Discourse and long-term effects of isolated and combined structured input and structured output on the acquisition of the English causative form

Alessandro Benati and Maria Batziou

Abstract
The present study explores discourse and long-term effects of structured input and structured output when delivered in isolation or in combination on the acquisition of the English causative. Research investigating the effects of structured input has indicated that it is the causative variable in the positive effects of processing instruction. To provide answers to the questions formulated in this study, one classroom experiment was carried out. Sixty-eight school-age Greek learners (aged 10-12) participated. The participants were randomly assigned to three groups: structured input only group (*n*=22); structured output only group (*n*=22); combined structured input and structured output group (*n*=24). Subjects who scored lower than 50% in the pre-tests were included in the final data collection. Instruction lasted for three hours. The design included a delayed post-tests battery (immediate, three weeks after instruction, twenty-four weeks after instruction). The assessment tasks included an interpretation and production task at discourse-level. The results indicated that learners who received structured input both in isolation and in combination benefitted more than learners receiving structured output only. These two groups were able to retain instructional gains three weeks after instruction (short-term effects) in all assessment measures. Long-term effects after six months were also investigated.

1. Background

1.1. Introduction

One of the key assumptions in second language acquisition theory and research is that there is not a one-to-one correspondence between the linguistic input learners are exposed to and their internal language system (Authors, XXXX). Only a small proportion of input is actually taken in (intake) by learners during comprehension and subsequently made available for the internal developing system. The input processing theoretical framework (VanPatten, 1996, 2004, 2015a), is concerned with two conditions: the conditions in which L2 learners process input and ultimately make from-meaning mappings (when learners hear the sentence *I talked to my teacher* and understand that *talked* means that the action is in the past a form-meaning connection is made); and the grammatical roles L2 learners assign to nouns based on their position in an utterance. VanPatten's theory of input processing consists of two main

overarching principles addressing different aspects of processing. The first principle called 'The Primacy of Meaning Principle' asserts that when L2 learners are exposed to input, they are primarily concerned with extracting meaning before they process grammatical forms. The second principle called 'The First Noun Principle' asserts that the order in which L2 learners encounter sentence elements is a powerful factor in assigning grammatical relations amongst sentence elements. According to this processing principle, learners, in the attempt to make moment-by-moment computation of sentence structure during comprehension, would process the first element they encounter in the sentence as the subject of the sentence.

Processing Instruction (PI henceforth) is a pedagogical intervention to grammar instruction which derives from the input processing theoretical model described above. The characteristics of this pedagogical intervention have been described in detail in previous work (Authors, XXXX; Farley, 2005; Lee & VanPatten, 1995; VanPatten, 2015a, 2015b; Wong, 2004). A first characteristic of PI is that learners are required to simultaneously focus on form to get meaning to improve their ability to process the right information and make the right form-meaning connections during comprehension. This is a different function from simply noticing a form in the input which means being consciously aware of something in the input. For example, learners might hear the word 'played' and notice that it is different from either 'plays' or 'playing.' However, they might not connect immediately the ending (verbal inflection) with the concept (e.g. pastness, present) that the action has already taken place, takes place at the moment or is taking place. PI is fundamentally different from other input-based pedagogical interventions to grammar instruction such as input enhancement. This is because, enhancing a feature in the input might help learners to notice that feature but it does not necessarily mean that learners actually link meaning

with form or that it facilitates how they are able to build sentence structure as part of parsing. A second characteristic of PI is that it focuses on the processing of morpho-phonological units as well as sentences. The measures used to assess 'processing' are therefore interpretation measures. In PI research rules are not tested (e.g., grammaticality judgment tasks, fill-in-the-gap, etc.) but the ability to process is. In PI research, when production measures are used, they are not meant to show that speaking or writing skills are developing as a result of its beneficial effects. PI alters the way in which the input is processed by learners, which in turn might have an effect on their developing system. PI is not an intervention meant to assist in skill development, but it might help learners to access the right information (form or structure) in order to express meaning. To summarize, PI is a pedagogical intervention to grammar instruction that uses a particular type of input to push learners away from non-optimal processing strategies so that they are more likely to make correct form-meaning connections or parse sentences appropriately (compute basic structure in real time) during comprehension. It is mainly concerned with the processing of morpho-phonological units in input strings and the development of underlying linguistic representation.

Wong (2004, p.33) asserts that ''the goal of PI is to help L2 learners derive richer intake from input by having them engage in structured input activities that push them away from the strategies they normally use to make form-meaning connections''. Most simply put, a main objective of PI is to ensure that L2 learners process forms and structures (one at a time) correctly and efficiently in the input they receive. PI consists of two main components: learners are given explicit information about a linguistic structure of form and the particular processing strategy that may negatively affect their picking up of the form or structure during language processing; learners

are pushed to process (not producing) the form or structure during structured input tasks. More information on the nature and characteristics of structured input tasks is presented later in this paper.

