate and Delayed)

Ser/estar production	Pretest Item Difficulty	Immediate Posttest Item Difficulty	Delayed Postes Item Difficulty
Mean	49.77	50.90	49.532
Median	48.25	51.68	50.15
Mode	39.30	57.78	50.15

As Table 41 shows, all three instruments had very similar means and medians, which would suggest that there would be no instrument-associated reason for groups to perform better in the delayed posttest. In view of the figures above, we conclude that the increase in scores exhibited by Groups 1 [-EI][-TE SI] and 5 (Control) appear to originate in the repeated testing.

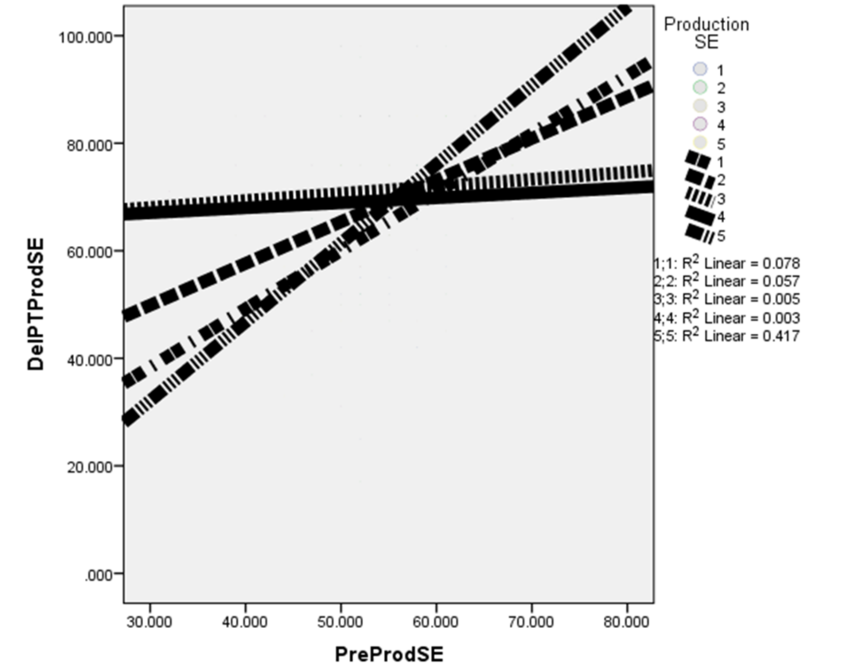
Levene's test was non-significant, $F(4, 126) = 1.15$; $p = .336$, and therefore the assumption of homogeneity of variance has been met.

Given the closeness in scores exhibited in the non-adjusted means, it will come as no surprise that no group effect was found, $F(4,121) = 1.720$; $p = .150$, (Table 42). However, given the increase in scores that Groups 1[-EI][-TE SI] and 5 (Control) experienced, it is possible that the lack of an effect is not due to a loss of gains but rather due to the better performance of the two lowest-scoring groups. A main effect was found for the covariate, $F(1,121) = 9.80$; $p < .01$, suggesting that prior knowledge did have an impact on the scores in the delayed posttest. Finally, no interaction was found between Group and prior knowledge, $F(1,121) = 1.569$; $p = .187$, suggesting that whatever effect the covariate may have had, it affected all participants within their respective conditions equally.

Table 42: ANCOVA analysis of unadjusted means for ser/estar production (Delayed Posttest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3594.400[a]	9	399.378	1.701	.096	.112
Intercept	1368.186	1	1368.186	5.826	.017	.046
Group	1615.617	4	403.904	1.720	.150	.054
PreProdSE	2303.631	1	2303.631	9.809	.002	.075
Group * PreProdSE	1474.314	4	368.578	1.569	.187	.049
Error	28416.775	121	234.849			
Total	642388.000	131				
Corrected Total	32011.176	130				

Plotting the regression lines for each group (Figure 29), we see indeed that with the exception of Groups 3 [+EI] [-TE SI] and 4 [+EI][+TE SI], which exhibit no effect of prior knowledge, line orientation is similar for all groups. However the slopes are not parallel, which, despite the numeric analysis, would indicate that certain levels of prior knowledge result in an increase in scores. Although not to a significant level, it seems that Groups 1[-EI][-TE SI], 2 [-EI] [+TE SI] and, especially, Group 5 (Control) saw their performance enhanced three weeks after treatment as a function of their prior knowledge.



Group 1 [-EI][-TESI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 29: Regression slopes for ser/estar production (Delayed Posttest)

Worthy of note with respect to the regression of the slopes at the immediate posttest stage is the trend exhibited by Group 2 [-EI][+TE SI]. It was evident from the graphic plotting of regression lines at immediate posttest that at least for some members of Group 2 [-EI][+TE SI] the combination of treatment and prior knowledge resulted in underperformance. Three weeks later, the slope for this group exhibits a clear upward trend. This change presumably was brought about by the fact that those learners whose proficiency was upset by the treatment reverted to relying on their own prior knowledge, whether due to the absence of reinforcement or because they could not rehash the knowledge derived from the treatment.

After controlling for prior knowledge, the scores remained, once again, virtually identical and the ordering of the groups was, therefore, the same (Table 43): Groups 3 [+EI][-TE SI], 4 [+EI][+TE SI], 2 [-EI][+TE SI], 1 [-EI][-TE SI] and 5 (Control) in that order.

Table 43: Adjusted means for ser/estar production ANCOVA (Delayed Posttest)

Production SE	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TE]	64.322[a]	2.809	58.761	69.883
2 [-EI][+TE]	68.458[a]	2.802	62.910	74.006
3 [+EI][-TE]	71.179[a]	3.010	65.220	77.138
4 [+EI][+TE]	69.287[a]	3.193	62.965	75.609
5 (Control)	67.235[a]	3.356	60.591	73.879

Inferential analyses of the adjusted means concurred with the analyses of the unadjusted ones and yielded no Group effect (Table 44).

Table 44: ANCOVA analysis of adjusted means for ser/estar production (Delayed Posttest)

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Contrast	721.337	4	180.334	.768	.548	.025
Error	28416.775	121	234.849			

6.9. Ser/Estar Picture matching – Immediate Posttest

In the next section we will be looking at the results from the ser/estar picture-matching task immediately after treatment. As before, we start by looking at the unadjusted means (Table 45), which reveal a hierarchy largely in concert with previous ser/estar and OVS tasks. In this case Group 4 [+EI][+TE SI] was the highest scoring group, with 85.9 logits, followed by Groups 2 [-EI][+TE SI] and 3[+EI][-TE SI], with means of 81.55 and 82.83 respectively. Quite unlike in previous cases, Group 5 (Control) follows with a mean of 75.75 and finally Group 1 [-EI][-TE SI] emerges as the lowest-scoring group.

Table 45: Unadjusted means for ser/estar picture matching ANCOVA (Immediate Posttest)

Picture matching SE	Mean	Std. Deviation	N
1 [-EI][-TE SI]	66.93750	18.850751	32
2 [-EI][+TE SI]	82.83333	14.417622	30
3 [+EI][-TE SI]	81.55556	12.882108	27
4 [+EI][+TE SI]	85.90909	11.096585	22
5 (Control)	75.75000	23.623973	20
Total	78.12214	17.755183	131

A glance at Figure 30 below reveals that all groups stood at around 62 and 65 logits at pretest, and that all treatments generated some gains, most obviously for Groups 2 [-EI][+TE SI] and 4[+EI][+TE SI].

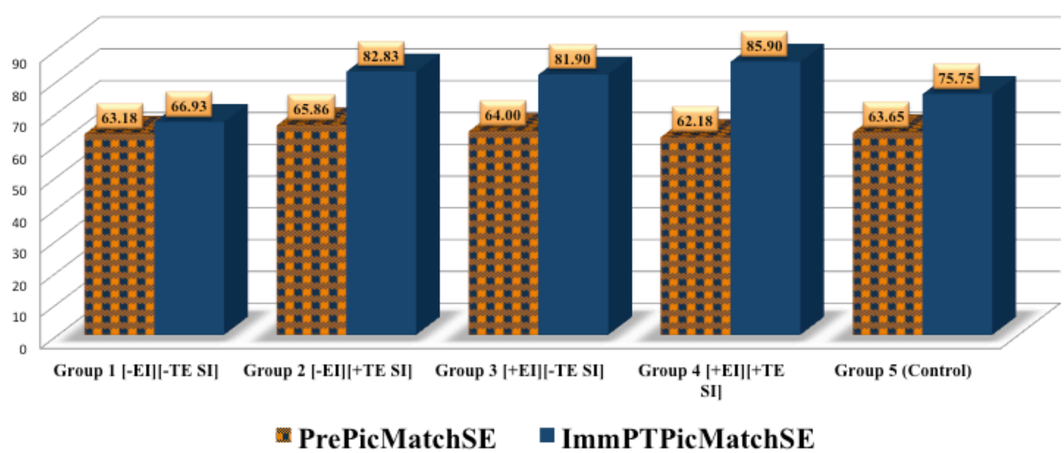


Figure 30: Mean ability scores for ser/estar picture matching (Pretest and Immediate Posttest)

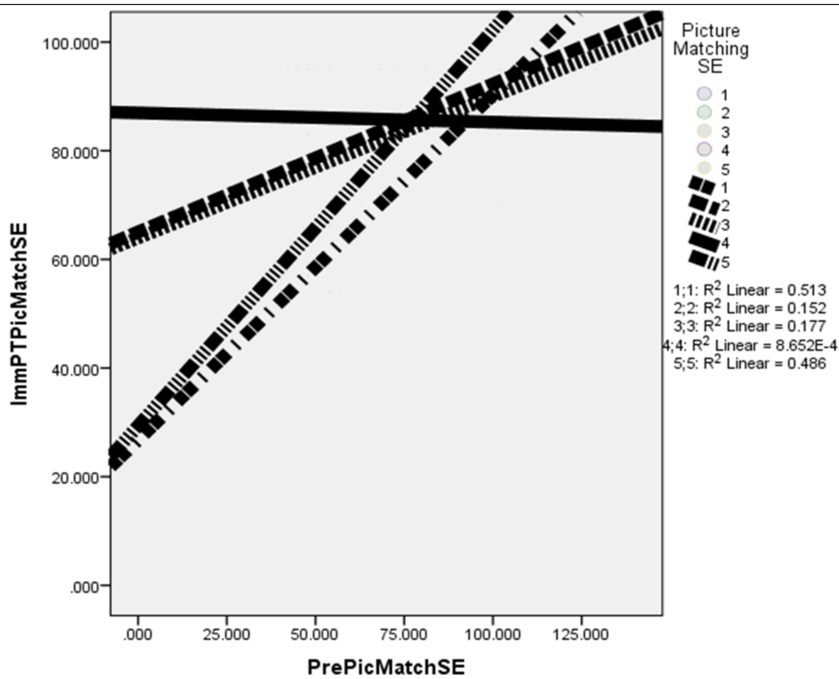
Levene's test was not significant, $F(4, 126) = .87$; $p = .468$, suggesting that the assumption of homogeneity of variance had been met.

Analysis of the unadjusted means (Table 46) revealed a Group effect, F (4,121) = 8.40, $p < .001$, as well as a significant effect for prior knowledge, F (1,121) = 40.96, $p < .001$ and an interaction between Group and prior knowledge, F (4,121) = 4.928, $p < .01$.

Table 46: ANCOVA analysis of unadjusted means for ser/estar picture matching (Immediate Posttest)

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	18928.833[a]	9	2103.204	11.540	.000	.462
Intercept	34432.450	1	34432.450	188.922	.000	.610
Group	6123.792	4	1530.948	8.400	.000	.217
PrePicMatchSE	7466.156	1	7466.156	40.965	.000	.253
Group * PrePicMatchSE	3592.575	4	898.144	4.928	.001	.140
Error	22053.212	121	182.258			
Total	840484.000	131				
Corrected Total	40982.046	130				

After plotting the regression lines (Figure 31), we see that lines for Groups 1 [-EI][-TE SI], 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 5 (Control) have the same orientation. That said, it is clear from the crossed slopes that the extent to which prior knowledge enhanced learner performance was far greater in the case of Groups 1 [-EI][-TE SI] and 5 (Control), than for Groups 2 [-EI][+TE SI] and 3 [+EI][-TE SI], as the slopes for the former two groups are far steeper. Something in the treatment interacted with a certain level of prior knowledge possessed by participants in Group 1 [-EI][-TE SI] and 5 (Control). Given the nature of conditions 1 [-EI][-TE SI] and 5 (Control), it may be that the pronounced upward trend owes precisely to the absence of treatment; in other words, it is possible that upon receiving no information or practice that could help learners deduce the rules governing the structures at hand, whatever prior knowledge they may have possessed was accountable for whatever positive performance they had. Finally, Group 4 [+EI][+TE SI] once again gives evidence of no prior knowledge effect, which may well respond to the fact that having the proper access to both +EI and +TE practice, learners in Group 4 [+EI][+TE SI] were largely able to overcome gaps in their initial knowledge.



Group 1 [-EI][-TE SI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 31: Regression slopes for ser/estar picture matching (Immediate Posttest)

Adjustment for prior knowledge caused little change in the means (Table 47). All scores varied from the unadjusted figures in less than one logit, which resulted in the same ordering of groups, with Group 4 [+EI][+TE SI] at the top, followed by 2 [-

EI][+TE SI] and 3 [+EI][-TE SI] and Groups 5 (Control) and 1 [-EI][-TE SI] in fourth and fifth position.

Table 47: Adjusted means for ser/estar picture matching - ANCOVA (Immediate Posttest)

Picture matching SE	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 [-EI][-TE SI]	67.750[a]	2.391	63.016	72.484
2 [-EI][+TE SI]	82.606[a]	2.467	77.722	87.490
3 [+EI][-TE SI]	80.760[a]	2.627	75.559	85.961
4 [+EI][+TE SI]	85.870[a]	2.900	80.130	91.611
5 (Control)	76.344[a]	3.021	70.364	82.325

The differences in ser/estar picture matching tasks yielded a significant effect for Group, F (4,121) = 7.65; $p < .001$ (Table 48).

Table 48: ANCOVA analysis of adjusted means for ser/estar picture matching (Immediate Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	5580.024	4	1395.006	7.654	.000	.202
Error	22053.212	121	182.258			

Pairwise comparisons (Table 49) revealed that the differences, once again, originated in the contrast between Group 1 [-EI][-TE SI] and Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI]. Group 5 (Control) was not different from either Group 1 [-EI][-TE SI] or the higher-scoring groups. As was the case in previous analyses, then, it would seem that it absence of EI and TE practice Group 1 [-EI][-TE SI] could not perform significantly better than the control. In addition, the lack of difference between the scores of Group 5 (Control) and Groups 2[-EI][+TE SI], 3 [+EI][-TE SI], and 4 [+EI][+TE SI] seems to indicate that none of the groups learned very much, as shown by the fact that none of the treatments yielded an effect significantly larger than the practice effect in the control group.

