The relative effects of isolated and combined structured input and structured output on the acquisition of the English causative forms.

Abstract
The present study explores the 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 processing instruction and meaning output-based instruction has provided some interesting and sometimes conflicting results. Additionally, there are a number of issues (e.g., measuring a combination of structured input and structured output, measuring discourse-level effects) that have not been fully and clearly addressed. To provide answers to the questions formulated in this study, two classroom experiments were carried out. In the first study, fifty-four Chinese university students (age 18-20) participated. The participants were randomly assigned to four groups: structured input only group (n=13); structured output only group (n=15); combined structured input and structured output group (n=16); control group (n=10). In the second study, thirty school-age Greek learners (age 10-12) participated. The participants were randomly assigned to three groups: structured input only group (*n*=10); structured output only group (*n*=10); combined structured input and structured output group (*n*=10).

Only subjects who participated in all phases of each experiment and scored lower than 60% in the pre-tests were included in the final data collection. Instruction lasted for three hours. The control group received no instruction on the causative structure. Interpretation and production tasks were used in a pre-test and post-test design. The design included a delayed post-test battery (3 weeks after instruction) for both experiments. In the first study, the assessment tasks included an interpretation and production task at sentence-level, and an interpretation task at discourse-level. In the second study, an additional discourse-level production task was adopted along with the interpretation discourse-level task. 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 later in all assessment measures.

Keywords
Structured input, structured output, English causative, sentence-level tasks, discourse-level tasks.

1. Background

1.1. Introduction

One of the key issues in second language acquisition (SLA henceforth) concerns of the role instruction (VanPatten and Benati, 2015). A great deal of classroom-based research in instructed SLA has investigated the role of different input and output-based pedagogical interventions. Language learners must be able to develop the underlying linguistic system of a target language and also the ability to use that linguistic system for communication.

Processing instruction (PI henceforth) is a pedagogical intervention predicated on VanPatten's input processing theory (VanPatten, 2015a). This theory consists of two overarching principles addressing two different aspects of processing: processing input for meaning (The Primacy of Meaning Principle, VanPatten, 2004); and placing agents and subjects in a first noun position (The First Noun Principle, VanPatten, 2004). The main focus of PI is to help learners to develop their internal linguistic system by ensuring that they process forms/structures correctly and appropriately in the input they receive. Research within the PI framework has repeatedly demonstrated (Benati and Lee, 2015) that when input is meaningful and takes into account of learners' processing strategies, it can affect positively the acquisition of formal features of the target language. There is enough empirical evidence to support the view (Lee and Benati, 2009; Benati and Lee, 2010; Benati and Lee, 2015) that PI is better than traditional output-based instruction (explicit information about the target feature followed by mechanical output-based instruction) in increasing learner's rate of processing language input. At the same time, PI has an effect on learners' developing system and what they can access for production (VanPatten and Cadierno, 1993; Cadierno, 1995; Benati, 2001; Cheng, 2004; VanPatten and Wong, 2004). However, one of the objections raised by Cadierno (1995) was whether the same results can be found when PI is compared to a more meaning-output based pedagogical intervention.

1.2. Research measuring PI vs. Meaning output-based instruction

Farley (2001, 2004) was the first who compared the effects of PI versus meaning output-based instruction (MOI) on the acquisition of the Spanish subjunctive. The

activities for the MOI intervention were based on the tenets of structured-output activities (see structured-output activities guidelines in the design section of this paper) proposed in Lee and VanPatten (1995).

In his first study (2001) twenty-nine subjects enrolled in a fourth-semester Spanish course were assigned to two different instructional treatments: processing instruction and meaning-based output instruction. The instructional period lasted for two days. Pre-tests and post-tests consisted of an interpretation and production tasks. The results showed that PI had a greater effect than MOI on how L2 learners interpret subjunctive forms in Spanish. Both instructional treatments had similar effects on how L2 learners produce sentences containing those forms.