Research within the PI framework has repeatedly demonstrated (Authors, XXXX; Authors, XXXX; Authors, XXXX) that when input is meaningful and considers of learners' processing strategies, it can affect positively the acquisition of formal features of the target language. PI has an effect on how learners interpret target features affected by a processing problem and at the same time, it has an effect on learners' developing system and what they can access for production (VanPatten & Cadierno, 1993; Cadierno, 1995; Authors, XXXX; Cheng, 2004; VanPatten & Wong, 2004). Within PI, the structured input component is the causative variable responsible for the way learners process input and what they can access for production (see results from studies investigating the effects of PI and its components in VanPatten & Oikennon, 1996; Authors, XXXX; Sanz, 2004; Wong, 2004; Authors, XXXX). This has been proved and observed in different processing principles, languages, linguistic items and assessment task. Structured input activities represent the most significant variable within the PI pedagogical intervention.

1.2. Research measuring discourse effects for PI

There is a large body of research in PI (Authors, XXXX; Authors, XXXX; Leeser, forthcoming) which provides evidence that learners who received this type of instruction performed significantly better on sentence-level interpretation tasks than learners receiving other type of instruction such as traditional instruction (paradigmatic explanations of rules followed by drill practice) or less mechanical output-based interventions to grammar instruction (e.g. structured-output tasks, more

later). In addition, PI can cause equal improvement compared to output-based approaches in learner's performance in different sentence-level production tasks (oral and written). Can PI effects be measured on discourse-level interpretation and production tasks?

Authors, XXXX assessed the effects of PI using discourse-level interpretation tasks. They showed that PI on English past tense led to improved scores on both a sentence-level and discourse-level interpretation task. Authors, XXXX showed that PI on Japanese passive constructions led to improved scores on two different discourse-level interpretation tasks (immediate effects), both presented aurally. Preliminary results from this line of enquiry (Authors, XXXX) indicate that the PI group makes measurable gains not only in the interpretation sentence-level tasks but also in discourse-level tasks confirming Lee's original hypothesis (2004). According to Lee (2004:322) ''PI will yield significant improvement on discourse level interpretation tasks''.

VanPatten and Sanz (1995) and Sanz (1997) showed that PI on Spanish direct object pronouns led to improved scores on oral and written video-based retellings and oral and written structured interview. Subsequent research confirmed this finding, Sanz (2004) and Sanz and Morgan-Short (2004), again with PI on Spanish direct object pronouns, found positive effects on oral video retellings. Other types of discourse-level assessment tasks have been used in PI research. Cheng (2002, 2004) found positive effects for PI on Spanish copular verbs using a picture-based guided composition. Authors, XXXX used a guided composition and found a positive effect for PI on the Spanish subjunctive after *cuando*. Empirical findings from these studies have indicated that PI is effective not only at the sentence level but at the discourse level production tasks (immediate and short-term effects).

Research examining the effects of PI using discourse-level instrumentation is in the minority and certainly future work could well investigate how learners process and produce discourse containing different target forms and whether these effects are durable.

1.3. Research measuring short and long-term effects for PI

The original PI studies conducted by VanPatten and Cadierno (1993) and Cadierno (1995) demonstrated that the effects of PI were sustained one month after instruction. Subsequent PI studies have included delayed post-testing that shows learners retain the benefits of PI for a short-term (short-term is a time period from one to four weeks). PI effects are retained one week after instruction (Cadierno, 1995; Authors, XXXX; Authors, XXXX; Morgan-Short & Bowden, 2006; VanPatten & Cadierno, 1993), two weeks after instruction (Farley, 2001a, 2001b, 2004a, 2004b), three weeks after instruction (Authors, XXXX; Cheng, 2002, 2004), 24 days after instruction (Toth, 2006), and four weeks after instruction (Authors, XXXX; Cadierno, 1995; Keating & Farley, 2008; VanPatten & Cadierno, 1993).

Long-term effects for PI have overall indicated that the effects of instruction diminish over time, but they do not disappear completely. L2 learners receiving PI retained some, if not all, of what they had learned after six weeks (VanPatten, Inclezan, Salazar & Farley, 2009; VanPatten, Farmer & Clardy, 2009; Marsden & Chen, 2011), ten weeks (Kasprowicz & Marsden, 2017), fourteen weeks later (Marsden 2006), and after eight months (VanPatten & Fernández, 2004). VanPatten and Fernández (2004) measured the effectiveness of PI immediately after instruction and again eight months later. The learners' scores on the immediate posttest showed statistically significant improvement from the pretest. Although the scores from the

immediate posttest to the delayed posttest had significantly diminished, the scores eight months later were statistically higher than the pretest scores. The main finding from these long-term effects study is that L2 learners retained some of the benefits of instruction eight months after the instruction took place and demonstrated those benefits through their performance on both an interpretation and production sentence-level tasks. Long-term effects for PI have been measured only in few studies and further research should investigate long-term effects including the measurement of discourse-level tasks.