Table 49: Pairwise comparison for ser/estar picture matching ANCOVA (Immediate Posttest)

(I) Picture matching SE	(J) picture matching SE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
1 [-EI][-TESI]	2[-EI][+TE SI]	-14.856*	3.435	.000	-24.680	-5.033
	3[+EI][-TE SI]	-13.010*	3.552	.004	-23.167	-2.853
	4[+EI][+TE SI]	-18.120*	3.758	.000	-28.867	-7.374
	5(Control)	-8.594	3.853	.275	-19.611	2.422
2 [-EI][+TESI]	1[-EI][-TE SI]	14.856*	3.435	.000	5.033	24.680
	3[+EI][-TE SI]	1.846	3.604	1.000	-8.458	12.151
	4[+EI][+TE SI]	-3.264	3.807	1.000	-14.150	7.621
	5(Control)	6.262	3.900	1.000	-4.890	17.414
3 [+EI][-TESI]	1[-EI][-TE SI]	13.010*	3.552	.004	2.853	23.167
	2[-EI][+TE SI]	-1.846	3.604	1.000	-12.151	8.458
	4[+EI][+TE SI]	-5.110	3.913	1.000	-16.298	6.077
	5(Control)	4.415	4.003	1.000	-7.032	15.863
4 [+EI][+TESI]	1[-EI][-TE SI]	18.120*	3.758	.000	7.374	28.867
	2[-EI][+TE SI]	3.264	3.807	1.000	-7.621	14.150
	3[+EI][-TE SI]	5.110	3.913	1.000	-6.077	16.298
	5(Control)	9.526	4.187	.247	-2.447	21.499
5 (Control)	1[-EI][-TE SI]	8.594	3.853	.275	-2.422	19.611
	2[-EI][+TE SI]	-6.262	3.900	1.000	-17.414	4.890
	3[+EI][-TE SI]	-4.415	4.003	1.000	-15.863	7.032
	4[+EI][+TE SI]	-9.526	4.187	.247	-21.499	2.447

6.10. Ser/Estar Picture matching – Delayed Posttest

As for the durability of the gains for the ser/estar distinction in picture matching, the unadjusted means below exhibit scores lower than the ones obtained in the immediate posttest, although scores were higher than at pretest (Table 50).

Figure 32 reveals that all experimental groups experienced considerable loss from immediate to delayed posttest.

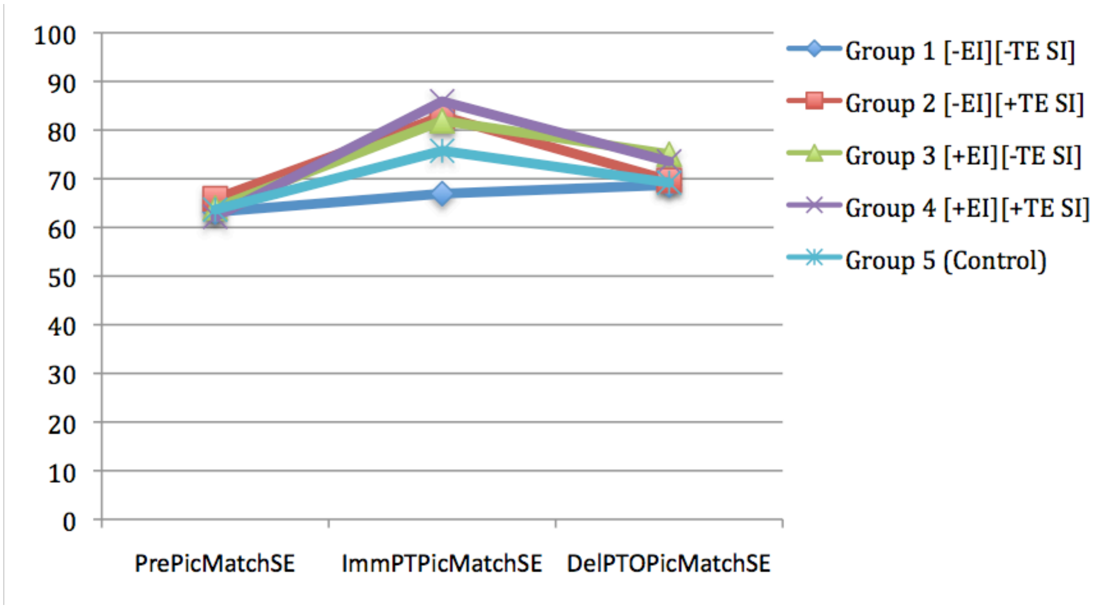


Figure 32: Development over time for ser/estar picture matching.

As shown in Figures 33 and 34, descriptively, Group 3 [+EI][-TE SI] emerged as the group that retained the largest portion of gains, going from 81.55 to 75.03 logits and was the highest scoring group in the delayed posttest. It was followed by the other [+EI] group, Group 4 [+EI][+TE SI], whose score was 73.54 (from 85.90 in the posttest), and by Groups 2 [-EI][+TE SI] and 5 (Control), with almost identical means around the 70 logits. Interestingly, Group 2 [-EI][+TE SI] presented the greatest loss going from nearly 83 logits to 69.66. Finally, Group 1[-EI][-TE SI], once again in fifth position with almost the exact same score as in the immediate posttest.

Table 50: Unadjusted means for ser/estar picture matching ANCOVA (Delayed Posttest)

Picture matching SE	Mean	Std. Deviation	N
1 [-EI][-TE SI]	68.71875	17.035512	32
2 [-EI][+TE SI]	69.66667	18.532978	30
3 [+EI][-TE SI]	75.03704	13.642359	27
4 [+EI][+TE SI]	73.54545	19.291566	22
5 (Control)	69.20000	12.288634	20
Total	71.12214	16.489168	131

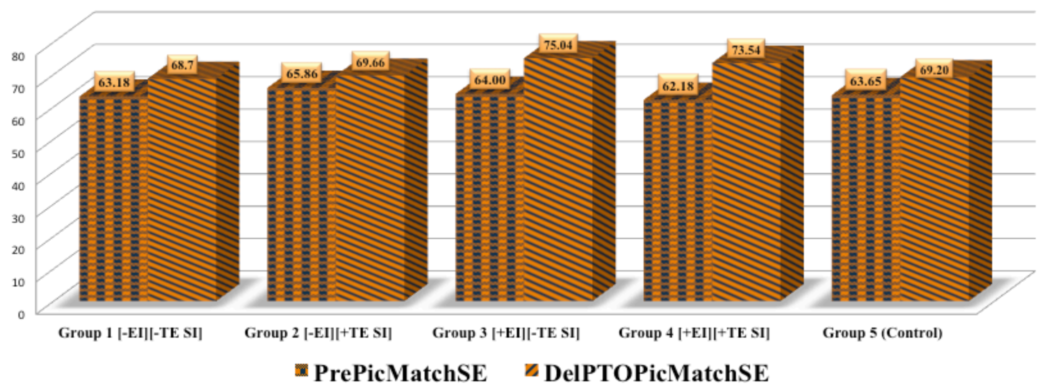


Figure 33: Mean ability scores for ser/estar picture matching (Pretest and Delayed Posttest)

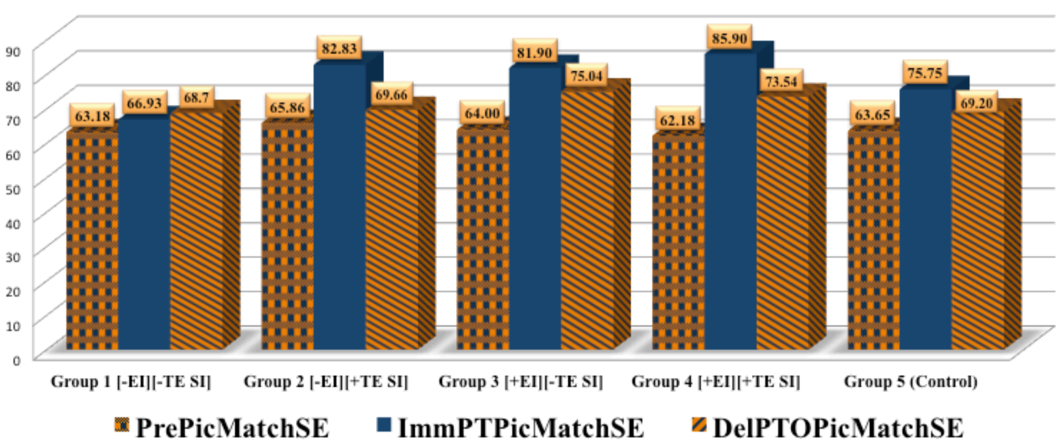


Figure 34: Mean ability scores for (Pretest, Immediate and Delayed Posttest)

The data from ser/estar picture matching task at delayed posttest presented homogeneous variance in errors across groups, as shown by the non-significant Levene's test, $F(4, 126) = 1.38$; $p = .244$.

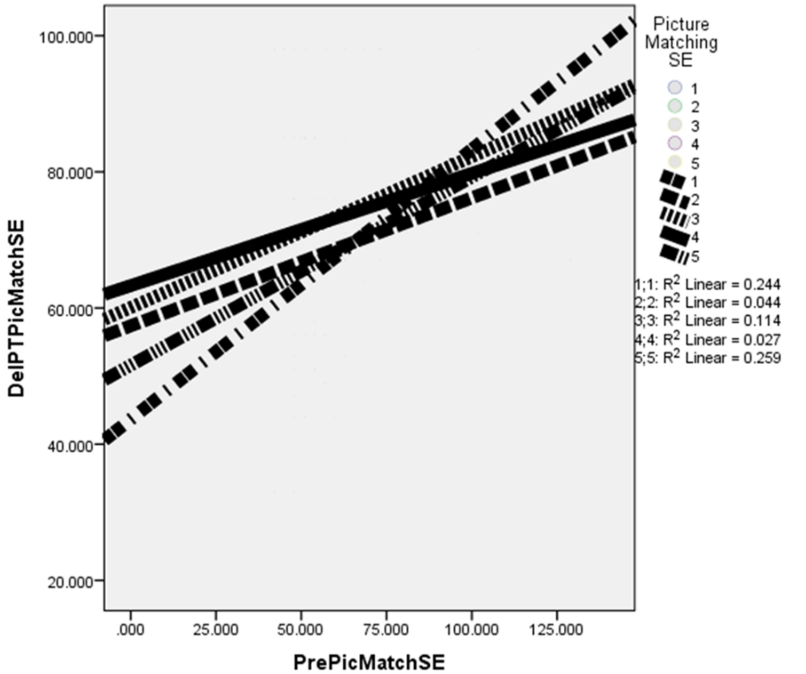
The differences in means across groups in the delayed posttest yielded no significant effect, $F\ (4, 121) = .619$; $p = .650$, indicating that the improvement experienced by Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] immediately after treatment had been lost three weeks later (Figure 51). Similar results were obtained in the delayed posttest of ser/estar production and thus, it would appear that although [+TE SI] and [+EI] treatments generate benefits for both OVS and ser/estar, gains obtained for OVS are more robust than those for ser/estar. Still, a significant effect for prior knowledge was found, $F\ (1, 121) = 13.09$; $p < .001$, which suggested that initial familiarity with the ser/estar distinction had an impact on the scores that were retained. Finally, the lack of interaction between Group and prior knowledge, $F\ (4, 121) = .412$; $p = .799$, indicated that prior knowledge had a parallel impact across groups.

Table 51: ANCOVA analysis of unadjusted means for ser/estar picture matching (Delayed Posttest)

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	5004.107[a]	9	556.012	2.217	.025	.142
Intercept	35437.213	1	35437.213	141.319	.000	.539
Group	621.229	4	155.307	.619	.650	.020
PrePicMatchSE	3283.892	1	3283.892	13.096	.000	.098
Group * PrePicMatchSE	413.625	4	103.406	.412	.799	.013
Error	30341.939	121	250.760			
Total	697991.000	131				
Corrected Total	35346.046	130				

As Figure 35 shows, three weeks after treatment, the effects of prior knowledge appear to be comparable for Groups 2 [-EI][+TE SI] and 3 [+EI][-TE SI], and arguably even Group 4 [+EI][+TE SI], whereas Groups 1 [-EI][-TE SI] and 5 (Control), as had been the case in the immediate posttest, give evidence of having had their performances boosted as a function of prior knowledge to a greater degree. Presumably, as a result of the effects of the treatment wearing off over time, it appears that the effects of prior knowledge for Group 4 [+EI][+TE SI] became evident and more similar to those exhibited by Groups 2 [-EI][+TE SI] and 3 [+EI][-TE SI].

Despite some commonalities, the regression slopes for this particular task were rather different at the immediate and delayed posttest stages. Whereas immediately after treatment, regression slopes for Groups 2 [-EI][+TE SI] and 3 [+EI] [-TE SI] were disordinal with respect to those of Groups 1[-EI][-TE SI] and 5 (Control), after three weeks the interaction appears to be much lower and therefore the regression lines are far closer to parallel (despite some crossing) for those groups.



Group 1 [-EI][-TESI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 35: Regression slopes for ser/estar picture matching (Delayed Posttest)

As in previous cases, adjustment of the obtained means for prior knowledge resulted in very little changes to the delayed posttest scores, and the ordering of groups remained unaffected (Table 52).

Table 52: Adjusted means for ser/estar picture matching ANCOVA (Delayed Posttest)

picture matching SE	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TE SI]	69.225[a]	2.805	63.673	74.778
2 [-EI][+TE SI]	69.510[a]	2.894	63.781	75.238
3 [+EI][-TE SI]	74.361[a]	3.081	68.260	80.461
4 [+EI][+TE SI]	73.922[a]	3.401	67.189	80.655
5 (Control)	69.425[a]	3.543	62.410	76.440

After adjustment, inferential analyses yielded the same results as before; thus, no group effect was present (Table 53).

Table 53: ANOVA analysis of adjusted means for ser/estar picture matching (Delayed Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	696.849	4	174.212	.695	.597	.022
Error	30341.939	121	250.760			

6.11. Ser/Estar Sentence Interpretation – Immediate Posttest

Finally, we turn to the last task in the ser/estar sequence of outcome measures. Table 54 below presents the unadjusted means of the ser/estar sentence interpretation task at the immediate posttest stage. A number of things are by now familiar, such as the higher scores obtained by [+EI] groups. In this case it is Group 4 [+EI][+TE SI] that exhibits the highest mean score (94.93), far higher than in any of the previous tasks, whether ser/estar or OVS. Group 3 [+EI][-TE SI] comes in second, albeit with a far lower mean, namely, 73.55. Next we have Group 2 [-EI][+TE SI], at 69.11 and Group 5 (Control), with a surprisingly high mean (68.48) given the lack of treatment. Finally, the lowest scores where obtained by Group 1 [-EI][-TE SI], whose mean logit was 35.55.

Table 54: Unadjusted means for ser/estar sentence interpretation Rank ANCOVA (Immediate Posttest)

Sentence interpretation SE	Mean	Std. Deviation	N
1 [-EI][-TE SI]	35.55000	28.511447	30
2 [-EI][+TE SI]	69.11667	37.540250	30
3 [+EI][-TE SI]	73.55556	31.142456	27
4 [+EI][+TE SI]	94.93750	23.904844	24
5 (Control)	68.47727	38.612075	22
Total	67.00000	37.321190	133

The comparison of Group performance at pretest and immediate posttest (see Figure 36) for the ser/estar interpretation task is somewhat unusual. On the one hand we see, once again, that the performance of Groups 1 [-EI][-TE SI], 2 [-EI][+TE SI], and 5 (Control) drops in the immediate posttest. This trend was most acute for Group 1[-EI][-TE SI], for which immediate posttest scores were indeed very low. This is, of course, counterintuitive, as it would seem to indicate that participants knew less about the ser/estar distinction after receiving the treatment than before. A similar pattern was found in picture matching and sentence interpretation tasks for OVS, in which case the drop in performance also affected Groups 1[-EI][-TE SI] and 5 (Control). In this case, the reliability index of .70, was once again, if not spectacular, not suggestive of a problem with the task that could result in the sharp decrease in knowledge. As for misfitting participants, only two, both in Group 1 [-EI][-TE SI] gave evidence of random guessing, albeit that their indexes (1.77 and 1.64) were barely over the 1.5 threshold. In view of this, we propose, once again, that part of the reason for this may originate in the fact that for sentence interpretation the nature of the treatments can sometimes be confusing and lead learners to abandon any correct hypothesis they may have been working from originally. While the pretest was the first participation session in the study, the ser/estar sentence interpretation immediate posttest was administered at the end of session 2 (for second semester students) and at the end of session 4 (for first semester students). Interestingly, it appears that those groups receiving less EI-oriented treatments were more likely to lose focus. Such a scenario would not be entirely unreasonable, since participants in Groups 1 [-EI][-TE SI], 5 (Control) and to a lesser extent Group 2 [-EI][+TE SI], may well have found it more difficult to continue to make principled decisions systematically as they neither received rules to act by nor did they practice in environments that were particularly conducive to rule deduction (in the case of Groups 1, [-EI][-TE SI] and 5, Control).