One hundred and twenty nine undergraduate students from ten different sections of a fourth-semester Spanish grammar review course participated to Farley's second experiment (2004). They were assigned to two instructional groups (PI vs. MOI). Both groups were assessed following a pre-test/post-test design which included interpretation and production tasks. Farley's results differed from his previous study as both groups made equal and significant improvements on both the interpretation and the production tests. Farley attributed the equal performance of the two treatments to one main factor. The MOI treatment is different from traditional instruction as it does not contain mechanical drills practice and its communicative and interactive nature might have resulted in incidental input.

Benati (2005) investigated the effects of PI, traditional instruction and MOI on the acquisition of English past simple tense. The subjects involved in Benati's studies were Chinese (47 subjects) and Greek school-age learners (30 subjects) of English residing in their respective countries. The participants in both schools were divided

into three instructional groups. One interpretation and one production measure were used in a pre and post-test design. The results are very consistent in both studies and showed that PI had positive effects on the processing and acquisition of the target feature. In both studies the PI performed better than the traditional and MOI groups in the interpretation task and the three groups made equal gains in the production task.

Gely (2005) investigated the effects of PI and MOI on the acquisition of French imperfect tense. Thirty-three college students learning French (second year) participated in this experimental study. One group received PI, the second group MOI and the third group acted as a control group. Gely (2005) used three different tasks to measure the effects of instruction: an interpretation task and two production tasks (completion text and a written production task). The results of this study were very similar to Farley (2001) and Benati (2005). The PI group outperformed the other two groups on the interpretation task and the PI group performed equally to the MOI group on the production task.

Morgan-Short and Bowden (2006) carried out a study with forty-five first semester Spanish students assigned to three groups: PI, MOI and a control group. One interpretation and one production task were developed to measure the effects of the two instructional treatments on the acquisition of Spanish object pronouns. The results of this study are very similar to Farley's study (2004). However, the two authors argued that (2006:56) '' input-based practice leads to form-meaning connections evidence in both interpretation and production test gains, whereas in the MOBI, learners might have been able to establish or strengthen form-meaning connections via both output-based practice and the input they receive in instruction and feedback''.

In sum, a group of studies (Farley, 2004; Morgan-Short and Bowden, 2006) indicated that MOI could have equal effects to PI on interpretation and production tasks. However, another group of studies (Farley, 2001; Benati, 2005; Gely 2005) showed that PI had a greater effect than MOI on interpretation tasks, and equal effects on production. The main question is: Why the difference in outcome between these studies? Would the nature of the MOI treatment be the explanation for the conflicting results? Would the particular structures (Spanish subjunctive and Spanish object pronouns) investigated cause the results to be different?

Lee and Benati (2007a) attempted to address these questions by carrying out a parallel study where they contrasted the effects of PI and MOI, delivered in two different modes (via classroom instruction vs. computer terminals) on the acquisition of Italian subjunctive of doubt and opinion (forty-seven first semester university students) and French subjunctive of doubt (sixty-one second semester university students).

The results of the Lee and Benati study (2007a) indicated that PI is better than MOI in helping learners to interpret and produce the target features process no matter the form of delivery. The results obtained in Lee and Benati's (2007a) study might explain the difference in the results between the majority of studies comparing PI and MOI and Farley (2004) and Morgan-Short and Bowden's (2006) studies. As stated by Lee and Benati (2007a: 122) ''could be the fact that both treatments were compared through different modes of delivery. In the case of the meaning output-based instruction and processing instruction in the treatments delivered via computer terminals, learners did not receive input from an instructor or from interacting with other learners''. The group receiving the MOI treatment delivered via computer

terminals received no incidental structured input (i.e. when a learner's output works as input for the rest of the class).

VanPatten, Farmer and Clardy (2009) also conducted a study contrasting the effects of PI to MOI on the acquisition of Spanish object pronouns. One hundred and eight university students participated in this study. This study was a partial replication of a previous study by Keating and Farley (2008). Again the main purpose for the replication was that the MOI treatment in Keating and Farley (2008) was operationalized as PI plus output-based instruction (exposition of learners to incidental structured input). Because of the concern for the PI-like activities in Keating and Farley's (2008) material, VanPatten, Farmer and Clardy reworked their activities in which there were possible conflations of processing and meaning-based output instruction. Findings from VanPatten, Farmer and Clardy (2009) showed that when MOI is not combined with PI and therefore does not involve processing of input in any way it has no overall advantage over PI.