1.4 Research measuring the effects of structured input and structured output practices

The effects of PI have been compared to meaning output-based instruction (MOI). MOI involves the use of structured output activities which are meaningful activities in nature. These activities carry a meaningful context and the target forms are produced not with the sole intention of practicing the target item, but rather to communicate opinions, beliefs, or other information related to designated topics. The findings on the effects of PI compared to those of MOI are on learners' performance on the production task (e.g., Author, XXXX; Farley 2004a; Authors, XXXX). The findings of this line of investigations on the interpretations task are mixed. Those learners who receive PI sometimes score significantly higher on interpretation tasks than those who receive MOI (Authors, XXXX; Farley 2001; and sometimes make equal gains (Farley 2004b; Morgan-Short and Bowden 2006)). Authors, XXXX and VanPatten, Farmer and Clardy (2009) attributed the relative effects of MOI on interpretation tasks to the fact that in this group learners could have been exposed to incidental structured input. Authors, XXXX compared the effects of MOI and PI on the acquisition of the Italian and French subjunctive. The two treatments were

delivered via computer terminals and learners did not receive input from an instructor of from interacting with other learners. The PI group improved significantly form pre to post-tests on the interpretation and the production tasks. The MOI group only on the production task. The fact that learners did not receive any possible incidental structured input in the MOI treatment might explain the superiority obtained by the PI group in this study on the interpretation tasks

Despite the fact that there is a large number of studies (see Leeser forthcoming and Shintani 2014 for a meta-analysis of research on PI) which have investigated the relative effects of PI when compared to meaning output-based instruction, there is a lacuna in the research body of evidence as far as investigating structured input vs. other output-based approaches to grammar instruction in isolation or combination. Two recent studies have looked into the effects of structured input and PI in combination with different types of output-based instruction. Mystkowska-Wiertelak (2011) investigated the effects of three separate packets of instructional material on the acquisition of English reported speech. Seventy-four first year university students of English were divided into three groups. The first group received structured input + interpretation tasks practice. The second group received output-based instruction mostly mechanical in nature. The third group a combination of structured-input and output—based instruction (mostly mechanical). The target linguistic feature in this study was a complex grammatical phenomenon that requires the application of knowledge regarding morphology (e.g. tense changes), syntax (e.g. word order) and semantics (e.g. use of the appropriate introductory verb). The results of this study indicated that a combination of input and output-based instruction is most beneficial.

Despite the main findings of this study, there are a number of methodological shortcomings: (i) it is not clear why structured input was employed in order to tackle

the processing problems identified in the study (to interpret modal verbs appropriately); (ii) input was not operationalized only as structured input but it also included other interpretation tasks, which minimized the role of structured input; (iii) high pretests scores due to the advanced level of participants who were university students of English philology. Normally, within the PI research framework, participants who score higher than 50% or 60% are excluded from the data analysis since this implies an already existing knowledge of the feature; (iv) the operationalization of the treatments was rather confusing and not reliable. Structured input was operationalized differently. It included interpretation tasks that did not follow the structured input guidelines and principles and therefore did not aid learners to process the target form. It was not clear whether the researchers were comparing the same type of input-based instruction alone and in combination.

Kirk (2013) investigated the effects of PI alone versus PI with output-based meaningful instruction on the acquisition of the Spanish Subjunctive in three conjunctional phrases. Seventy intermediate and high-intermediate (aged 15-17) students (students' level was determined according to ACTFL proficiency guidelines) were assigned to five groups. Precisely, the high-intermediate students were assigned to three groups: one receiving only PI over three days of instruction (PI+PI+PI); the second receiving PI for two days and meaning-based output practice on the third day (PI+PI+O); and the last group a different sequence of PI and output-based practice (PI+O+ PI). The intermediate students were assigned to two groups: one receiving only PI (PI+PI+PI) and one receiving less PI and more output practice than the other experimental groups (PI+O+O). The results of this study indicated that all groups improved from pre-test to post-tests in both the interpretation and production tasks, showing that PI is an effective instructional intervention. The study also shows that

output-based instruction does not enhance or hinder the effects of PI. One of the limitations of this study is the lack of control group.

2. Motivation and research questions

A body of research has compared the effects of the components of PI (explicit information and structured input practice) in order to determine the source(s) of its effects on language development. This work affirms the importance of structured input activities. A different line of research has investigated discourse effects. Preliminary results seem to show that PI is also effective when measured by interpretation and production discourse-level tasks. However, these discourse-level studies have only measured immediate and short-term effects for instruction. Empirical studies measuring PI short and long-term effects have provided important insights into the long-term effects of this pedagogical intervention. The number of studies documenting the short-term effects of PI is far greater than the ones that document long-term effects.