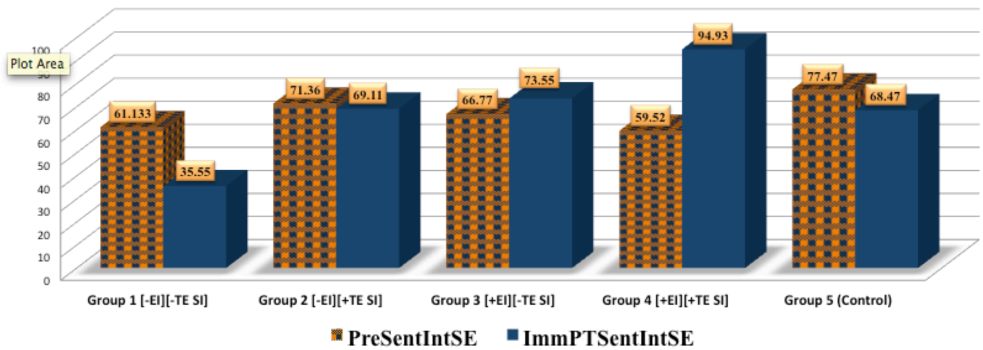


Figure 36: Mean ability scores for ser/estar sentence interpretation task (Pretest and Immediate Posttest)

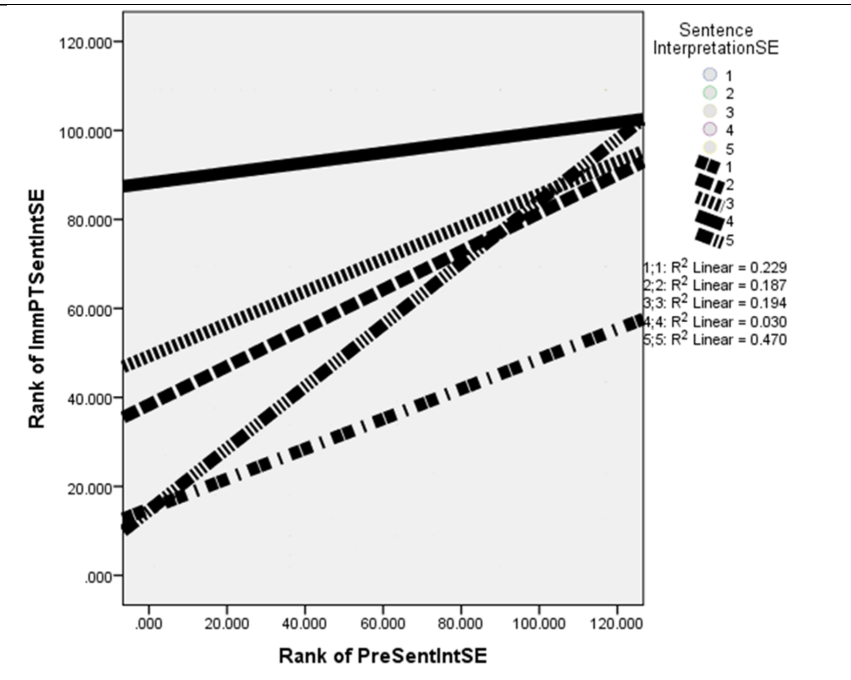
Results from Levene's test, $F(4, 128) = 1.67$; $p = .160$, showed that error variance across groups was not significanty different and therefore that the homogeneity of variance assumption was met.

Based on the unadjusted means, the analysis yielded a significant Group effect, $F\ (4, 123) = 7.13$; $p < .001$, as well as a main effect for prior knowledge, $F\ (1, 123) = 33.153$; $p < .001$. No interaction between Group and previous knowledge was found (see Table 55).

Table 55: Analysis of unadjusted means for ser/estar sentence interpretation (Immediate Posttest)

Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	82796.346[a]	9	9199.594	11.197	.000	.450
Intercept	50980.327	1	50980.327	62.046	.000	.335
Group	23456.441	4	5864.110	7.137	.000	.188
RPreSent	27239.845	1	27239.845	33.153	.000	.212
Group * RPreSent	5382.385	4	1345.596	1.638	.169	.051
Error	101062.654	123	821.648			
Total	780896.000	133				
Corrected Total	183859.000	132				

After plotting the regression slopes (Figure 37) it became evident that all lines exhibited the same orientation, i.e., the different treatments all affected the different groups in a similar way. As we can see in the graph, however, there appears to be a somewhat different effect of initial knowledge for the participants in Group 5 (Control) compared to the treatment groups. Given that there was no relevant treatment intervening between the pretest and the immediate posttest for the control group, it is logical that posttest results for this group strongly reflect those of the pretest. Lines for the rest of the groups are quite parallel, with Group 4 [+EI][+TE SI] displaying a less pronounced upward trend, again, perhaps due to the fact that a more comprehensive treatment made prior knowledge less determinant for participants in this group.



Group 1 [-EI][-TE SI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 37: Regression slopes for ser/estar sentence interpretation (Immediate Posttest)

Contrary to previous ser/estar analyses, after factoring out the effect of prior knowledge, the mean scores for sentence interpretation at the immediate posttest stage did vary significantly, although the ordering of groups remained the same (see Table 56). We see that the mean score for Group 5 (Control) goes down considerably, suggesting again that a large portion of their scores in the immediate posttest originated in prior familiarity with the ser/estar distinction. This is also consistent with the regression slope exhibited by Group 5 (Control), which was steeper than that of the rest of the groups', presumably indicating something in the treatment activated prior knowledge for certain participants in the group, which resulted in higher scores for those particular subjects.

Table 56: Adjusted means for ser/estar sentence interpretation Rank ANCOVA (Immediate Posttest)

Sentence interpretationSE	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 [-EI][-TE SI]	37.526[a]	5.290	27.055	47.997
2 [-EI][+TE SI]	67.235[a]	5.270	56.804	77.666
3 [+EI][-TE SI]	73.636[a]	5.517	62.717	84.556
4 [+EI][+TE SI]	95.784[a]	5.979	83.948	107.620
5 (Control)	61.192[a]	6.349	48.624	73.760

Analysis of the adjusted means yielded a main effect for Group, F (4, 123) = 14.082; $p < .001$ (see Table 57).

Table 57: Rank ANCOVA analysis of adjusted means for ser/estar sentence interpretation (Immediate Posttest)

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Contrast	46282.443	4	11570.611	14.082	.000	.314
Error	101062.654	123	821.648			

Post-hoc analyses (Table 58) located the usual differences between Group 1 [-EI][-TE SI] and Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI]. In this case, the difference between Groups 1 [-EI][-TE SI] and 5 (Control) also appeared to be marginally significant. In addition, although no difference was found between [+EI] groups (Groups 3 and 4), scores for Groups 2 [-EI][+TE SI] and 4 [+EI][+TE SI] were significantly different. This would indicate that in sentence interpretation for ser/estar, receiving EI in addition to TE practice did result in additional gains.

Table 58: Pairwise comparison for ser/estar sentence interpretation Rank ANCOVA (Immediate Posttest)

(I) sentence interpretationSE	(J) sentence interpretationSE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1[-EI][-TESI]	2[-EI][+TE SI]	-29.708*	7.467	.001	-51.052	-8.365
	3[+EI][-TE SI]	-36.110*	7.643	.000	-57.958	-14.262
	4[+EI][+TE SI]	-58.257*	7.983	.000	-81.079	-35.436
	5(Control)	-23.666*	8.264	.049	-47.289	-.042
2 [-EI][+TESI]	1[-EI][-TE SI]	29.708*	7.467	.001	8.365	51.052
	3[+EI][-TE SI]	-6.401	7.629	1.000	-28.209	15.406
	4[+EI][+TE SI]	-28.549*	7.970	.005	-51.332	-5.766
	5(Control)	6.043	8.251	1.000	-17.543	29.629
3 [+EI][-TESI]	1[-EI][-TE SI]	36.110*	7.643	.000	14.262	57.958
	2[-EI][+TE SI]	6.401	7.629	1.000	-15.406	28.209
	4[+EI][+TE SI]	-22.148	8.135	.074	-45.403	1.108
	5(Control)	12.444	8.411	1.000	-11.599	36.487
4 [+EI][+TESI]	1[-EI][-TE SI]	58.257*	7.983	.000	35.436	81.079
	2[-EI][+TE SI]	28.549*	7.970	.005	5.766	51.332
	3[+EI][-TE SI]	22.148	8.135	.074	-1.108	45.403
	5(Control)	34.592*	8.722	.001	9.661	59.523
5 (Control)	1[-EI][-TE SI]	23.666*	8.264	.049	.042	47.289
	2[-EI][+TE SI]	-6.043	8.251	1.000	-29.629	17.543
	3[+EI][-TE SI]	-12.444	8.411	1.000	-36.487	11.599
	4[+EI][+TE SI]	-34.592*	8.722	.001	-59.523	-9.661

6.12. Ser/Estar Sentence Interpretation – Delayed Posttest

Regarding the durability of the gains recorded for the ser/estar interpretation task, as Table 59 shows, a number of things changed rather dramatically from immediate to delayed posttest in this task. First, the highest-scoring group was Group 3 [+EI][-TE SI], which, not only retained but increased by two logits its mean ability score. This is still consistent with the overall picture found in this study in that an [+EI] group is in the lead and gives evidence of most durable gains.

Table 59: Unadjusted means for ser/estar sentence interpretation Rank ANCOVA (Delayed Posttest)

Sentence interpretation SE	Mean	Std. Deviation	N
1 [-EI][-TE SI]	60.33333	35.240145	30
2 [-EI][+TE SI]	63.10000	33.835148	30
3 [+EI][-TE SI]	75.33333	27.411746	27
4 [+EI][+TE SI]	67.77083	32.367784	24
5 (Control)	70.34091	32.519733	22
Total	67.00000	32.448035	133

Figure 38 shows that at the delayed posttest stage all group scores regressed back to roughly those obtained at pretest.

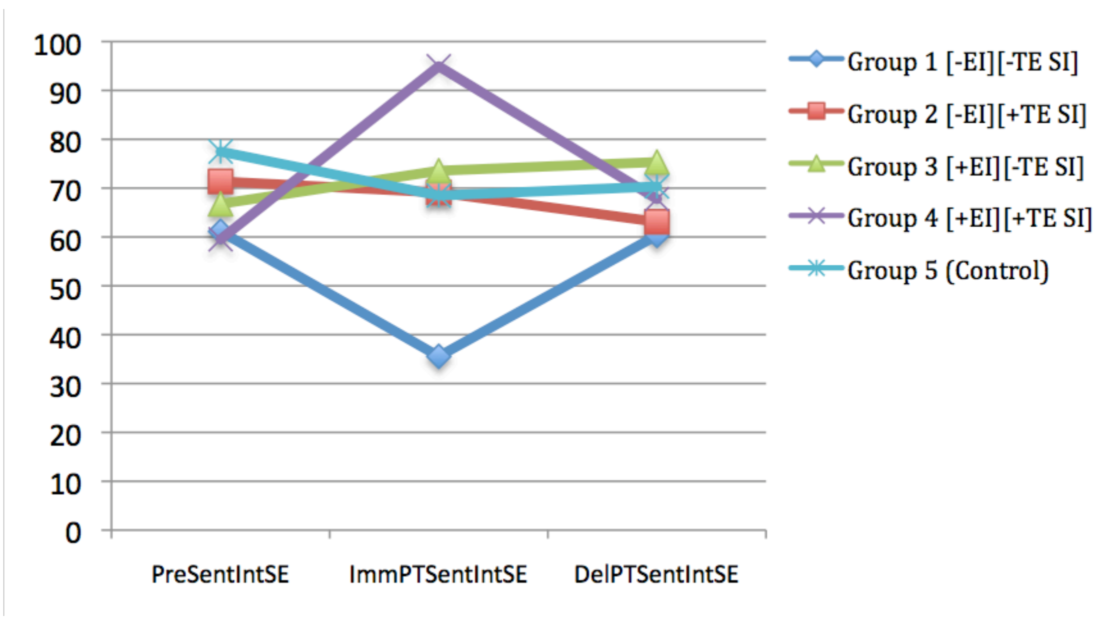


Figure 38: Development over time for ser/estar sentence interpretation

Oddly enough, the second highest scoring group, according to unadjusted means, was Group 5 (Control), with a mean of 70.34 (higher by about 2 logits than it had been in the immediate posttest). In third position we find Group 4 [+EI][+TE SI], with a seemingly steep loss, at 67.70 from the previous 94.93. Finally, we have

Groups 2 [-EI][+TE SI] and 1 [-EI][-TE SI], in that order, the first one with a mean of 63.60 (a 6 logit loss) and the second one exhibiting a performance dramatically higher (30 logits) than at the immediate posttest stage. These very irregular results become obvious from looking at Figures 39 and 40:

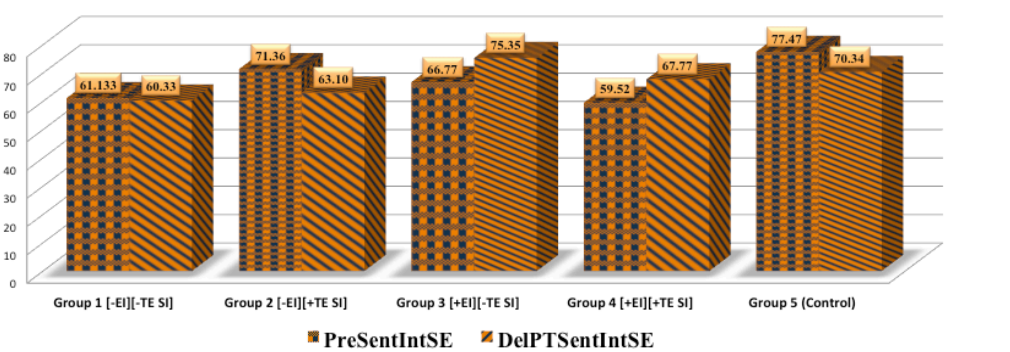


Figure 39: Mean ability scores for ser/estar sentence interpretation (Pretest and Delayed Posttest)

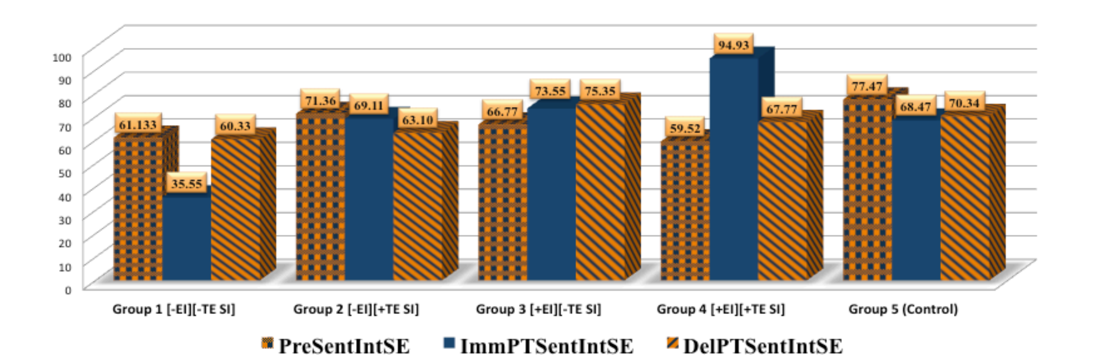


Figure 40: Mean ability scores for ser/estar sentence interpretation (Pretest, Immediate and Delayed Posttest)

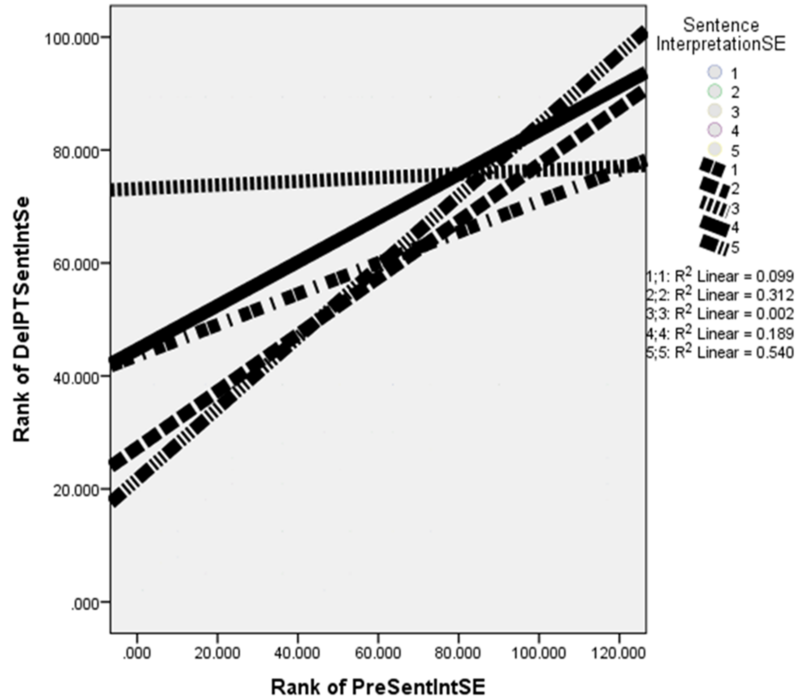
Results from Levene's test were marginally significant, $F(4, 128) = 2.51$; $p < .05$, which constituted a violation of the homogeneity of variance assumption. Given this violation and the interaction between time and prior knowledge (see Figure 35) that emerged upon analyzing unadjusted means, results from the present test should be interpreted with extreme caution.