The overall results from the studies reviewed in this section seem to provide a number of contributions to the on-going debate on the effects of PI when contrasted to MOI:

- PI is overall superior to MOI to alter input processing problems and subsequently to have an impact on learners' developing system and what learners can access under controlled situations;
- Structured input is the main component and the causative variable responsible for the way learners process input and what they can access for production (see also studies investigating the effects of PI and its components in VanPatten and Oikennon, 1996; Benati, 2004a, 2004b; Sanz, 2004; Wong,

2004; Lee and Benati, 2007b) as in studies comparing PI vs. MOI all groups received the same amount and type of explicit information and the only difference was the actual pedagogical treatment (structured input versus structured output activities).

1.3. Research measuring the effects of structured input, PI and output-based instruction in isolation and in 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 shortcomings: (i) it is not clear how and why structured input was implemented to aid learners circumvent default processing strategies (i.e. The Lexical Preference Principle, The Non-redundancy Principle, and The Sentence Location Principle) when coping with

reported speech; (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 pre-tests 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 might be rather confusing and not reliable. Structured input was operationalized differently in the first group compared to the third combination group, while output-based instruction was exactly the same both in the isolated and in the combined group. It is not clear whether we are comparing the same type of 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 students (students' level was determined according to ACTFL proficiency guidelines) were assigned to three groups: one receiving only PI over three days 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 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.

2. Motivation and research questions

The results from classroom studies comparing PI versus MOI have shown that the MOI treatment, despite the way it is delivered, is not successful in bringing similar effects to those brought by PI. It is plausible that in the case of MOI treatment delivered via classroom instruction (Farley, 2004; Keating and Farley, 2008) participants might have been affected by incidental structured input. Structured input component has been proved to be the main factor in the improved performance measured through interpretation and production tasks.

While research on combined PI and output-based instruction has provided some interesting results, 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 has beneficial effects? Would learners receiving structured input be able to interpret discourse and produce the target feature in less controlled situations?

The aim of the present study is twofold:

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 the effects of SI and SO practice on both sentence and discourse-level interpretation and production tasks.

Two specific questions were formulated:

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

Q2: What are the immediate and delayed 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

In order to address the questions of this study, two separate experimental classroom studies were carried out. In the first experiment, fifty-four participants (native speakers of Chinese, aged 18-20) were enrolled in an early intermediate English course in a British university (original pool sixty-four subjects). They were randomly assigned to four groups: SI only (*n*=13); SO only (*n*=15); SI+SO (*n*=16); control group (*n*=10). In the second experiment, thirty participants (native speakers of Greek, aged 10-12) were enrolled in an early intermediate English course in a private school in Greece (original pool thirty-four subjects). They were randomly assigned to three groups: SI only (*n*=10); SO only (*n*=10); SI+SO (*n*=10). No control group was available for the second study. Participants were removed from the final data pools if they had contact with English outside class and if they scored over 60% on the pre-tests. Only those participants who had participated in all phases of the experiment were included in the final data analyses.

In both experimental studies, instruction lasted for three hours over two consecutive days in a pre-test and post-test design. The control group received no instruction during the first classroom experiment. In both studies, 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

3.2. Target grammar feature

The target grammar feature selected for this study was the passive English causative. It was primarily chosen 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 *Jane had her dress mended last Monday*, learners would process *Jane* as the person who actually mended the dress. This 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 parsing English causative correctly and appropriately. Previous research within the input processing framework has provided evidence for the positive effects of SI in altering the First Noun Principle and helping learners to correctly interpret and produce sentences containing the target feature (VanPatten and Cadierno, 1993; VanPatten and Oikkenon 1996; VanPatten and Wong 2004; Morgan-Short and Bowden 2006).