Research investigating a combination of structured input/PI and output-based instruction has provided some interesting and sometime mixed results. However, no empirical study has measured 'pure' structured input practice vs. structured output practice in isolation or in combination.

Despite the different lines of investigation within the PI research agenda, there are still a number of issues that have not been addressed: (a) Would a balanced combination of structured input and structured output-based instruction have beneficial effects? Would learners receiving structured input only practice be able to interpret and produce discourse? Would the effects of structured input tasks have long-terms effects?

The aim of the present study is threefold:

a) To compare and contrast three instructional treatments: structured input (SI) only, structured output (SO) only, and a combination of structured input and structured output (SI+SO);

b) To measure short-term effects of SI and SO practice (separately and in combination) on discourse-level interpretation and production tasks;

c) To measure long-term effects of SI and SO practice (separately and in combination) on discourse-level interpretation and production tasks.

Two specific questions were formulated:

Q1: What are the short-term effects of SI, SO and SI+SO, on the acquisition of the English causative form as measured with discourse-level interpretation and production tasks?

Q2: What are the long-term effects of SI, SO and SI+SO on the acquisition of the English causative form as measured with discourse-level interpretation and production tasks?

3. Design

3.1. Participants and Procedure

To address the questions of this study, an experimental classroom study was carried out. Sixty-eight participants (native speakers of Greek, aged 10-12) were enrolled in an early intermediate English course in a private school in Greece. They were from six different classes of similar level. They were randomly assigned to three groups by drawing names out of a box: SI only (n=22); SO only (n=22); SI+SO

(*n*=24). Each group was taught separately. No control group was used. Participants were removed from the final data pools if they scored over 50% on the pre-tests. Only those participants who had participated in all phases of the experiment were included in the final data analyses.

Instruction lasted for three hours over two consecutive days (1.5 hours x day) in a pretest and posttest design (immediate, short and long-term effects). The regular classroom instructors were trained in the use of the instructional material and they acted as facilitators during the experiment (see an overview of the study in Figure 1.).

FIGURE 1 ABOUT HERE

The researcher controlled the curriculum so that the learners received no further instruction or practice with English causative for a period of six months after instruction. This was achieved by stipulating with the School that the target feature was not to be taught during this period. Despite this, there is always a chance that, although this is distribute equally across the groups, that learners might have been exposed to the target feature during normal classroom talk. A questionnaire was also given at the end of the study to check for possible exposure to the target item outside class.

3.2. Target grammar feature

The target grammar feature selected for this study was the passive English causative. It was chosen for two main reasons. Firstly, because there is no such a structure in Greek. In order to say *John had the dishes done* L1 speakers of Greek would use a structure that sounds like an active structure (something like: *John asked someone to do the dishes*). Secondly, because it is affected by the First Noun Principle

(VanPatten, 1996, 2002, 2004). According to this principle, L2 learners tend to assign agent status to the first noun or pronoun they encounter in a sentence. For example, in the sentence *Lina had her dress mended last Monday*, learners would process *Lina* as the person who actually mended the dress. The use of this default processing strategy would cause misunderstanding and delay in the acquisition of the target feature and word order pattern. The main goal of SI would be to aid learners parse English causative correctly and appropriately. Previous research within the input processing framework has provided evidence for the positive effects of SI in circumventing the First Noun Principle and helping learners to correctly interpret and produce sentences containing the target feature (VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten &Wong, 2004; Morgan-Short & Bowden, 2006; Kasprowicz and Marsden, 2017).

3.3. Instructional treatments

Three instructional treatments were used in this experiment. The two sets of material were piloted before their use in the final experiment. They were balanced in terms of activity types, number of target features, duration of activities, vocabulary items (high frequency and familiar items) and use of visuals. At the end of each pilot some adjustments were made to balance the two packets of material. The activities were provided to learners on worksheets. No explicit instruction about the target feature was provided. At no time did the participants receive either explanation or feedback about the target form. Participants in the three groups were only informed whether they were right or wrong but no explanation was given at the end of each activity.