Inferential analysis of the mean differences in unadjusted means yielded a significant effect, F (4, 123) = 2.715; $p < .05$, that, as we will see, disappears once the means have been adjusted for previous knowledge (see Table 60). The influence of prior knowledge was also significant, F (1, 123) = 28.303; $p < .001$. No significant interaction was found between prior knowledge and Group for the delayed posttest scores.

Table 60: Rank ANCOVA analysis of unadjusted means for ser/estar sentence interpretation (Delayed Posttest)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	34428.179[a]	9	3825.353	4.500	.000	.248
Intercept	53529.716	1	53529.716	62.975	.000	.339
Group	9231.298	4	2307.824	2.715	.033	.081
RPreSent	24057.921	1	24057.921	28.303	.000	.187
Group * RPreSent	7350.437	4	1837.609	2.162	.077	.066
Error	104551.321	123	850.011			
Total	736016.500	133				
Corrected Total	138979.500	132				

It appears from the graphic plotting of the regression (Figure 41) lines that the effect of prior knowledge, which was uniform at the immediate posttest stage for ser/estar sentence interpretation, was not equal for all groups three weeks after treatment. As was the case at the immediate posttest stage, Group 5 (Control) gave evidence of a strong correlation with prior knowledge. This was also the case for Group 2 [-EI][+TE SI], for whom scores went down from immediate to delayed posttest, and where prior knowledge played a larger role on the delayed than on the immediate posttest. Likewise, Group 4 [+EI][+TE SI] also gave evidence of a greater influence of prior knowledge in their delayed posttest performance as compared to the immediate posttest stage, which could owe to the fact the benefits of the treatments wearing off, it was prior knowledge that governed decision making in the latter testing session. An opposite effect was exhibited by Groups 3 [+EI][-TE SI] and 1 [-EI][-TE SI]. Group 3 [+EI][-TE SI] showed no effect of prior knowledge after three weeks, presumably because the [+EI] treatment created new knowledge for the learners in this group, and after three weeks, the benefits having been maintained in this group, prior knowledge ceased to be a factor. Finally, Group 1 [-EI] [-TE SI] was clearly affected by prior knowledge in a different way, as shown by its slope crossing that of Groups 2 [-EI][+TE SI] and 5 (Control). As the inclination of the slope does not appear to have changed much from immediate to delayed posttest for Group 1 [-EI][-TE SI], it would seem that the increase in impact of prior knowledge Groups 2 [-EI][+TE SI] and 5 (Control) experienced, did not obtain for Group 1 [-EI][-TE SI].



Group 1 [-EI][-TESI]	Group 2 [-EI][+TE SI]	Group 3 [+EI][-TE SI]	Group 4[+EI][+TE SI]	Group 5 Control

Figure 41: Regression slopes for ser/estar sentence interpretation (Delayed Posttest)

As we saw for this same task at the immediate posttest stage, adjustment of means for previous knowledge caused changes in the scores and, in this case, also in the ordering of groups (see Table 61). Most notably, the usual order is restored in that it is the two [+EI] groups that occupy the first two positions. Group 5 (Control), now the third highest, continues to exhibit a surprisingly high score, although after adjustment it is considerably lower (7 logits). Finally Groups 1 [-EI][-TE SI] and 2 [-EI][+TE SI], with very similar means and rather moderate losses were fourth and fifth respectively.

Table 61: Adjusted means for ser/estar sentence interpretation - Rank ANCOVA (Delayed Posttest)

Sentence interpretationSE	**Mean**	**Std. Error**	**95% Confidence Interval**	
			Lower Bound	Upper Bound
1 [-EI][-TE SI]	61.939[a]	5.380	51.289	72.589
2 [-EI][+TE SI]	60.909[a]	5.360	50.300	71.518
3 [+EI][-TE SI]	75.341[a]	5.611	64.234	86.447
4 [+EI][+TE SI]	70.668[a]	6.082	58.629	82.706
5 (Control)	63.764[a]	6.458	50.981	76.547

Once means had been adjusted for prior knowledge, and unlike the results from the analysis with the unadjusted means, no significant group effect was found (Table 62). Given the results obtained in the immediate posttest, these findings indicated that although [+TE SI] and [+EI] treatments generated gains, these are differentially long-lasting depending on the structure in question. Thus, while OVS gains appeared to largely prevail after three weeks, ser/estar gains had largely faded by the time participants were given the delayed posttest.

Table 62: Analysis of adjusted means for ser/estar sentence interpretation (Delayed Posttest)

	Sum of Squares	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Contrast	4194.394	4	1048.599	1.234	.300	.039
Error	104551.321	123	850.011			

Given the normality and homogeneity of slopes violations present in the data for this test, a secondary analysis was deemed appropriate in order to ascertain the lack of effect. As shown by Tables 63 and 64, McKean's Robust ANCOVA concurred with the Rank ANCOVA, i.e., a main effect for Group was found, F (4, 4) = 11.426; $p < .001$.

Table 63: Results of Robust ANCOVA for ser/estar sentence interpretation (Immediate Posttest)

	DF	RD	MRD	F	p-value
Groups	4	291.5093	72.87732	11.426276	0.0001
Homog	4	172.2297	43.05742	6.750879	0.0001

Table 64: Results of Robust ANCOVA for ser/estar sentence interpretation (Delayed Posttest)

	DF	RD	MRD	F	p-value
Groups	4	152.4785	38.11963	7.311167	0.0001
Homog	4	139.4320	34.85800	6.685601	0.0001

In the case of the delayed posttest (Table 64), however, McKean's Robust ANCOVA does find an effect for Group, $F\ (4, 4) = 7.311$; $p < .001$, which was not the case for the Rank ANCOVA. The result from the Robust ANCOVA would suggest that gains obtained from the treatment were maintained for the three weeks until the delayed posttest. As we mentioned when comparing results from Rank and Robust ANCOVAs in OVS production, there seems to be an indication here that the Rank ANCOVA is indeed more conservative than the Robust ANCOVA. Given that the present Robust ANCOVA takes into account the heteroscedastic nature of our distribution, we are inclined to rely on its results over those obtained from the Rank ANCOVA. Unfortunately, no pairwise comparisons for the Robust ANCOVA are available and as such we cannot test the significance of the difference between the groups, e.g. between those that descriptively score the highest (Groups 3 and 4) and the others.

Chapter 7: Discussion of Results and Conclusions

7.1. Discussion of Results

Given the large number of analyses reported on above, a summary of all results is in order. As Figure 42 shows, all tasks for both structures yielded a main effect for Group at the immediate posttest stage. For OVS, Groups 1 [-EI][-TE SI] and 5 (Control) obtained scores that were statistically lower than those of Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI], the only exception being OVS production, where no difference was found between Groups 1 [-EI][-TE SI] and 2 [-EI][-TE SI]. This was largely the same for ser/estar, where the same differences emerged with respect to Group 1 [-EI][-TE SI] across all three tasks.

Structure	Immediate Posttest	Group effect?	Location of Statistical Differences
OVS	Production	YES	G1 [-EI][-TE SI] ≠ G3[+EI][-TE SI] & G4[+EI][+TE SI] G2 [-EI][+TE SI] ≠ G3[+EI][-TE SI] & G4[+EI][+TE SI] G5 (Control) ≠ G3[+EI][-TE SI] & G4[+EI][+TE SI]
	Picture Matching	YES	G1 [-EI][-TE SI] ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI] G5 (Control) ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
	Sentence Interpretation	YES	G1 [-EI][-TE SI] ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI] G5 (Control) ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
Ser/estar	Production	YES	G1 [-EI][-TE SI] ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI] G5 (Control) ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
	Picture Matching	YES	G1 [-EI][-TE SI] ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
	Sentence Interpretation	YES	G1 [-EI][-TE SI] ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI], G4[+EI][+TE SI] & G5 (Control) G2 [-EI][+TE SI] ≠ G4[+EI][+TE SI] G5 (Control) ≠ G1 [-EI][-TE SI] & G4[+EI][+TE SI]

Figure 42: Summary of the results (Immediate Posttest)

The behavior of the control group for ser/estar was less straightforward, as its higher scores resulted in a lack of difference between Group 5 and any other group in picture matching and a level of ability statistically higher than that of Group 1 [-EI][-TE SI] in sentence interpretation.

In general, then, results for the immediate posttest were highly homogeneous, showing a systematic superiority for [+EI] groups. With the slight exception of ser/estar sentence interpretation, a generalized effect for treatments consisting of TE only was also recorded. The superiority of the TE only treatment, (received by Group 2 [-EI][+TE SI]), appeared to be somewhat less robust in the case of OVS, as its effects obtained only for the two receptive tasks included in the treatment but did not extend to production. In the case of the ser/estar, the TE only treatment was superior to that of Group 1 [-EI] [-TE SI] both in receptive and productive tasks. In other words, it was always on par with EI treatments. This difference strongly suggests that the type of task and the nature of the processing problem that a given linguistic structure represents will have an impact on the effectiveness of instructional approach such as providing explicit grammar instruction or task-essential practice.

All analyses unanimously indicated that a treatment consisting solely of EI generates gains comparable to those generated by a treatment combining EI and TE. This was shown by the lack of statistical difference between Groups 3 [+EI][-TE SI] and 4 [+EI][+TE SI] across all analyses and structures. From the purely experimental point of view, this means that according to the present results EI is sufficient for benefits to obtain and, given the presence of EI, TE is superfluous. From the point of view of language instruction, however, one may want to argue the issue from a less parsimony-oriented angle. As there is no indication in the performance of Group 4 [+EI][+TE SI] that suggests that receiving EI and TE practice resulted in an undue taxing of participants' attention span, it may be wiser from the instructional point of view to incorporate both components in hopes that over an extended period of time their benefits may add to one another or simply that they may complement each other. In addition, overall, our results also indicated that in the cases where TE generated gains, these were comparable to the gains generated by treatments combining TE and EI. Nevertheless, there was one instance, i.e., ser/estar in the sentence interpretation task, in which the treatment combining both yielded results that were statistically superior to those obtained by the TE only group. We take this as further evidence for what we advanced earlier in this same section, namely, that different processing problems and different task types invite and benefit from different instructional practices.

Structure	Delayed Posttest	Group effect?	Location of Statistical Differences
OVS	Production	Yes	G5 (Control) ≠ G2 [-E][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
	Picture Matching	Yes	G1 [-EI][-TE SI] ≠ G3[+EI][-TE SI] & G4[+EI][+TE SI] G5 (Control) ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
	Sentence Interpretation	Yes	G1 [-EI][-TE SI] ≠ G3[+EI][-TE SI] G5 (Control) ≠ G2 [-EI][+TE SI], G3[+EI][-TE SI] & G4[+EI][+TE SI]
Ser/estar	Production	No	N/A
	Picture Matching	No	N/A
	Sentence Interpretation	Yes[25]	Unavailable

Figure 43: Summary of the results (Delayed Posttest)

As Figure 43 shows, the main effect for group, present across all tasks and structures at the immediate posttest stage, also prevailed in the delayed posttest for all three OVS tasks but only for one of the ser/estar tasks. This finding suggests that although PI, whether in the form of [+TE SI] SI or [+EI] and [+TE SI] SI, can generate gains for diverse kinds of processing problems, the durability of those gains differs depending on the linguistic target. The present data only allow for speculation on this point, but it seems reasonable to assume that the durability of the gains may differ as a function of the difficulty of the rules governing the targeted linguistic phenomena. In this respect, one might argue that whereas the notions of 'doer' and

[25] The result reported for ser/estar Sentence Interpretation at Delayed Posttest is the one yielded by the Robust ANCOVA. The reader may recall that the Rank ANCOVA yielded no effect for Group at the delayed posttest stage. Unlike the Rank ANCOVA, the Robust ANCOVA took into account the heteroscedastic nature of our data. For that reason, its outcomes were considered most appropriate.

‘recipient’ are fairly straightforward and transparent, an aspectual difference involving a distinction between ‘inherent’ and ‘circumstantial’ is rather obscure. Including terms like ‘inherent’ or ‘circumstantial’ in EI calls for supporting definitions that explain to learners what such notions mean. This is hardly the case for notions such as ‘doer’ and ‘recipient’. While judging whether a participant is doing or receiving the action is a decision that can be unequivocally made based on direct evidence, the same cannot be said for judging whether a feature is inherent or not. One might argue, then, that in the absence of EI, the rule that hinging upon the notion of inherence may have been more difficult than one hinging upon the doer/recipient distinction. That being the case, it may be argued that while OVS represents more a question of automatizing newly acquired knowledge, ser/estar represents a conceptual struggle, that once mastered, will yet need to be automatized.

In relation with the above, OVS treatments under TE conditions presented response options that were not only mutually exclusive, but most importantly, exhaustive. Combined with feedback, this reduces to one the number of possible interpretations learners could entertain with respect to the target item. Upon encountering a prompt sentence “Lo besa la chica” (*him kisses the girl*) and two pictures, one where a boy kisses a girl and another where a girl kisses a boy, as the two response options, there is only one possible response: “Lo” can only be the recipient of the action. Creating an equally exhaustive scenario for ser/estar is not quite as simple. First, circumstantial and inherent are not necessarily mutually exclusive and, as we know from the variationist literature (Geeslin 2000) on the copula distinction, learners draw from a myriad of cues when faced with the need to choose between ser and estar. This means that there may be a limit to what we can successfully operationalize (both experimentally and instruction-wise) in a task-essential manner, which may eventually translate in a decreased likelihood that perfect form-meaning mapping will take place, or a longer time until criterion is reached, and/or a decrease in the durability of any gains derived.