3.3. Instructional treatments

Three instructional treatments were used in both experiments. They were balanced in terms of number of target features, duration of activities, vocabulary items (high frequency) and use of visuals. No explicit instruction about the target feature was provided. At no time did the participants received either explanation or feedback

about the target form. Participants in the three groups were only informed whether or not they were right or wrong but no explanation was given.

The SI treatment

The SI treatment contained in total six activities (see sample in Figure 2.) 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; 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 four 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. 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 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) others must respond to the content of the output; and, 6) the learners

must have some knowledge of the form or structure. 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. As stated by Lee and VanPatten (1995:121), 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 six activities (three SI and three SO activities) selected from the SI and SO treatments.

3.4. Assessment tasks and scoring

A pre-test and post-test split block design was used in both experiments. Pretests were administered to all four groups a week before the beginning of the instructional period. Immediate post-tests and three-week delayed post-tests were administered to all participants to measure treatments’ effects. The assessment tasks consisted of a sentence-level aural interpretation and a written production task, and a discourse-level aural interpretation and a written production task. Three versions of each assessment task were developed and balanced in terms of difficulty and vocabulary. In the first experiment, one sentence-level interpretation, one sentence-level production and one discourse-level interpretation task were used. In the second experiment, one discourse-level interpretation, and one discourse-level production task were adopted.

The sentence-level interpretation tasks consisted of 20 sentences (10 target items and 10 distractors). The participants had to listen to the sentences and decide (interpret) who was the agent of the action. They could choose between the 'names' of the person, 'someone else', or they could tick the 'not sure' option. No repetition was provided so that the test would measure real-time interpretation. The raw scores were calculated as follow: 0 point for an incorrect response; 1 point for a correct response.

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. Participant 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. Participants received 1 point for each correct selection and 0 point for each incorrect one.

The sentence-level production task was developed to measure a learner's ability to produce correct English causative forms at sentence-level. It consisted of ten short sentences (seven target items and three distractors). Participants had to write a sentence using correct target forms. They had five minutes to complete the task. Correct and accurate forms received 1 point, and incorrect ones received 0 point.

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.

4. Results

4.1. Sentence-level interpretation data (Study 1)

A one-way ANOVA was conducted on the pre-test scores. The analysis showed no significant differences between the four groups before instruction (F (3,54) = .403 p = .239). Any differences found after instruction will be attributed to the effects of instruction. Table 1 provides the descriptive statistics for learners' performance on the sentence-level interpretation task. The descriptive statistics show the means of the four groups in the sentence-level interpretation task (pre-test, post-test and delayed post-test). The SI group and the SI+SO groups clearly improved from pre-test to post-tests scores.

TABLE 1. ABOUT HERE

A repeated-measures ANOVA was used on the raw scores of the sentence-level interpretation task. It showed a significant main effect for Treatment (F (3,54) = 286.456, p< .000); a significant main effect for Time F (3,54) = 115.111, p< .000; and significant interaction between Treatment and Time F (3,54) = 49.839, p< .001. Given the significant main effect for instructional treatment post-hoc tests were conducted to compare the group's scores from the pre-test to the post-tests. The post-hoc Tukey test showed that the SI group and the SI+SO groups were equal (p = .059) and significantly different than the SO (p = .000) and Control group (p = .000). There was no difference between the SO group and Control group (p = .107).

To investigate possible delayed effects, a second ANOVA was conducted on the raw scores of the two sentence-level interpretation post-tests. The results showed a significant main effect for Treatment (F (3,54) = 311.527, p< .000). The post-hoc Tukey test showed that the SI group and the SI+SO groups again equal (p = .057) and significantly different than the SO (p = .002) and Control group (p = .000). There was no difference between the SO group and Control group (p = .636).

The results from the sentence-level interpretation task demonstrated that only the SI and SI+SO groups gained in their ability to interpret English causative forms presented at the sentence-level. These gains were maintained over a period of three weeks. The SO group and the control group made no significant gains.