The SI treatment

Structured input is the practice component of PI. During structured input tasks, learners are pushed to process the targeted form or structure through tasks in which the input is manipulated in particular ways to push learners to become dependent on the form or the structure to get meaning. Structured input tasks created for this study were of two types: referential and affective. Referential tasks are those for which there is a right or wrong answer and for which the learner must rely on the targeted grammatical form to get meaning. Affective structured input tasks are those in which learners express an opinion, belief, or some other affective response and are engaged in processing information about the real world. The SI treatment, developed for this experiment, contained in total forty target items in ten activities (see sample in Figure 2.) It was developed according to the following guidelines provided by Lee and VanPatten, (1995) and Farley (2005) for developing structured input activities: 1) present one thing at a time (one target feature at a time); 2) keep meaning in focus; 3) move from sentence to discourse 4) use both written and oral input; 5) have learners do something with the input; and, 6) keep learners' processing strategies in mind. More specifically, it consisted of eight referential and two affective activities (both aural and written input). Activities were structured in a way so that learners relied on the causative structure to correctly understand meaning in the input. They were developed so that they aided learners to circumvent the First Noun processing principle by manipulating word order and contrasting passive (English) causative structures to SVO order active structures where the first noun was the causer/agent of the action. All the activities were communicative and meaningful and learners were asked to interpret input correctly. No activities were included where learners had to produce the target grammar feature.

FIGURE 2. ABOUT HERE

The SO treatment

The SO treatment contained in total forty target items in six activities (see sample in Figure 3.) developed according to the following guidelines provided by Lee and VanPatten (1995) to develop structured output activities: 1) present one thing at a time; 2) Keep meaning in focus; 3) move from sentence to discourse 4) use both written and oral production; 5) other learners must respond to the content of the output; and, 6) the learners must have some knowledge of the form or structure. Learners included in the final pool had a limited knowledge of the target form as only learners who scored up to 50% of the maximum score available in the pre-tests were included in the final pool. Each activity contained four steps that pushed learners to produce both written and oral output. All activities were meaningful and communicative in nature and no mechanical practice was included. They were working in pairs or groups to complete each task. Each activity included familiar items and a list of vocabulary items was also provided. As stated by Lee and VanPatten (1995, p.121), structured output has two characteristics: ''involves the exchange of previously unknown information; requires learners to access a particular form or structure in order to express meaning''.

FIGURE 3. ABOUT HERE

The SI+SO treatment

This treatment contained in total forty target items in eight activities (five SI -twenty items- and three SO activities- twenty items) selected from the SI and SO treatments. Learners in this group received SI practice first and this was followed by structured-output tasks.

3.4. Assessment tasks and scoring

A pretest and posttest split block design was used. There were four versions of the interpretation test and the production test so to allow different distributions across the different stages of the experiment (pre and posttests, see figure 4. for an overview of test distribution. All the tests were piloted to ensure same level of difficulty. Pretests were administered to all experimental groups a week before the beginning of the instructional period. Immediate posttests, three-week and twenty-four weeks delayed posttests were administered to all participants to measure treatments' and time effects. The assessment tasks consisted of a discourse-level aural interpretation and a discourse-level production task. Four versions of each assessment tasks were developed and balanced in terms of difficulty and vocabulary.

FIGURE 4 ABOUT HERE

The discourse-level interpretation task was developed to measure the ability of learners to interpret correct English causative forms when these forms are embedded in discourse. Participants had to listen to a story which was presented into three segments each containing three target items and two distractors. The task had nine target items and six distractors in total. A booklet was constructed for the discourse-level interpretation task. Learners heard the story segment only once, then turned into the appropriate answer sheet (pictures showing two different characters doing the same action), and they had to decide who was performing the action or opt for *Not Sure* if they could not understand who the agent was. Participants received 1 point for each correct selection and 0 point for each incorrect one.

The discourse-level production task was developed to measure learner's ability to produce correct English causative forms at discourse-level (text re-reconstruction). It

contained five target items (instances of the English causative form). Participants had to revise some key vocabulary items before the beginning of the assessment. The instructor read a story at normal pace providing information about the context at the beginning. As the instructor read the story, it was also projected on the whiteboard enabling participants to read and listen at the same time. After listening to the story, participants were asked to re-write the story they had just heard. They had 5 minutes to re-construct the story with the help of some prompts. For each correct use of the target structure one point was awarded. However, if the participants had used the correct structure of the target item but the wrong form of the verb (i.e., an infinitive instead of a participle) they were awarded half a point. No points were awarded if the target item was not produced even if meaning was conveyed properly and the agent was identified correctly (e.g.: if the sentence in the story was: "she had the walls painted pink" and the participant wrote "They painted the walls pink"). The maximum score for the discourse level tasks was five points as this was the maximum number of target items learners needed to complete the story.

4. Results

4.1. Discourse-level interpretation data

A one-way ANOVA was conducted on the pre-test scores. The analysis showed no significant differences among the three groups before instruction (F (2,68) = 2.113 p = .138). Any differences found after instruction will be attributed to the effects of instruction. Table1 provides the descriptive statistics for learners' performance on the discourse-level interpretation tasks. The descriptive statistics show the means of the three groups in the discourse-level interpretation task (pre-test, post-test and delayed

post-tests). The SI group and the SI+SO groups clearly improved from pre-test to post-test scores.