Although the main effect for Group was maintained from immediate to delayed posttest for OVS, the pairwise contrasts showed some changes. Indeed Figure 37 reveals that the differences with regard to Group 5 (Control) remained intact whereas for Group 1 [-EI][-TE SI] only the differences with the [+EI] groups were maintained. Clearly, this indicates that scores waned to the point that the performance of [+TE SI] without EI and sometimes even [+TE SI] with EI was comparable to that of participants who received a far less substantial treatment, such as the one administered in Group 1 [-EI][-TE SI]. That said, unlike the performance of Group 1 [-EI][-TE SI], the performance of Groups 2 [-EI][+TE SI], 3 [+EI][-TE SI] and 4 [+EI][+TE SI] always remained superior to that of the control Group, indicating that those three treatment effects are relatively durable, in spite of the sometimes much weaker scores on the delayed posttest than on the immediate posttest.

In OVS production, a difference that was not present at the immediate posttest stage arose between Group 2 [-EI][+TE SI] and 5 (Control) as a result of an increase in the score of the treatment group. As mentioned earlier in the paper, the 14-logit increase experienced by Group 2 [-EI][+TE SI] in OVS production is surprising by itself, given the lack of treatment in the three weeks separating the immediate from the delayed posttest. This finding is especially puzzling as the increase occurs in a task type not included in the treatment. In the Results section we advanced that this may owe to the TE nature of the treatment received by Group 2 [-EI][+TE SI]. However, upon looking at the picture matching and sentence interpretation tasks we see no evidence of the same phenomenon, instead finding that mean scores for Group 2 [-EI] [+TE SI] go down from immediate to delayed posttest. Looking at the overall scores obtained by Group 2 [-EI][+TE SI] across tasks, it would appear that the means for production are perceptibly lower than those for picture matching and sentence interpretation. That being the case, it could be argued that perhaps the incremental learning advanced as an explanation for the higher means in Group 2 [-EI][+TE SI] OVS production at the immediate posttest is simply only noticeable in the task for which treatments were less directly beneficial. However, given that the pools of participants across tasks were not identical, this argument is questionable[26].

In fact, we see that for picture matching and sentence interpretation, Group 2 [-EI][+TE SI], far from increasing in the three weeks between posttreatment testing, saw their means diminish. Although both of these tasks showed a difference between Group 2 [-EI][+TE SI] and Group 1 [-EI][-TE SI] (the group whose mean logit of ability was closest to that of Group 2 [-EI][+TE SI]) at immediate posttest, the difference was no longer present after three weeks. Though this certainly does not strengthen our case with regards to the improvement exhibited by Group 2 [-EI][+TE SI], it remains an interesting finding in that it would appear to align with results from OVS production, where the lack of difference between Group 2 [-EI][+TE SI] and

[26] The n size for OVS tasks was 132 for production, and 130 for both Picture Matching and Sentence Interpretation. The pool of participants for production and picture matching differed by 12 people. For Production and Sentence Interpretation it differed by 15 people. Finally the pools for picture matching and Sentence Interpretation differed with respect to 10 participants. In all the cases the differences were spread across groups 1-5.

Group 1 [-EI][-TE SI] in both posttests led us to claim that gains from [+TE SI] only treatments appeared to be less robust than those obtained from [+EI] Groups. This may owe to the fact that recalling a rule derived from TE practice may be more difficult than recalling an externally provided rule, as in the latter case learners may be able to use explicit recollection of concepts in the EI to retrieve the rule.

7.2. Discussion of Hypotheses

In the following section we will discuss the results obtained from the analyses above in light of the hypotheses listed in Chapter 5. For that purpose, we repeat hypotheses 1-5 below:

H1: In the absence of TE and EI, PI (i.e., SI only in this case) will result in no benefits.

At the end of our literature review, we proposed that SI activities in previous research fit the description of TE and posited that such a condition was critical for the benefits of SI to obtain in the absence of EI. Findings from the present study lend full support to the claim that TE is indeed a crucial component for SI to result in gains. At the immediate posttest stage Group 1 [-EI][-TE SI], the experimental group that received SI only (i.e., no EI), and in non-TE conditions, performed in a manner comparable to that of the control group (where participants never received any EI or practice related to the target items of the study at all) in all six tasks implemented in the study. At the delayed posttest stage, this lack of difference was also present in the three OVS tasks (this was not the case for ser/estar, as gains on the copula distinction disappeared for all task). It was also the case that experimental treatments for the rest of the groups (Groups 2 [-EI][+TE SI], 3 [+EI] [-TE SI] and 4 [+EI][+TE SI]) resulted in scores significantly higher than those of Group 1 [-EI][-TE SI]. To be precise, at the immediate posttest stage [+EI] groups scored statistically higher than Group 1 [-EI][-TE SI] in five of the six tasks in the study (the exception being OVS production). At the delayed posttest level, those differences were maintained with respect to Group 3 [+EI] [-TE SI] in two of the three OVS tasks (picture matching and sentence interpretation), and with respect to Group 4 [+EI][+TE SI] in one of the OVS tasks (picture matching).

The effects of time aside, these findings are a solid indication of the fact that the treatment devoid of EI and TE was no more effective than no treatment at all. It is important to remember that practice items for participants in Group 1 [-EI][-TE SI], while non-TE, still always depicted *lo* and *la* as the theme and sentence-initial position, and ser and estar representing inherent and circumstantial conditions, respectively. In addition, as may be recalled, half of the practice items in Group 1 [-EI][-TE SI] drew on number agreement and therefore finding the critical information to perform successfully involved focusing on the presence or absence of the plural morpheme attached to the target forms of our study. Despite consistent exposure to the target items in a particular context and with a particular meaning, and despite the fact that practice environment promoted focusing on the target items themselves, participants in Group 1 [-EI][-TE SI] clearly did not make significant progress towards connecting either one of our linguistic targets to their respective meanings. On the one hand, these findings suggest that unless attentional resources are properly channeled, the combination of exposure and practice does result in substantial gains. On the other hand, and by the same token, the present results are indicative of the direct relationship that binds attention and acquisition, and they cast a vote in favor of instructional methods that effectively raise learner's awareness of the relationship between form and meaning.

Hypothesis 1 also implies that previous results reported in the PI literature, which concluded that EI did not offer any additional benefits for PI, would only obtain if practice conditions were TE. Findings from Group 1 [-EI][-TE SI] have a bearing on this claim to the extent that they suggest the acquisition resulting from TE-less SI practice with no explicit instruction either does not generate benefits of any substance. In addition, the statistical difference obtained across all tasks in both structures at the immediate posttest level between Groups 1 [-EI][-TE SI] and 3 [+EI][-TE SI] suggests that in the absence of TE, EI does indeed have a beneficial effect. The only difference between the treatments received by Groups 1 [-EI][-TE SI] and 3 [+EI][-TE SI] was that the latter received EI, to which participants had access a total of 5 times throughout the treatment sessions. We might add that the benefits obtained as a result of EI also gave evidence of having longer-term durability, at least for some, perhaps less difficult, types of processing problems, as shown by the fact that the significant superiority of Group 3 [+EI][-TE SI] over Group 1[-EI][-TE SI] still emerged at the delayed posttest stage for both of the receptive OVS tasks.

In all, then, in the absence of a TE practice environment, lack of EI resulted in no benefits regardless of the processing problem or task type whereas its provision resulted in statistically higher gains. It seems reasonable to believe that not having received external information as to the rules that govern the linguistic phenomena at hand, and given that the practice received was not conducive to deduction of such rules, learners in Group 1 [-EI][-TE SI] were unable to derive any knowledge from the treatment that would help them make principled decisions, which resulted in lack of learning. Our data also appear to suggest that the [-EI][-TE] treatment confused participants. This was shown by the fact that in both the sentence interpretation tasks and the OVS picture matching task, Group 1 [-EI][-TE] scored lower in the immediate posttest than in the pretest. In the case of OVS picture matching, the immediate posttest score was lower by 5 logits, which, may be argued is relatively mild. However, in the case of the sentence interpretation tasks the difference was of 14 (OVS) and 26 logits (ser/estar). The decrease in scores is worthy of note in that it suggests that in the absence of TE, not only is SI ineffective but it may very well even be counterproductive. By the same token, it seems to follow from our results that when the practice environment did not allow for rule deduction, learners relied on EI to solve the tasks, which in turn resulted in learning. Further arguments on this point will be offered later in this section, when findings with respect to hypothesis 3 are discussed.

Findings with regard to Hypothesis 1 have implications for conclusions advanced by VanPatten and Oikennon (1996) and others, which claimed that SI was the component of PI that accounted for benefits recorded for this approach. Given that there does not seem to be a consensus in the literature as to whether SI should by definition be TE (Sanz 2004 and Sanz & Morgan-Short 2004 appear to be the only authors who claim SI should be TE), and in view of present results, further qualification of previous claims with regards to the benefits offered by SI is in order.

Although our data do not in any way deny that SI without EI can generate benefits, they clearly indicate that it is only SI of a very particular kind, namely TE SI, that results in a beneficial effect.

H2: When practice is TE, EI will not be necessary for PI benefits to obtain. SI will suffice because it is TE.

Based on our results, TE practice was sufficient to make up for EI in all three ser/estar tasks but only obtained for the picture matching and sentence interpretation tasks in OVS. Thus, our data largely support Hypothesis 2 but suggest that TE sometimes fails to produce the results that EI obtained.

In terms of durability, at least as OVS is concerned, any significant advantage recorded for Group 2 [-EI][+TE SI] over Group 1 [-EI][-TE SI] at the immediate posttest vanished in all six tasks by the time the delayed posttest was administered. This was also true for ser/estar in as much as all the differences from the immediate posttest disappeared by the delayed posttest. That said, the TE only treatment was statistically superior to that of the control group in two of the three OVS tasks (picture matching and sentence interpretation) and one of the ser/estar tasks (production) in the immediate posttest. In the same vein, the gains obtained by the TE only group yielded superior scores to the control group in all three of the OVS tasks in the delayed posttest. Thus, Hypothesis 2 was partially supported in that EI was, depending on the task, not necessary for benefits to obtain. However, those benefits not always being on par with the ones generated by EI, it cannot easily be argued that when SI is TE, metalinguistic information is superfluous.

In all, this suggests that the effectiveness of TE SI varies depending on the processing problem posed by the target form. In our case, and contrary to previous findings from VanPatten & Cadierno (1993) and VanPatten & Oikkenon (1996), it appears that, for OVS sentences, the benefits derived from the TE SI did not transfer to task types beyond those featured in the practice. This exception aside, however, it is rather clear from the present results that when SI is TE, benefits ensue without the need for EI at least immediately after treatment.

In view of the findings we reported with respect to Group 1 [-EI][-TE SI], the effectiveness of TE SI invites speculation as to what exactly may have triggered learning in this condition. First of all, due to our design we can confidently track the source of learning in Group 2 [-EI][+TE SI] to the TE nature of the treatment In other words, it seems that learning was brought about when the number of possible meanings that could be attached to the target form was reduced by manipulating the context of the task on the basis of learners' default processing strategies. This was achieved, in part, by providing only two response options as possible interpretations for the prompt. Most crucially, one of the two interpretations featured the meaning that would result from the application of learners' default processing strategy. Feedback (a feature present in all groups), would then either reaffirm learners' initial interpretation, or act as a garden path-like effect, thus bringing attention to the fact that the adopted interpretation should be dismissed in favor of the other one. In other words, the manipulated environment combined with feedback allowed learners to map the form to its correct meaning by process of elimination. In addition, the treatment provided learners with further opportunities to put their newly acquired knowledge into practice enough times that a difference in performance was perceptible in exit tasks. While our results clearly indicate that learning occurred and while our design does allow us to draw some conclusion as to how that learning may have come about, yet another interesting question that naturally follows from these results and that will remain unanswered is whether this learning was explicit or implicit in nature. Though not necessarily central to the present study, this issue becomes of interest in light of the ongoing debate in our field concerning what type of mechanisms are involved in adult SLA. Participants in Group 2 [-EI][+TE SI] were never provided with EI, and yet they clearly improved from with respect to Group 1 [-TE SI][-EI] and Control, performing on par with the remaining experimental groups. This may prompt two premature conclusions: (1) that explicit knowledge of grammatical rules is therefore not essential for learning, as certain types of practice (i.e., TE practice) may guarantee similar enough results with the added advantage of circumventing the need for metalinguistic explanations; (2) that in the absence of EI, the learning that occurred must have been implicit in nature. Both of those conclusions would be unwarranted. On the one hand, from the description above, the possibility remains that what TE SI actually accomplishes is to provide an environment where learners can inductively figure out all by themselves what otherwise EI would overtly spell out for them (DeKeyser 2003). In that view, it would not be so much the case that knowledge of underlying rules is not essential (quite the opposite, in fact), but rather that such rules, under very particular circumstances, do not need to be externally provided. The fact that no EI was provided to learners in Group 2 [-EI][+TE SI] does not rule out the possibility that they formulated such rules on their own and in a conscious manner. As a matter of fact, in a setting where participants are put in the situation of solving a task, it is only reasonable to think that learners would have quite deliberately engaged in finding the rules necessary to perform successfully. Be that as it may, however, on the former point the present paper presents evidence to empirically support both that EI is indeed essential under certain practice conditions, as well as that its provision is beneficial for learning. On the latter point, results from this paper do not allow for a categorical conclusion one way or the other. Given that this study did not include any exit task to try and determine whether learning in Group 2 [-EI][+TE SI] had happened with or without awareness, nothing in our data allows us to deem whether the learning that resulted from TE SI is explicit or implicit.

Before moving on to discuss Hypothesis 3, the disparity with regard to OVS production results between the present study and previous ones deserves further discussion. As mentioned earlier, unlike in VanPatten & Cadierno (1993) and VanPatten & Oikkenon (1996), gains from SI (in this case operationalized as TE) in OVS, did not appear to transfer to task types not featured in the treatment, i.e., the production task. Production tasks in previous PI studies differed slightly from one another and from the one included in our study[27]. Both VanPatten and Cadierno (1993) and VanPatten and Oikkenon (1996) implemented production tasks that were

[27] We limit the present discussion with regard to previous production findings to studies featuring OVS structures as their target form. It should be noted, however, that skill transfer to production was also reported by Benati (2004), Farley (2004) and Wong (2004), with the Italian future tense, the Spanish subjunctive and the French negation particle *de*, respectively. All these studies implemented production outcome measures in the form of fill in the gap tasks. The number of items in the production task was 5, 9 and 6 respectively. In addition to the low number of observations in the production task, results Benati (2004) are hard to interpret in that the five items were blanks within the same paragraph, a fact that would have compromised the need for independent observations required by the analysis of variance he used to analyze his data.

more open-ended than the one we report on here. OVS production for both those studies involved completing the second part of a sentence based on visual cues and it involved providing both the object pronoun and the verb (VanPatten and Cadierno 1993, p. 233). One might initially think that the reason behind the disparity in results may originate in this difference between tasks, but this is not entirely intuitive as one might argue that a more controlled environment in the present study would have favored the chances of our participants. The present production task was, as mentioned earlier in the paper, as close to multiple-choice as a production task can get. The reader may recall that learners were given a choice between *lo* and *la* and asked to fill the space with the appropriate one to match the prompt image. Thus, though the task differed from picture matching and sentence interpretation, the level of production was minimal, and the lesser production burden our task posed should have resulted in greater likelihood that skills obtained from the treatment would transfer.

A possible explanation for the difference in results is that while the present study featured a production task comprising 20 target items, exit tasks measuring production in VanPatten & Cadierno (1993) and VanPatten & Oikkenon (1996) featured only 5 items. The low number of observations collected may have had an impact on the quality of the measurement.