4.2. Sentence-level production data (Study 1)

A one-way ANOVA was carried out on the pre-test scores. The analysis showed no significant differences between the four groups before instruction (F (3,54) = .343 p = .165). Any differences found after instruction will be attributed to the effects of

instruction. Table 2 provides the descriptive statistics for learners' performance on the sentence-level production task. The descriptive statistics show the means of the four groups in the sentence-level production task. The SI group, the SO group and the SI+SO groups clearly improved from pre-test to post-test scores.

TABLE 2. ABOUT HERE

A repeated-measures ANOVA was used on the raw scores of the sentence-level production task. The results from the statistical analysis showed a significant main effect for Treatment (F (3,54) = 114.357, p< .000); a significant main effect for Time (3,54) = 66.958, p< .000; and significant interaction between Treatment and Time F (3,54) = 51.902, p< .000. A post-hoc Tukey test yielded the following contrasts: SI group, SO group and the SI+SO groups were equal (p =. 205) and significantly different than the Control group (p = .000).

To investigate possible delayed effects, a second ANOVA was conducted on the raw scores of the sentence-level production post-tests. The results showed a significant main effect for Treatment (F (3,54)= 115.642, p< .000). The post-hoc Tukey test showed again that the three instructional groups were equal (SI=SO=SI+SO, p =. 945) and significantly different than the Control group (p = .000).

The results from the sentence-level production task demonstrated that all the instructional treatments (SI, SO and SI+SO) made equal gains in their ability to produce English causative forms at sentence-level from pre-test to post-test. These gains were maintained over a period of three weeks.

4.3. Discourse-level interpretation data (Study 1)

A one-way ANOVA was conducted on the pre-test scores. The analysis showed no significant differences between the four groups before instruction (F (3,54) = .435p = .225). Any differences found after instruction will be attributed to the effects of instruction. Table 3 provides the descriptive statistics for learners' performance on the sentence-level interpretation tasks. The descriptive statistics show the means of the four groups in the discourse-level interpretation task (pre-test, post-test and delayed post-test). The SI group and the SI+SO groups clearly improved from pre-test to post-test scores.

TABLE 3. ABOUT HERE

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 (3,54) = 231.740, p< .000); a significant main effect for Time F (3,54) = 118.469, p< .000; and significant interaction between Treatment and Time F (3,54) = 98.525, p< .000. Given the significant main effect for instructional treatment post-hoc tests were conducted to compare the group's scores from the pre-test to the post-tests. The post-hoc Tukey test showed that the SI group and the SI+SO groups were equal (p = .074) and significantly different than the SO (p = .000) and Control group (p = .000). There was no difference between the SO group and Control group (p = .444).

To investigate possible delayed effects, a second ANOVA was conducted on the raw scores of the two discourse-level interpretation post-tests. The results showed a significant main effect for Treatment (F (3,54) = 88.469, p< .000). The post-hoc Tukey test showed that the SI group and the SI+SO groups were again equal (p= .237) and significantly different than the SO (p = .000) and Control group (p = .000). There was no difference between the SO group and Control group (p = .994).

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 maintained over a period of three weeks. The SO group and the control group made no significant gains.

4.4. Discourse-level interpretation data (Study 2)

The one-way ANOVA conducted on the pre-test scores showed no significant differences between the three groups before instruction (F (2,30) = 12.403 p = .110). Any differences found after instruction will be attributed to the effects of instruction. Table 4 provides the descriptive statistics for learners' performance on the discourse-level interpretation 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. There was also a minimum improvement of the SO group in the first post-test.

TABLE 4. ABOUT HERE

A repeated-measures ANOVA was used on the raw scores of the discourse-level interpretation task. It showed a significant main effect for Treatment (F (2,30) = 84.626, p< .000); a significant main effect for Time F (2,30) = 78.970, p< .000; and significant interaction between Treatment and Time F (2,30) = 33.801, p< .001. Given the significant main effect for instructional treatment post-hoc tests were conducted to compare the group's scores from the pre-test to the post-tests. The post-hoc Tukey test showed that the SI group and the SI+SO groups were equal (p= .921) and significantly different than the SO (p =. 000).