<center>TABLE 1. ABOUT HERE</center>

A repeated-measures ANOVA was adopted on the raw scores of the discourse-level interpretation task. It showed a significant main effect for Treatment (F (2,68) = 69.744, $p<$ .000); a significant main effect for Time F (2,68) = 22.300, $p<$ .000; and significant interaction between Treatment and Time F (2,68) = 18.244, $p<$ .000. Given the significant main effect for instructional treatment post-hoc tests were conducted to compare the groups' scores from the pre-test to the post-tests. The post-hoc Scheffe test showed that the SI group and the SI+SO groups were not statistically different (p = .204) and significantly different than the SO (p = .000).

To investigate possible delayed effects, a second repeated measured ANOVA was conducted on the raw scores of the discourse-level interpretation post-tests (immediate effects vs. three weeks delayed effects vs. twenty-four weeks). The results showed a significant main effect for Treatment (F (2,68) = 37.507, $p<$ .000) no significant main effect for Time F (2,68) = 28.451, $p<$ .102; and no significant interaction between Treatment and Time F (2,68) = 14.654, $p<$ .158. The post-hoc Scheffe test showed that the SI group and the SI+SO groups were not statistically different (p= .117) and significantly different than the SO (p = .000).

The results from the discourse-level interpretation task indicated that only the SI and SI+SO groups gained in their ability to interpret English causative forms presented at the discourse-level. These gains were durable and maintained over a period of six months. Although a slight decrease was observed in the two delayed

posttests compared to the immediate posttest scores this was not statistically significant. The SO group made no significant gains.

4.2. Discourse-level production data

The one-way ANOVA conducted on the pre-test scores showed no significant differences between the three groups before instruction (F (2,68) = 1.223 p = .801). Any differences found after instruction will be attributed to the effects of instruction. Table 2 provides the descriptive statistics for learners' performance on the discourse-level production task. The descriptive statistics show the means of the three groups in the discourse-level interpretation task in the immediate post-test and delayed post-tests. The SI group and the SI+SO groups improved from pre-test to post-test scores. The SO group did not improve.

TABLE 2. ABOUT HERE

A repeated-measures ANOVA was used on the raw scores of the discourse-level production task. It showed a significant main effect for Treatment (F (2,68) = 10.004, p< .000); a significant main effect for Time F (2,68) = 5.772, p< .003; and significant interaction between Treatment and Time F (2,68) = 1.007, p< .004. Given the significant main effect for instructional treatment post-hoc tests were conducted to compare the groups' scores from the pre-test to the post-tests. The post-hoc Scheffe test showed that the SI group and the SI+SO groups were similar (p= .401) and significantly different than the SO (p =. 000).

To investigate possible delayed effects, a second repeated measured ANOVA was conducted on the raw scores of the discourse-level production post-tests (immediate effects vs. three weeks delayed effects vs. twenty-four weeks). The results showed a

significant main effect for Treatment (F (2,68) = 17.151, p< .000) no significant main effect for Time F (2,68) = 12.139, p< .171; and no significant interaction between Treatment and Time F (2,68) = 8.279, p< .101.

The results from the discourse-level production task demonstrated that only the SI and SI+SO groups gained in their ability to produce the English causative forms presented at the discourse-level. The SI and SI+SO groups maintained these gains over a period of six months.

5. Summary of findings

The first research question was: What are the short-term effects of SI, SO and SI+SO, on the acquisition of the English causative form as measured with discourse-level interpretation and production tasks? The results of the interpretation discourse-level task clearly indicated that SI and a combination of SI+SO helps learners to process the English causative forms correctly and appropriately (immediate effects). The improvement found for the SI and SI+SO groups was maintained three weeks after instruction. Both the SI and SI+SO instructions were more effective than SO only in interpreting the target form embedded in discourse. The results of the production discourse-level task indicated that the two instructional groups (SI and SI+SO) performed better than the SO only group (immediate post-tests). This advantage was maintained after three weeks.

The second research question was: What are the long-term effects of SI, SO and SI+SO on the acquisition of the English causative form as measured with discourse-level interpretation and production tasks?

The results of the interpretation discourse-level task clearly indicated that the improvement found for the SI and SI+SO groups was maintained six months after instruction. The results of the production discourse-level task echoed the ones obtained for the interpretation discourse-level task and indicated that the two instructional groups (SI and SI+SO) maintained the good effects of instruction after six months, whereas the SO group did not. Both SI and SI + SO groups' scores slightly decreased at the three-week and twenty-four-week test, and immediate to twenty-four-week test and this is in-line with previous findings investigating delayed effects for processing instruction.

6. Discussion and Conclusion

The motivation of this study was to investigate the relative effects of SI and SO in isolation and in combination using both interpretation and production discourse-level tasks and measuring short and long-term effects. The findings from the interpretation and production discourse-level tasks will be interpreted and discussed separately in this section.