In addition, the difference in production results may also be partially accounted for by the different analysis used to implement the data. While the present study analyzed all results, including production, by way of ANCOVA with pretest scores as the covariate, both VanPatten & Cadierno (1993) and VanPatten & Oikkenon (1996) resorted to repeated-measures ANOVA. The stricter control over the variable of prior knowledge that was achieved by factoring out pretest scores in the present study may have contributed to the disparity recorded in results.

All those considerations aside, findings regarding Hypothesis 2 appear to have implications for the interpretation of two previous studies on PI. First, our results largely coincide with Sanz & Morgan-Short (2004), in the sense that there appears to be solid indication in the outcomes of our analysis to suggest that "exposing learners to task-essential practice is sufficient to promote acquisition" (p. 35). Second, these findings also have implications for VanPatten and Oikkenon's (1996) claim that SI is sufficient for acquisition to occur. If SI gains can essentially be attributed to TE, then the somewhat inconsistent and less durable effects of TE gains in the present study beg for an important nuance: it appears that the benefits obtained from TE SI practice may not always be as solid as those obtained from the provision of EI. Thus, though TE SI may well be sufficient for benefits to ensue, this does not mean that gains from a TE treatment and gains from an EI treatment are qualitatively indistinguishable. These conclusions will be further justified by the findings with regard to hypothesis 3, which we report below.

H3: EI alone will result in greater gains than SI when SI is not TE.

Results from the present experiments with regard to the effect of EI are clear: Group 3 [+EI][-TE SI] obtained scores statistically higher than those obtained by Group 1 [-EI][-TE SI] for all six tasks, regardless of the structure. This hypothesis was also supported by the difference between the EI only group and Group 5

(Control), which obtained in five out of the six tasks implemented in the study (all except ser/estar sentence interpretation). As for the durability of such gains, Group 3 [+EI][-TE SI] maintained its superiority with respect to the control group in all three OVS tasks, while remaining superior to Group 1[-EI][-TE SI] in both the picture matching and sentence interpretation OVS. Thus, the superiority of EI over TEless SI stated in Hypothesis 3 was consistently supported across all six tasks at the immediate posttest while at the delayed posttest Hypothesis 3 was still partially supported.

It should be added that according to our results, together with the other [+EI] group (Group 4[+EI][+TE SI], a close second), Group 3 [+EI] [-TE SI] emerged as the most consistent and most durable of our conditions. These findings lend further support to the claim that EI does indeed have a meaningful role in SLA, as it would appear that unless practice is TE, gains do not obtain without it. These findings acquire further significance upon considering that, depending on the target form, incorporating TE in practice tasks may prove challenging. This finding also represents a clear departure from the great majority of PI studies. With the exception of Farley (2004) and Fernández (2008), both of which reported EI appeared to be beneficial in the acquisition of the Spanish subjunctive, PI studies so far had consistently found the role of EI to be superfluous if practice was SI. Based on our results, the claim that EI is unnecessary would hold only for some structures and provided that SI practice is TE. The present results have to interpreted bearing in mind that while previous PI studies maintained TE constant across all conditions, our positive effect for EI emerges in the absence of TE. Thus, once again, while our findings do not in any way bear on the claim that SI can be effective, they do suggest that not all practice that fits the current definition of SI is able to offset the absence of EI. In addition to the nature of the practice, a factor that may have contributed to the present results relates to the way in which EI was administered in previous studies. As mentioned in passing earlier in the present paper, to the best of our knowledge, none of the PI studies conducted thus far allowed participants access to EI more than once; the only contact with EI took place prior to administration of the treatment. In addition, in previous studies, experimenters would often read the EI to participants, (which would obviously have an impact on the level of retention of the information), or they would ask learners to read the EI themselves but then fail to control for whether such directions had been followed or not. While the functionality of the software used in the present experiment did not allow for verification that EI had been properly attended to either, learners were warned at the beginning of each EI slide that (a) the information they were about to read was necessary to perform successfully in the tasks to follow and (b) that after each EI slide they would be tested on the information they had just read.

The unnaturally limited access to EI and the lack of control over attention may very well have been fully accountable for its consistent lack of effect in this literature (along with having TE practice as an alternative source of information). Exposing learners to EI just once, often auditorily only, and before learners had any inkling of what they may want to pay attention to, seems neither ecologically valid nor experimentally imperative. This study recognizes this as an effort to control how long each learner was exposed to EI, and yet, it contends that such control assumes learners may dwell on EI so long as to skew results, which would seem less of a concern than, for example, controlling that EI has been attended to at all.

These findings represent an additional layer to our results regarding Hypothesis 1 and Hypothesis 2. Our results indicate that when SI is not TE, EI turned out to be beneficial across all tasks and for both structures, an effect that was rather robust over time. In view of our findings from Group 2 [-EI][+TE SI], we may also add that the EI treatment received by Group 3 [+EI][-TE SI] appeared to be more effective than the TE treatment received by Group 2 [-EI][+TE SI]. We base this conclusion on three facts: First, OVS production results at the immediate posttest stage revealed a significant difference between Group 2 [-EI][+TE SI] and Group 3[+EI][-TE SI], indicating that the level reached by the latter group was significantly higher. Second, the benefits of the TE treatment manifested themselves less consistently than those from the EI treatment, as shown by the lack of significance between Groups 1 [-EI][-TE SI] and 2 [-EI][+TE SI] in OVS production and the statistical significance between Groups 1 [-EI][-TE SI] and 3 [+EI][-TE SI]. Finally, the durability of the results obtained from the EI treatment appeared to be more robust over time than those obtained from the TE treatment. Evidence for this is found in the fact that while any significant advantage recorded for Group 2 [-EI][+TE SI] over Group 1 [-EI][-TE SI] at the immediate posttest vanished by the time the delayed posttest was administered, the significant advantage obtained by the [+EI] only treatment of Group 3 [+EI][-TE SI] remained unaffected for OVS picture matching and sentence interpretation. Thus, our results appear to indicate that even when SI was TE, as it was for Group 2 [-EI][+TE SI], the gains obtained appear to be inferior to those obtained by EI alone on a number of counts. Such findings run counter to Sanz and Morgan-Short's (2004) claim suggesting that "explicit information may not necessarily facilitate second language acquisition and [that] exposing learners to task-essential practice is sufficient to promote acquisition." (p. 35).

Thus, in addition to advancing that not all types of SI can dispense with EI, we also contend that even when SI is TE, gains obtained from EI are both more robust and more durable. We argue that these results owe to the repeated exposures participants had to the metalinguistic information, as well as to the implementation of follow-up questions after provision of rules.

H4: Under TE conditions, EI will result in greater gains than TE alone.

The combination of EI and TE gave clear evidence of having a facilitative effect for learning, in the sense that Group 4 [+EI][+TE SI], where participants received EI before and during the TE practice treatment, was consistently superior to Group 1 [-EI][-TE SI] and the Group 5 (Control). It was only in two out of the six tasks and only at the immediate posttest stage, however, that complementing TE practice with EI resulted in ability scores that were statistically higher than those obtained through TE alone. More specifically, it appears from our results that EI offered additional benefits to TE practice in OVS production, as well as in ser/estar sentence interpretation at the immediate posttest stage, (as shown by the significant difference recorded between Group 2 [-EI][+TE SI] and Group 4 [+EI][+TE SI] in both of those tasks). Though this effect was far from obtaining across the board, then, Hypothesis 4 was partially supported. Of note is that according to our data, the positive effect of additional EI provision was perceptible for both types of processing problems represented in this study. We hasten to add that this our claim in favor of EI should not be confused with the endorsement of metalinguistic information as the only or even the main source of input language learners require. All of our conditions involved some sort of SI, in other words, practice. Instead, and based on our data, the contention of this paper is simply that, contrary to what previous studies have concluded, metalinguistic information is of value to learners and should therefore not be shunned in classrooms. Opponents of metalinguistic information often argue that the gains originating from treatments relying on EI result in declarative knowledge, which, although captured in outcome measures like the ones used in this study, is not indicative of language learning. Our claim in favor of EI provision is on the belief that knowledge derived from EI-based treatments can be and often is a necessary precursor for any other more automatized knowledge to ever be possible. It is within that framework that we suggest EI as a necessary and beneficial resource that should not be excluded from classrooms.

The reason why this facilitation occurred in production for OVS, as well as in sentence interpretation for ser/estar, but not for the rest of the tasks, is unclear. If nothing else, however, this fact constitutes further evidence that different task-types and processing problems necessitate different approaches and combinations in order for learners to thrive.

The gains derived from the addition of EI to TE, comparing Group 2 [-EI][+TE SI] and Group 4 [+EI][+TE SI], showed a very similar level of robustness to time. Like Group 2 [-EI][+TE SI], at the delayed posttest, Group 4 [+EI][+TE SI] lost the superiority it had over Group 1 [-EI][-TE SI] in production and sentence interpretation OVS. Unlike in the TE alone condition, however, in Group 4 [+EI][+TE SI] scores for picture matching were still statistically higher than those for Group 1 [-EI][-TE SI] at the time the delayed posttest was administered. In all, then, in addition to sometimes generating significantly greater gains than TE alone, it appears that benefits resulting from the combination of EI and TE may also be slightly longer-lasting than those obtained from TE alone treatments.

To the best of our knowledge, our findings in support of EI as an additional beneficial factor when practice is TE, run counter to all previous studies comparing PI to SI. In addition, our results with regard to Hypothesis 4 are also at odds with claims from Sanz & Morgan-Short (2004), who advance that "when learners are asked to complete a task in which they are presented with structured input and task-essential practice items, supplementary information about the language form provided a priori does not enhance their ability to use the form in subsequent interpretation or production measures" (p. 69). We propose, once again, that the more ecologically valid manner in which EI was administered in the present paper may explain the difference between our results and previously reported ones.

A final item worthy of note in relation to the effects of the [+EI] [+TE SI] condition, is that its effects both at the immediate and delayed posttest stages were virtually identical to those of Group 3 [+EI][-TE SI], where learners received EI but no TE practice. Thus, though complementing TE practice with EI gave evidence of resulting in significantly higher gains for some of the tasks in this study, it seems from our results that complementing EI with TE practice may be just as efficient as provision of metalinguistic information alone.

H5: Benefits obtained from [+TE SI] only practice will prevail across time.

Our prediction was that benefits generated by TE practice would, as had been the case in previous studies, endure across time. Indeed, according to our results this was true in as much as Group 2 [-EI][+TE SI] maintained its superiority with respect to the control group (Group 5) in all three OVS tasks. However, over the course of three weeks benefits obtained by the TE only condition with respect to Group 1 [-EI][-TE SI] vanished. This was in fact the general tendency shown by all the experimental groups that experienced significant gains immediately after treatment. Namely, the trend was for gains to wane down to the point where they were no longer statistically different from Group 1[-EI][-TE SI] (which, as the reader may remember, always performed on par with the Group 5), but to be maintained at a level still significantly higher than those obtained by the control group. Still, Group 2 [-EI][+TE SI] exhibited gains that appeared to be more prone to decay over time than the two [+EI] conditions. Whereas Group 2 [-EI][+TE SI] lost its difference with respect to Group 1 [-EI] [-TE SI] in all tasks, Group 3 [+EI][-TE SI] maintained them for OVS picture matching and sentence interpretation, and Group 4 [+EI][+TE SI] maintained it for OVS picture matching. The reason behind this difference in durability could originate in the fact that recalling knowledge obtained from the TE treatment may be more difficult, because learners find it hard to remember such knowledge without a terminological hook to hang it on. In this view, while learners in the [+EI] conditions may be able to retrieve a terminology item that might in turn be transparent enough to help them recall the entire rule, learners in Group 2 [-EI][+TE SI] would likely need to recall the entire rule itself, without relying on linguistic concepts that may help jog their memory.

Finally, we may add that the impact on durability seems to be a function of the processing problem as well as the treatment, as shown by the more severe losses exhibited by gains recorded for ser/estar. Our findings suggest that Hypothesis 5 was partially supported in cases were automaticity may be at the core of the issue. On the contrary, when the problem to be overcome is conceptual (ser/estar), Hypothesis 5 was not supported at all.

7.3. Conclusion and Future Directions

The primary goal of this paper was to gauge the role that TE played in the learning obtained from SI practice, and how its presence or absence affected the need for EI provision. In this respect all our results coincide in indicating that TE is a crucial component of SI. In the case of both ser/estar and OVS and regardless of the task, this study showed that in order for SI to be effective, learners had to engage in TE practice. On the one hand, this manifested itself in the fact that non-TE SI practice without EI consistently resulted in performance comparable to that of participants who did not receive any practice on OVS or ser/estar at all. On the other hand, evidence for the importance of TE was found in the fact that non-TE SI consistently resulted in levels of ability inferior to those exhibited by participants who received TE SI. As for the reasons behind the superior performance of the group receiving TE, we argue that the virtue of TE lies in its capacity to direct learner attention to the target item by constraining the number of possible interpretations the target form could take, which in turn results in an increased likelihood that a form-meaning mapping will be created.

Our results concerning the fundamental role played by TE are important for two reasons: First, they represent a clear indication of the extent to which SLA rests on attention to form-meaning mappings, a fact advanced previously by several studies in the SLA literature (Rosa & O'Neill 1999 and Rosa & Leow 2004). Second, these findings are of value in that they provide critical information for proper qualification of claims that were previously made and which have had direct implications for instructed SLA. Namely, previous PI studies concluded that SI, defined as practice that has been manipulated to put a grammatical item in a privileged position (VanPatten 1993 and Lee & VanPatten 1995) such that learners are pushed to process it (Lee & VanPatten 1995), was an effective means of addressing certain learnability problems, such as OVS structures (VanPatten & Oikkennon 1996 and others). Our results suggest that putting the grammatical item in question in a privileged position and pushing for it to be processed are certainly necessary but not always sufficient conditions for learning to occur. Findings from this paper indicate that SI must go even further and comprise practice manipulated in such a way that processing is not only pushed for, but rather made essential for successful completion of the task. In other words, only SI in which "the task cannot be successfully performed unless the structure is used" (Loschky & Bley-Vroman 1993, p. 32) results in form-meaning mapping and learning. In what pertains to language instruction, then, results from this paper appear to suggest that TE practice should be incorporated in classrooms, particularly in order to achieve noticing of forms that tend to go unlearned due to their low salience. Although the question as to whether TE practice can be devised for certain linguistic targets looms large, making instructors aware of the potential that TE offers would seem both reasonable and necessary.

In the same vein, the present paper found that claims previously made in the PI literature pertaining to the role of EI in learning were also in need of further qualification in light of our findings with regard to SI. While numerous studies have reported a positive effect of EI on learning (e.g., Alanen 1995; DeKeyser 1995; De Graaff 1997; Ellis 1993; and Robinson 1996, 1997), with few exceptions, PI studies have commonly found that when practice is SI, EI is not necessary and it does not offer any additional benefit. In this regard, our results show that EI can be dispensed with only as long as SI is TE but can even be beneficial if added to certain TE practice. Thus, our findings indicate that when the practice environment is constrained enough for learners to deduce the form-meaning connection and the rule governing the target item on their own, external provision of metalinguistic information is not necessary for learning to occur. This nuance becomes particularly important in view of the fact that due to the difficulty that is often involved in creating TE conditions, learners may not always receive TE practice. Thus, as far as instructed SLA is concerned, EI should not be considered in the least moot, but rather it should be perceived as a fundamental element particularly when practice does not meet TE conditions. Most crucially, however, our findings also revealed that depending on the processing problem and the type of task, EI can sometimes produce benefits superior to those obtained by TE practice alone. This is critical evidence not only in as much as it further defines the core and the limitations of SI, but most importantly because it dispels the notion that EI holds no additional value for instructed learners. Our results clearly indicate that depending on the nature of the learnability problem, learners do benefit from additional support to complement practice.