To investigate possible delayed effects, a second ANOVA was conducted on the raw scores of the two discourse-level interpretation post-tests. The results showed only a significant main effect for Treatment (F (2,30) = 77.402, p< .000). The post-hoc Tukey test indicated that the SI group and the SI+SO groups again equal (p =. 545) and significantly different than the SO (p =. 002).

The results from the discourse-level interpretation task were similar to the one obtained in study 1 and demonstrated that only the SI and SI+SO groups gained in their ability to interpret English causative forms presented at the discourse-level. The SI and SI+SO groups maintained these gains over a period of three weeks.

4.5. Discourse-level production data (Study 2)

The one-way ANOVA conducted on the pre-test scores showed no significant differences between the three groups before instruction (F (2,30) = 16.456p = .210). Any differences found after instruction will be attributed to the effects of instruction. Table 5 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 production task in the immediate post-test and delayed post-tests. The SI group and the SI+SO groups have improved from pre-test to post-test scores. There was no improvement for the SO group.

TABLE 5. 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 (2,30) = 124.636, p< .000); a significant main effect for Time (2,30) = 86.669, p< .000; and significant interaction between Treatment and Time F (2,30) = 45.162, p< .000.

Given the significant main effect for instructional treatment post-hoc tests were conducted to compare the group's scores from the pre-test to the post-tests. The post-hoc Tukey test showed that the SI group and the SI+SO groups were equal (p= .829) and significantly different than the SO (p =. 000).

To investigate possible delayed effects, a second ANOVA was conducted on the raw scores of the two discourse-level production post-tests. The results indicated a significant main effect for Treatment (F (2,30) = 14.362, p< .000). A post-hoc Tukey test conducted on the treatment revealed the following contrasts: SI group and the SI+SO groups were equal (p =. 998) and significantly different than the SO (p =. 001).

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

5. Summary of findings

The first research question was: What are the immediate and delayed effects of SI, SO and SI+SO, on the acquisition of the English causative form as measured with sentence-level interpretation and production tasks? The results of interpretation sentence-level tasks in the first study clearly indicated that SI and a combination of SI+SO helps learners to process the English causative forms correctly and appropriately. The improvement found for the SI and SI+SO groups was maintained three weeks after instruction. These instructional groups were more effective than the SO and Control groups in interpreting the target form embedded in a sentence. The results of the production sentence-level task indicated that the three instructional

groups (SI, SO, and SI+SO) equally improved from pre-tests to post-tests and were statistically better than the Control group. The advantage found for these three groups was maintained for a period of three weeks after the end of the instructional treatment.

The second research question was: What are the immediate and delayed 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 interpretation discourse-level tasks in the first and second study clearly indicated that SI and a combination of SI+SO are more effective than the SO and the control group in interpreting correct English causative forms at discourse level. The improvement found for the SI and SI+SO groups was maintained three weeks after instruction. The results of the production discourse-level task in study 2 echoed the one obtained for the interpretation discourse-level tasks in study 1 and 2 and indicated that the two instructional groups (SI and SI+SO) improved from pre-tests to post-tests, whereas the SO group did not. The advantage found for these two groups was maintained over a three-week period.

6. Discussion and Conclusion

The overarching question of this study was to measure the relative effects of SI and SO in isolation and in combination using both sentence and discourse-level interpretation and production tasks. The findings from the interpretation sentence-level task (study 1) provide further empirical support for the view that SI is better than SO (also used in MOI studies and reviewed earlier in this paper) 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 evidence to support the view that SI alone is sufficient to improve learner's performance as the SO only groups made no gains in the interpretation sentence-level task. Similar results were obtained for the interpretation discourse-level task. In both study 1 and study 2, the SI group and SI+SO group outperformed the SO group providing further evidence of the impact of SI on input processing. One possible explanation of these results might be found in the nature and purpose of SI activities. As outlined by Wong (2004:35) SI ''push learners to abandon their inefficient processing strategies for more optimal ones''. SI significantly improves learners' interpretation of both sentence and discourse containing English causative forms. The SI activities learners received were at sentence-level. However, learners in the SI group were also able to interpret the target forms at discourse-level. 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 task. In both interpretation tasks the SI group and the SI+SO group were able to maintain the gains made over a three weeks period.