The findings from the interpretation discourse-level task provide empirical support for the view that SI only is better than SO only in altering the way learners processed input. SI is a better form of pedagogical intervention than SO in helping learners to process and interpret English causative forms and providing 'good' intake for the developing system. The SI group and SI+SO group shared the SI component and this provides further evidence to support the view that SI alone is sufficient to improve learner's performance as the SO only group made no gains in the interpretation discourse-level task. One possible explanation of these results might be found in the nature and purpose of SI activities. As outlined by Wong (2004, p.35) SI ''push

learners to abandon their inefficient processing strategies for more optimal ones''. SI significantly improves learners' interpretation of discourse containing English causative forms. In the interpretation task, the SI group and the SI+SO group were able to maintain the gains made over a period of six months.

The main findings from the discourse-level production task indicate that only the SI and SI+SO groups made statistically significant gains from pretest to posttests and those gains were maintained over a period of six months. SO alone was not enough to provide learners with the ability to produce discourse containing the target feature. Only by altering learners' developing system would we get effects on tasks that were not practiced during instruction like in the case of the discourse-level production task. The findings from the production discourse-level task are in line with previous research findings in the PI/SI research agenda (Authors, XXXX; Authors, XXXX; see also Authors, XXXX; and Authors, XXXX). SI has clearly altered the way learners processed input and this had an effect on L2 learners' developing system and subsequently on what learners could access for production. These findings are consistent with the original findings from VanPatten and Cadierno (1993, p.240). ''Learners who receive instruction that attempts to alter input processing receive a double bonus: better processing of input and knowledge that is apparently also available for production''. This view is further supported by the results from the production discourse-level tasks where learners had to produce the target form in a less controlled situation. While the SI treatment is successful at circumventing processing problems and consequently having an impact on learners' developing system, the SO treatment was not successful at producing positive effects (circumventing the processing problem) on learners' performance. Output practice

might help with fluency and accuracy in production but it is not 'responsible' for getting the grammar into learners' head.

Overall, the main findings from this study make a number of theoretical and pedagogical contributions to the ongoing debate on the effects of SI and SO:

Firstly, the results of this experimental study confirm the key role of SI as an effective pedagogical intervention designed to alter processing problems such as the First Noun Principle. SO only is not successful in bringing about similar effects to those brought about by SI in interpretation discourse-level tasks. The findings from this study reaffirm the importance of input-based practice as a key pedagogical tool and make a contribution to the view that this practice should precede output practice (structured-input activities should precede structured-output activities). However, more research is needed to establish whether this should be always the case. Replication studies should also consider the combination SO + SI to investigate other possible combinations in instruction.

Secondly, SI is the causative variable for the change in performance of the groups. Not only SI is effective in developing learners' ability to process input (at discourse-level) but also has an impact on their developing system so that learners can access a linguistic feature in written production tasks under less controlled situations (discourse-level task). SI altered the way learners processed input and assisted in the developing of underlying knowledge. As VanPatten (2015b, p.100) argued ''it assists in the developing of underlying knowledge that can be tapped during the development of skill''. SO only practice did not have an effect on the ability for L2 learners to produce discourse containing English causative forms.

Thirdly, the results from this study provide new evidence on the short-term and long-term effects of SI practice. The effects of both the SI and the combination (SI +SO) treatments were measured and maintained over a period of three weeks (short-term effects) and six months (long-term effects). There were no effects for the SO treatment. The effects of SI are both durative (three weeks after instruction) and longitudinal (six months after instruction). If we look at the descriptive statistics in Table 1 and Table 2, we notice that the improvement of the SI group from the second (three weeks' delayed post-test) to the third post-test (six months' delayed post-test) was maintained and effects did not diminish for neither the interpretation nor the production discourse-level tasks.

Fourthly, the findings from this study lend support to a number of hypotheses formulated within the PI research framework (Authors, XXXX; Authors, XXXX). The positive results obtained in this study lend support to the so-called *Age Hypothesis* (Authors, XXXX) ''PI will be just as effective an intervention with young learners as it is with older learners''. Our results, with learners aged 10-12, confirmed previous research (Authors, XXXX; Marsden and Chen, 2011; Kasprowicz and Marsden, 2017) demonstrating the effectiveness of structured input tasks with young learners.

The findings from this study in using native speakers of other languages than English, also support the so-called *Native Language Hypothesis* (Authors, XXXX) ''PI will be effective for instilling target language specific processing strategies, no matter the native language of the learners''. The present study contributes to the expanding of the *Native Language Hypothesis* by adding Greek, a non-western language, to the current list of languages on which the effectiveness structured input practice has been observed.