In fact, in all, this paper found EI to be a more reliable source of benefits than TE. Though our data revealed that TE clearly had a facilitative effect, there was also some indication that such an effect may be less likely to emerge in tasks that learners did not encounter during the treatment. In contrast, the beneficial effects of EI were perceptible without exception for previously encountered and novel tasks alike. To this we should add the fact that the durability of the derived gains proved to be more robust when they stemmed from an EI treatment, as opposed to the TE one. We argue that this could originate in the fact that certain information may be difficult for learners to conceptualize, which may make retrieval and recall more difficult. In the same way, this paper also contends that the difference in results between this and other PI studies may lie in the way EI was administered. Based on our present results it is possible that the lack of EI effect reported in previous studies owed to the fact that learners had a single exposure to the metalinguistic information. Thus, EI does indeed have a facilitative effect but it has to be available several times throughout the practice session. Findings around EI that resulted from this study have direct implications for instructed second language acquisition. Perhaps still as a reaction to methodologies that relied exclusively on EI as a source of input at the expense of communication, it is very common to find language teaching professionals (whether instructors, teacher trainers or publishers) who strive to teach second languages completely forgoing metalinguistic information. Furthermore, it is often the case that that provision of rules is perceived as a malfeasance on the part of the instructor. Results from this paper, however, clearly indicate that EI is helpful to learners and make a case for its proper incorporation in classrooms. Undoubtedly, it is a well-known fact that learners, amongst other things, need ample opportunities for meaningful communication in the target language in order to become functional speakers. Thus, our intention is not to suggest that we should revert back to the Grammar Translation Method or that EI provision will result in fluent, pragmatically perfect second language learners. However, this study shows depriving learners of EI, especially in cases where practice conditions do not allow for rule deduction, is doing them a disservice. In the same way, it also shows that learners who received EI perform better than those who do not. Being able to perform in the outcome measures of a study differs substantially from being a full-fledged successful second language speaker. However, performing in such a controlled environment, we argue, does require proceduralizing the declarative knowledge obtained by way of EI. And that, presumably, is the first step in a longer process during which, by way of practice, automatization and native-like ability could be reached. Furthermore, it should be clarified that advancing EI as a useful tool in language instruction is not a claim against implicit knowledge, but rather a claim against the insistence for exclusion of metalanguage often espoused by proponents of implicit teaching methods. In a nutshell, the implications of our findings with respect to second language instruction are perfectly captured by the proverbial urge not to throw the baby out with the bath water. Based on our evidence, EI should be restored to the language classroom as yet another resource for instructors to exploit.

Finally, this paper also sought to determine whether the beneficial effects previously recorded for SI would obtain upon using this type of treatment on a different processing problem, such as that represented by the ser/estar distinction. In concert with results obtained from Cheng (2002), our results showed that TE SI condition did indeed have a facilitative effect for copula choice, and that its benefits were largely on par with those generated by treatments including EI. Added to previous results with the subjunctive in Spanish, the Italian future tense as well as the French negative particle *de*, these results appear to make a rather solid case for the effectiveness of SI as an instructional approach, suggesting that the benefits recorded for it so far are likely to obtain for learnability problems beyond those features in previous studies, at least when task-essentialness of the structure being practice is assured.

Appendices

Appendix A – EI for SER/ESTAR and OVS ([+EI] conditions)

***ser/estar* – EI (Slide 1 of 2) administered immediately before the treatment and after practice items 1, 4, 8 and 16)**

PLEASE READ THIS INFORMATION TO BE ABLE TO COMPLETE THE TAKS IN THIS EXPERIMENT.

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.

A) Do not click on button A. Click on Next Question directly to proceed to a question about this information.

***ser/estar* – EI (Slide 2 of 2) administered immediately before the treatment and after practice items 1, 4, 8 and 16)**

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.

A) Do not click on button A. Click on Next Question directly to proceed to a question about this information.

OVS – EI (Slide 1 of 2) administered immediately before the treatment and after practice items 1, 4, 8 and 16)

PLEASE READ THIS INFORMATION. There is a question on it in the next slide.

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.

Please click on NEXT QUESTION to continue to the question on the information above.

A) ignore this button

OVS – EI (Slide 2 of 2) administered immediately before the treatment, and after practice items 1, 4, 8 and 16)

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.

Click NEXT QUESTION to answer a question on this information

A) Ignore this button

Appendix B – Questions on EI ([+EI] conditions)

***ser/estar* – Question on EI Slide 1 administered after practice item 1**

English TO BE has two equivalents in Spanish, SER and ESTAR. SER is used for inherent, built-in traits.

A) True

B) False

***ser/estar* – Question on EI Slide 2 administered after practice item 1**

SER and ESTAR can be used interchangeably.

A) TRUE

B) FALSE

***ser/estar* – Question on EI Slide 1 administered after practice item 4**

We decided not to ask a question after this item.

A) Do not click on button A. Click on Next Question directly to proceed to a question about this information.

***ser/estar* – Question on EI Slide 2 administered after practice item 4**

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

***ser/estar* – Question on EI Slide 1 administered after practice item 8**

'El hombre <u>ESTÁ</u> serio' means that being serious is part of the man's personality.

A) TRUE

B) FALSE

***ser/estar* – Question on EI Slide 2 administered after practice item 8**

'El hombre <u>ESTÁ</u> 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

***ser/estar* – Question on EI Slide 1 administered after practice item 16**

'Inherent' and 'circumstantial' mean the same thing.

A) TRUE

B) FALSE

***ser/estar* – Question on EI Slide 2 administered after practice item 16**

We decided not to ask a question after this item.

A) Do not click on button A. Click on Next Question directly to proceed to a question about this information.

OVS – Question on EI Slide 1 administered after practice item 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

OVS – Question on EI Slide 2 administered after practice item 1

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

OVS – Question on EI Slide 1 administered after practice item 4

'Lo admira Sonia' means 'He admires Sonia'

A) TRUE

B) FALSE

OVS – Question on EI Slide 2 administered after practice item 4

False alarm! No question!

A) Ignore this button

OVS – Question on EI Slide 1 administered after practice item 8

'Lo admira Sonia' means 'Sonia admires him'

A) TRUE

B) FALSE

OVS – Question on EI Slide 2 administered after practice item 8

In 'LA inspira Raúl' the first word 'LA' designates the recipient of the inspiration

A) TRUE

B) FALSE

OVS – Question on EI Slide 2 administered after practice item 16

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

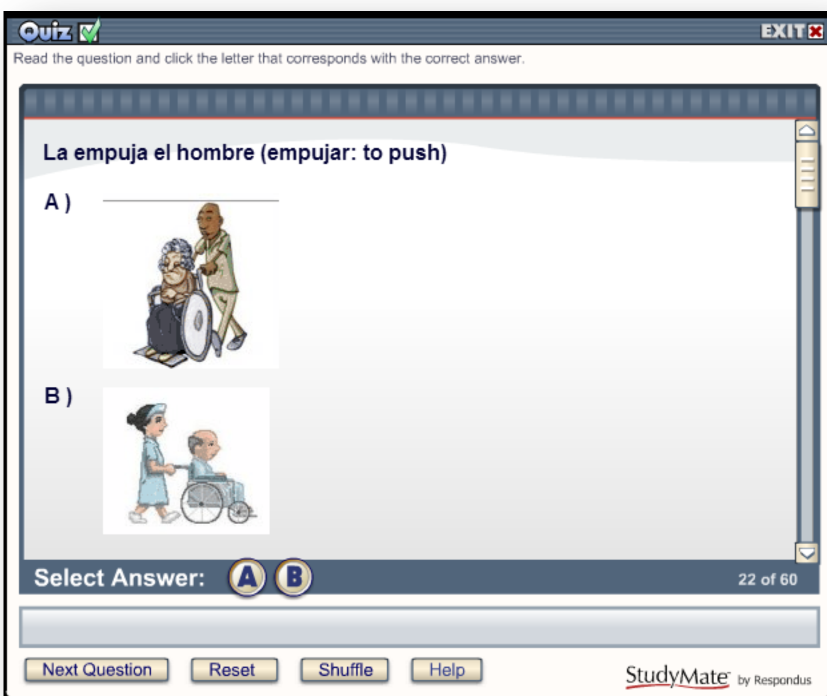
B) FALSE

OVS – Question on EI Slide 2 administered after practice item 16

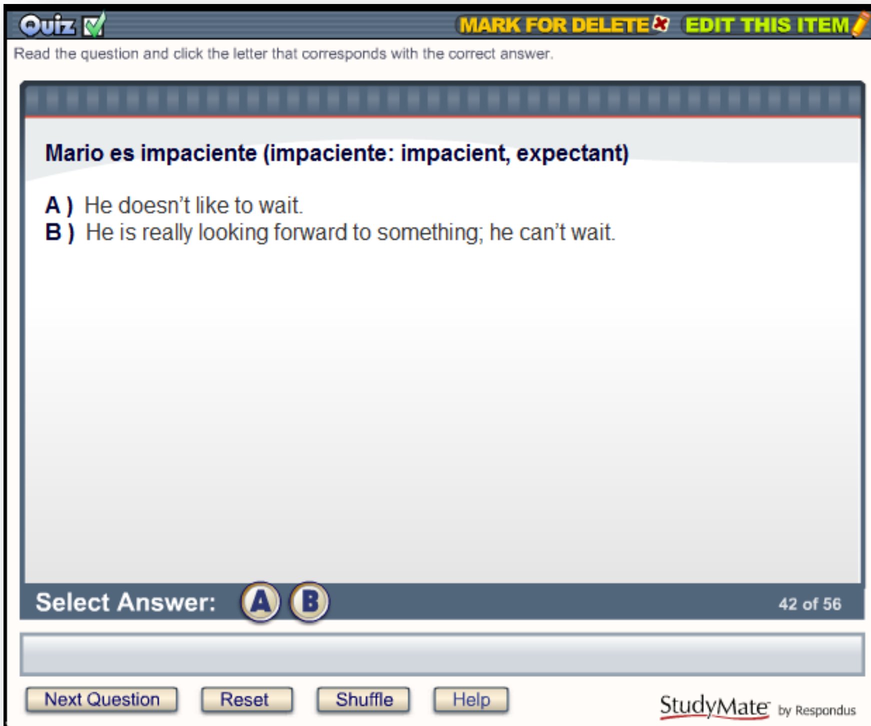
False alarm! No question!

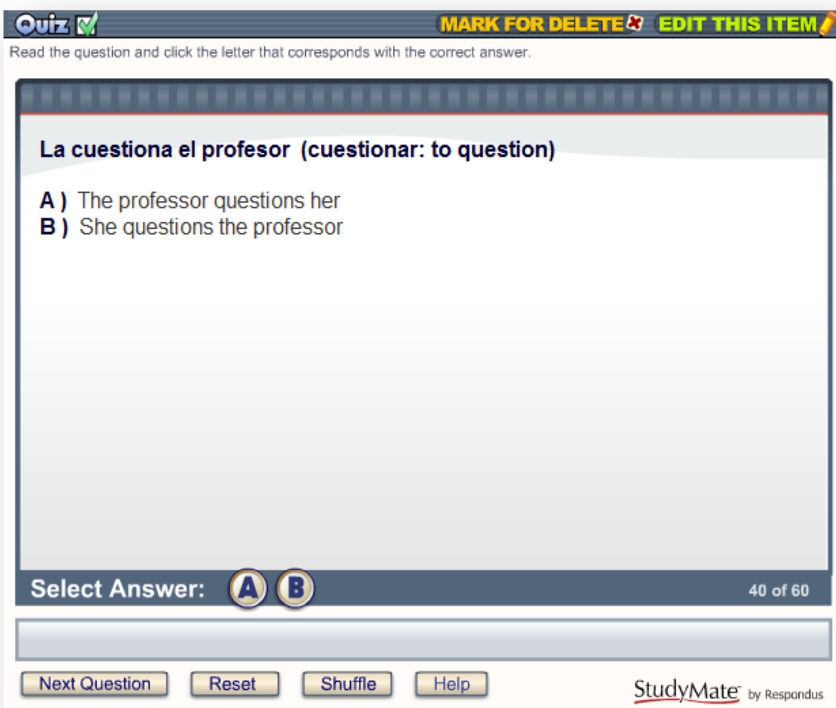
A) Ignore this button

Appendix C – Target items for [+TE SI] condition

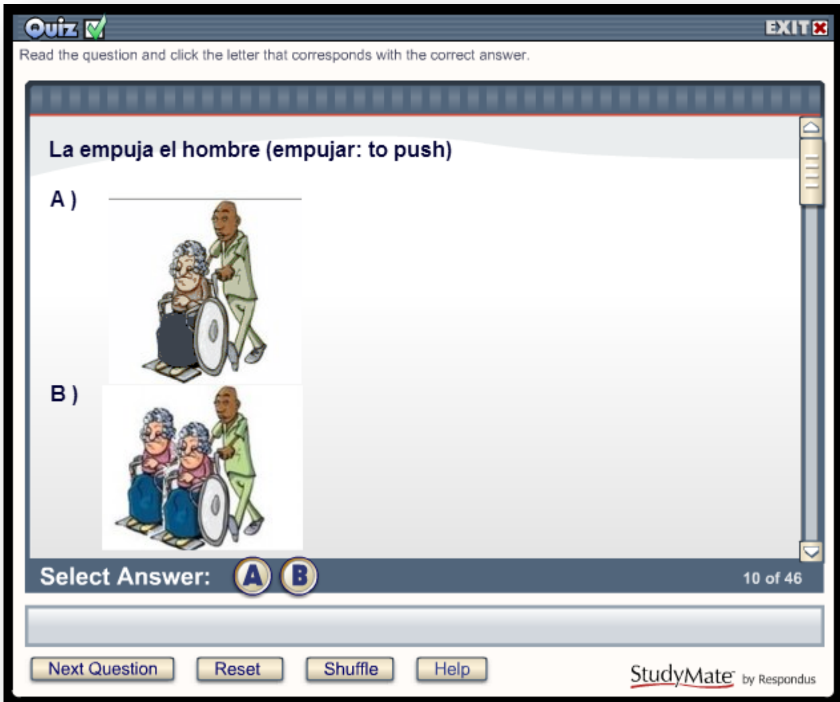


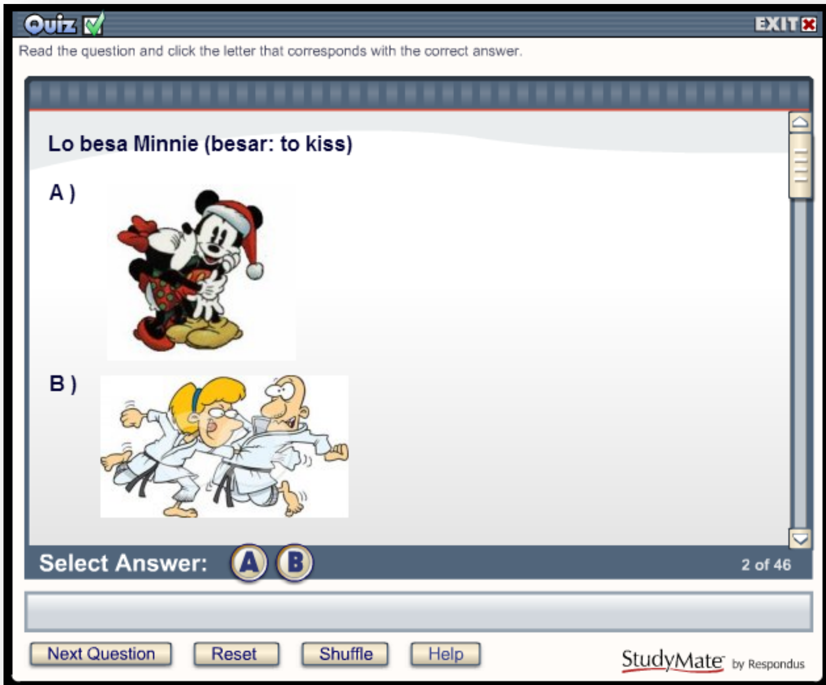

Quiz
EXIT
Read the question and click the letter that corresponds with the correct answer.
Está elegante (elegante: elegant)
A)
B)
Select Answer: A B
20 of 40
Next Question
Reset
Shuffle
Help
StudyMate by Respondus