Despite the positive outcomes of the present study, there are some limitations for the present study. The lack of a control group (due to attrition and other selection issues) is a methodological limitation. Future research should replicate this study using a control group and a larger sample of participants. This study measured the effects of SI and SO on the interpretation and production discourse-level tasks. This research is in the minority and certainly future work should further investigate discourse and long-term effects comparing SI with other instructional interventions (input or output-based). These effects should also be measured on more spontaneous production tasks that include time pressure, measure reaction time, and do not allow learners to monitor their responses. This research might also consider the role and effects of structured-input and structured-output tasks with or without explicit information.

Acknowledgements

7. References

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX)

Authors,XXXX

Authors,XXXX

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, 79, 179-93.

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 Spanish *ser* and *estar*: Forms with semantic-aspectual value. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and* commentary, (pp. 119-141). Mahwah, NJ: Erlbaum.

Farley, A. (2001a). The effects of processing instruction and meaning-based output instruction. *Spanish Applied Linguistics*, 5, 57-94.

Farley, A. (2001b). Authentic processing instruction and the Spanish subjunctive. *Hispania*, 84, 289-299.

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

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

Farley, A. (2005). *Structured input: Grammar instruction for the acquisition-oriented classroom*. New York: McGraw-Hill.

Kasprowicz, R., & Marsden, E. (2017). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, pages forthcoming.

Keating, G. and A. Farley. (2008). Processing instruction, meaning-based output instruction, and meaning-based drills: Impacts on classroom L2 acquisition of Spanish object pronouns. *Hispania*, 19, 639-650.

Kirk, R. 2013. The effects of processing instruction with and without output: Acquisition of the Spanish subjunctive in three conjunctional phrases. *Hispania*, 96, 153-169.

Lee, J. (2004). On the generalizability, limits, and potential future directions of processing instruction research. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and* commentary, (pp. 311-323). Mahwah, NJ: Erlbaum.

Lee, J. and B. VanPatten. (1995). *Making communicative language teaching happen.* New York: McGraw-Hill.

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Authors (XXXX).

Leeser, M. (Forthcoming). A research synthesis and meta-analysis of Processing Instruction.

Marsden, E. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning,* 56, 507-566.

Marsden, E., & Chen, H. Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, 61(4), 1058-1098.

Morgan-Short, K. and H.W. Bowden. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition,* 28, 31-65.

Mystkowska-Wiertelak, A. (2011). The effects of a combined output and input-oriented approach in teaching reported speech. *Research in Language, 9*, 112-126.

Sanz, C. (1997). Experimental tasks in SLA research: Amount of production, modality, memory, and production processes." In A. T. Pérez-Leroux and W. R. Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish: Production, processing and comprehension,* (pp. 41-56). Somerville, MA: Cascadilla Press.

Sanz, C. (2004). Computer delivered implicit versus explicit feedback in processing instruction. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and* commentary, (pp. 241-255). Mahwah, NJ: Erlbaum.

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

Shintani, N. (2014). The effectiveness of processing instruction and production-based instruction on L2 grammar acquisition: A meta-analysis. *Applied Linguistics*, 36 (3), 306-325.

Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. *Language Learning,* 56, 319-385.

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

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

Van Patten, B. (2015a). Input processing in adult SLA. In B. VanPatten and J. Williams, (Eds.), *Theories in second language* acquisition, (pp. 113-135). New York: Routledge.

VanPatten, B. (2015b). Foundations of processing instruction. *IRAL*, 53, 91-109.

VanPatten, B. and T. Cadierno. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition,* 15, 225-243.

VanPatten, B. and C. Sanz. (1995). From input to output: processing instruction and communicative tasks. In F.R. Eckman, D. Highland, P. W. Lee, J. Mileham and R. R. Weber (Eds.), *Second language acquisition theory and pedagogy,* (pp. 169-185). Mahwah, NJ: Erlbaum.

VanPatten, B. and S. Oikennon. (1996). Explanation vs. structured input in processing instruction. *Studies in Second Language Acquisition,* 18, 495-510.

VanPatten, B. and C. Fernández. (2004). The long-term effects of processing instruction. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and commentary,* (pp. 273-289). Mahwah, NJ: Erlbaum.

VanPatten, B. and W. Wong. (2004). Processing instruction and the French causative: Another replication. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and* commentary, (pp. 97-118). Mahwah, NJ: Erlbaum.

VanPatten, B., J. Farmer, and C. Clardy. (2009). Processing instruction and meaning-based output instruction: A response to Keating and Farley (2008). *Hispania,* 92, 116-126.

VanPatten, B., D. Inclezan, H. Salazar and A. Farley. (2009). Processing instruction and dictogloss: A study on object pronouns and word order in Spanish. *Foreign Language Annals,* 42, 557-575.

Authors (XXXX).

Wong, W. (2004). The nature of processing instruction. In B.VanPatten (Ed.), *Processing instruction: Theory, research, and* commentary, (pp. 33-65). Mahwah, NJ: Erlbaum.