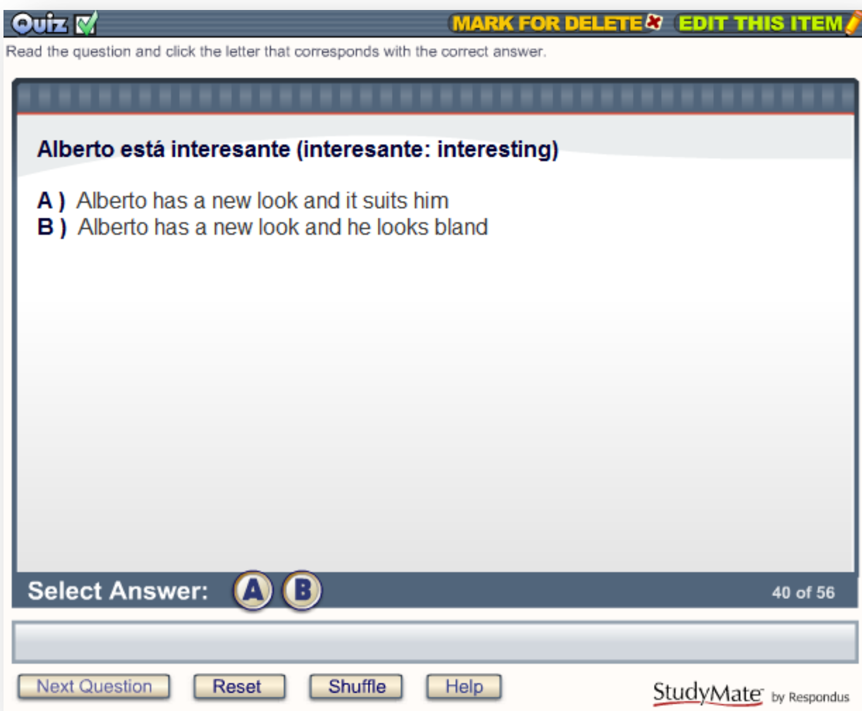

Quiz
MARK FOR DELETE
EDIT THIS ITEM
Read the question and click the letter that corresponds with the correct answer.
Mario es impaciente (impaciente: impacient, expectant)
A) He doesn't like to wait.
B) He is really looking forward to something; he can't wait.
Select Answer: A B
42 of 56
Next Question
Reset
Shuffle
Help
StudyMate by Respondus

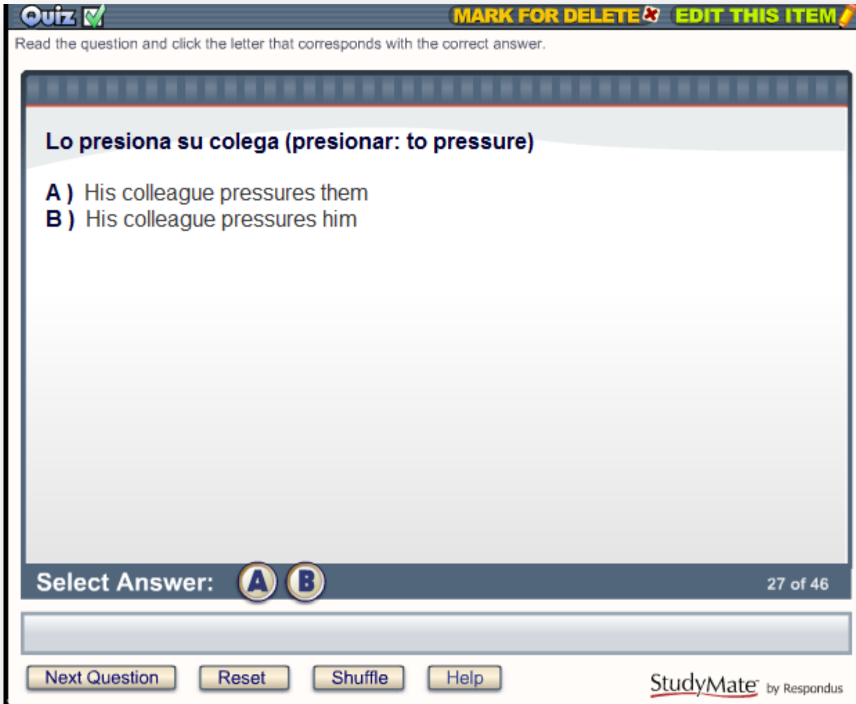

Quiz
MARK FOR DELETE
EDIT THIS ITEM
Read the question and click the letter that corresponds with the correct answer.
La cuestiona el profesor (cuestionar: to question)
A) The professor questions her
B) She questions the professor
Select Answer: A B
40 of 60
Next Question
Reset
Shuffle
Help
StudyMate by Respondus

Appendix D – Target items for [– TE] condition

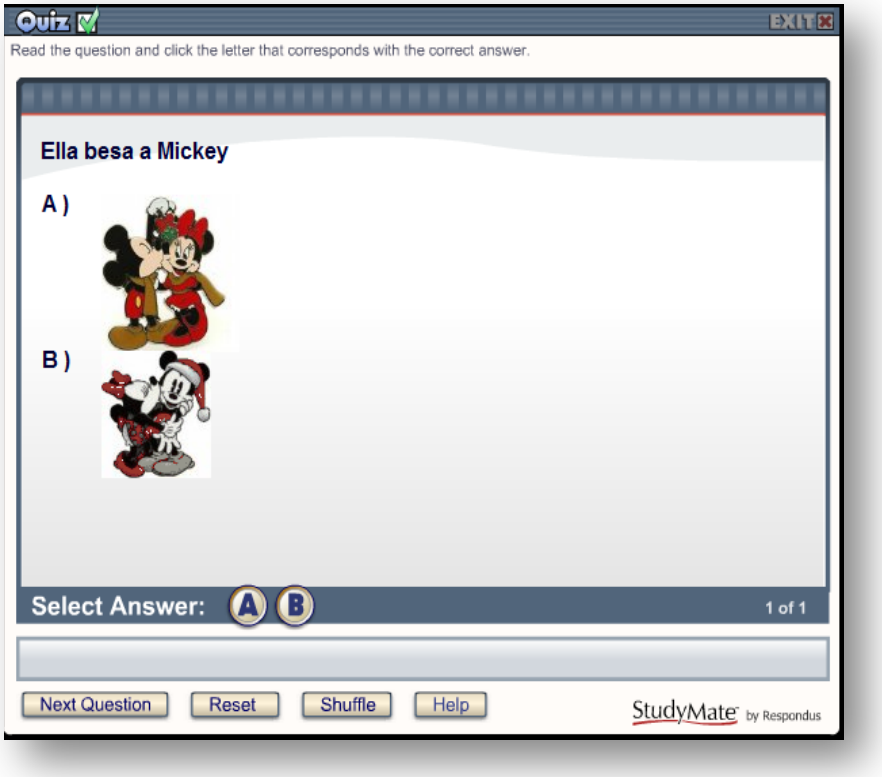



Quiz
EXIT
Read the question and click the letter that corresponds with the correct answer.
Lo besa Minnie (besar: to kiss)
A)
B)
Select Answer: A B
2 of 46
Next Question
Reset
Shuffle
Help
StudyMate by Respondus


Quiz
MARK FOR DELETE
EDIT THIS ITEM
Read the question and click the letter that corresponds with the correct answer.
Alberto está interesante (interesante: interesting)
A) Alberto has a new look and it suits him
B) Alberto has a new look and he looks bland
Select Answer: A B
40 of 56
Next Question
Reset
Shuffle
Help
StudyMate by Respondus


Quiz
MARK FOR DELETE
EDIT THIS ITEM
Read the question and click the letter that corresponds with the correct answer.
Lo presiona su colega (presionar: to pressure)
A) His colleague pressures them
B) His colleague pressures him
Select Answer:
A
B
27 of 46
Next Question
Reset
Shuffle
Help
StudyMate by Respondus

Appendix E – Distracters for OVS Treatment and Testing Instruments



7	Morado (morado: purple)	
	a. Suéter 	a. Radio 

39	Fill in the blank with **rojo** or **roja**
	Pimiento _______. (Pimiento: Pepper) 

Please pick the option that matches the sentence in bold.

1- Este periodista escribe artículos (escribir: to write)

a. This sentence refers to a woman b. This sentence refers to a man

Appendix F – Exit Task production

26	Fill in the gap with es or está
	____ limpio. (limpio: clean)

12	Fill in the gap with **lo** or **la**
	____ pisa. (pisar: to step on)

Look at the following sentences and fill in the gap with *lo* or *la*

- 1- _____ besa.

Look at the following sentences and fill in the gap with *es* or *está*

1- El hombre _____ guapo

Appendix G – Language Background Questionnaire

Survey of Participant Background and Prior Language Experience

Directions: Please answer each question to the best of your ability. Check or write-in your answers where indicated.

Confidentiality: This information will be kept strictly confidential. Your name or other identifying information will not be attached to this survey.

Biographical Information

1. Age: _____

2. Sex: _____ Male ______ Female

Language Learning History

3. What is your first or native language? _____________________________

4. Do you speak any languages other than English? Yes / No

If yes, please list the languages you know here:

____

5. If you speak any other languages, please state how you learned them (school, learned at home from parents or family, etc.)

____

6. Have you ever studied a foreign language before taking this course? Yes / No

If yes, please state what language and the number of years you took classes.

____

7. Have you ever been to a Spanish speaking country? Yes / No

If yes, please state the name of the country and the duration of the stay.

____

8. Have you ever lived in a country where a foreign language was spoken for an extended period of time? Yes / No

If yes, please name the country and the languages you were exposed to there.

____

Bibliography

Alanen, R. (1995) Input enhancement and rule presentation in second language acquisition. In R.Schmidt (Ed.), *Attention and awareness in foreign language learning and teaching* (Tech. Rep. No. 9; pp (259–302) Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.

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

Benati , A. (2001). A comparative study of the effects of processing instruction and outputbased instruction on the acquisition of the Italian future tense. *Language Teaching Research*, *5*, 95 – 127.

Benati, A. (2004) The effects of structured input activities and explicit information on the acquisition of the Italian future tense. In B. VanPatten (Ed.), *Processing Instruction: Theory, research, and commentary* (pp. 207–225) Mahwah, NJ: Erlbaum.

Berkovits, I., Hancock, G., & Nevvitt, J. (2000) Bootstrap Resampling Approaches for Repeated Measures Designs: Relative Robustness to Sphericity and Normality Violations. *Educational and Psychological Measurement, 60*, 877-892.

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

Briscoe, G. (1995). The acquisition of *ser* and *estar* by nonnative speakers of Spanish. Ph.D. dissertation, University of Pennsylvania.

Cheng, A. C. (2002). The Effects of Processing Instruction on the Acquisition of Ser and Estar. *Hispania*, 85, 308-323.

Cheng, A. C. (2004) Processing Instruction and Semantic Ser and Estar: Forms with Semantic Aspectual Values. In B. VanPatten (Ed.), *Processing instruction: Theory, research,and commentary* (pp. 119–141) Mahwah, NJ: Erlbaum.

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on language acquisition. *Studies in Second Language Acquisition*, 19, 249–276.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, 17, 379–410.

DeKeyser, R. (2003). Implicit and Explicit Learning. In M. Long and C. Doughty (Eds.), *The Handbook of Second Language Acquisition* (pp. 313-348). New York: Blackwell.

Ellis, N. (1993). Rule and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, 5, 289–318.

Farley, A. (2004) 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.

Gass, S. M. (1989). How do learners resolve linguistic conflicts? In S. Gass & J. Schacter (Eds.) *Linguistic perspectives on second language acquisition* (pp. 183-199). Cambridge: Cambridge University Press.

Geeslin, Kimberly L. (2000). "A New Approach to the Second Language Acquisition of Copula Choice in Spanish." *Spanish Applied Linguistics at the Turn of the Millennium: Papers from the 1999 Conference on the L1and L2 Acquisition of Spanish and Portuguese*. Ed. Ronald P. Leow and Cristina Sanz. Somerville: Cascadilla. 50–66.

Guntermann, G. (1992). An analysis of interlanguage development over time, part II: *Ser* and *estar*. *Hispania*, 75, 1294–1303.

Henry, N., Culman, H., & VanPatten, B. (2009). More on the Effects of Explicit Information in SLA. *Studies in Second Language Acquisition*, 31, 559-575.

Ho, R. (2006) *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. Taylor Francis Group. Hettmansperger, T.P., McKean, J. (2012) Robust Non-parametric Statistical Methods. (2nd Ed.) Taylor Francis Group.

Huitema, B. (2012) *The Analysis of Covariance and Alternatives*. New York, NY: John Wiley & Sons.

Lee, J. F., & VanPatten, B. (1995). *Making communicative language teaching happen*. San Francisco: McGraw-Hill

LoCoco, V. (1987). Learner comprehension of oral and written sentences in German and Spanish: The importance of word order. In B. VanPatten, T. R. Dovrak, & J. F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 116–129). Rowley, MA: Newbury House.

Lomax, R. G. (2007). *An introduction to statistical concepts for education and behavioral sciences* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Lomax, R. G. (2007). *Statistical concepts: A second course for education and the behavioral sciences* (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based learning. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, England: Multilingual Matters.

Paulston, C. (1972). Structural pattern drills: A classification . In H. Allen & R. Campbell (Eds.), *Teaching English as a second language* (pp. 129 – 138). New York : McGraw- Hill .

Robinson, P. (1996) Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition*, *18*, 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, 223–247.

Rosa, E., & Leow, R. (2004). *Awareness, Different Learning Conditions and Second Language Development*. Applied Psycholinguistics, 25, 269-292.

Rosa, E., & O'Neill, M. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition*, 21, 511–556.

Ryan, J., & Lafford, B. (1992). The acquisition of lexical meaning in a study abroad environment: *Ser* + *estar* and the Granada experience. *Hispania*, 75, 714–722.

Sanz, C. (2004). Computer Delivered Implicit Versus Explicit Feedback in Processing Instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 241–257) Mahwah, NJ: Erlbaum.

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.

Silvá-Corvalán, C. (1986). Bilingualism and language change: The extension of estar in Los Angeles Spanish, *Language*, 62, 587-608

Silvá-Corvalán, C. (1994) *Language contact and change: Spanish in Los Angeles.* Oxford. Clarendon Press.

VanPatten, B. (1984). Learners' comprehension of clitic pronouns: More evidence for a word order strategy. *Hispanic Linguistics*, 1, 57–67.

VanPatten, B. (1993) Grammar teaching for the acquisition-rich classroom. *Foreign Language Annals*, *26*, 435–450.

VanPatten, B. (1996) *Input processing and grammar instruction: Theory and research*. Westport, CT: Ablex.

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

VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5–31) Mahwah, NJ: Erlbaum.

VanPatten, B. (2007). Input processing in adult second language acquisition . In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 115 – 135). Mahwah, NJ : Erlbaum .

VanPatten, B., & Cadierno, T. (1993) Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*, 225–244.

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

VanPatten, B. (2010). Some Verbs Are More Perfect than Others: Why Learners Have Difficulty with ser and estar and What It Means for Instruction. *Hispania, 91,* 29-38.

Wilcox, R. (2005). An Approach to Ancova that Allows Multiple Covariates, Nonlinearity, and Heteroscedasticity. *Educational Psychology and Measurement, 65*, 442-450.

Wong, W. (2004). 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) Mahwah, NJ: Erlbaum.