The findings from the production sentence-level task in study 1 are in line with previous research findings in the PI/SI research agenda (Benati, 2004a, 2005; Benati and Lee, 2008; see also Lee and Benati, 2009; and Benati and Lee, 2015). The SI group, SO group and SI+ SO group performed equally. The evidence obtained in the sentence-level production task further suggests that SI not only have an impact on the way learners interpret sentences but also on the way that learners produce sentences containing the target form. SI has clearly altered the way learners processed input and this had an effect on their developing system and subsequently on what learners could access for production. These findings are in line with the original study conducted by

VanPatten and Cadierno (1993:240). ‘‘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. The main findings from the discourse-level production task (study 2) indicate that only the SI and SI+SO groups made statistically relevant gains from pre-test to post-tests and those gained were maintained over a period of three weeks. SO alone was not enough to provide learners with the ability to produce discourse containing the target feature.

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 the two experimental studies confirm the key role of SI as an effective pedagogical intervention designed to alter processing problems such as the First Noun Principle. SO is not successful in bringing about similar effects to those brought about by SI in interpretations tasks at both sentence and discourse-level. 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).

Secondly, SI is the causative variable for the change in performance of the groups in study 1 and 2. Not only SI is effective in developing learners’ ability to process input (at sentence and 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: 100) argued ‘‘ it assists in the developing of underlying knowledge that can be tapped during the development of skill’’.

Thirdly, the results from this study confirm the durative effects of SI over a period of three weeks.

Fourthly, the findings from this study lend support to a number of hypotheses formulated with the PI research framework (Benati and Lee, 2008: Benati and Lee, 2010). The positive results obtained in both studies with two different populations, lend support to the so-called *Age Hypothesis* (Benati and Lee, 2008:168) ‘‘PI will be just as effective an intervention with young learners as it is with older learners’’. Our results are based on adults (age 18-20) and young learners (age 10-12) from two different L1s and confirmed previous research (Lee and Benati, 2013) demonstrating the effectiveness of structured input tasks no matter the age.

The findings from this study in using native speakers of other languages than English, also support the so-called *Native Language Hypothesis* (Benati and Lee, 2008:166) ‘‘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 and Chinese speakers 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 small number of participants in study 2 and the lack of a control group (due to attrition and other selection issues) is something that needs to be addressed in a replication if findings from this study can be generalized. Although,

durative effects were measured by using delayed post-tests in the design of this study, further research should investigate long-term effects. This study measured the effects of SI and SO on the interpretation and production discourse-level tasks. Further research would need to continue to measure a variety of instructional interventions on discourse-level tasks and more spontaneous production tasks that include time pressure and do not allow learners to monitor their responses. Further research should also consider the role of structured-input and structured-output tasks with or without explicit information.

Acknowledgements

We would like to thank all the students who participated to the two studies for helping us in the collection of the data in the two studies. We also express our gratitude to Bill VanPatten, James Lee and three anonymous reviewers for their valuable comments and suggestions.

7. References

Benati, Alessandro. 2001. A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research* 5. 95-127.

Benati, Alessandro. 2004a. The effects of structured input and explicit information on the acquisition of Italian future tense. In Bill VanPatten (ed.). *Processing instruction: Theory, research, and commentary*, 207-255. Mahwah, NJ: Erlbaum.

Benati, Alessandro. 2004b. The effects of processing instruction and its components on the acquisition of gender agreement in Italian. *Language Awareness* 13. 67-80.

Benati, Alessandro. 2005. The effects of PI, TI and MOI in the acquisition of English simple past tense. *Language Teaching Research* 9. 67-113.

Benati, Alessandro & James F. Lee. 2008. *Grammar acquisition and processing instruction: Secondary and cumulative effects.* Bristol: Multilingual Matters.

Benati, Alessandro & James F. Lee. 2010. *Processing instruction and discourse*. London: Continuum.

Benati, Alessandro & James Lee. G. 2015. *Processing instruction: New insights after twenty years of theory, research and application*. Special Issue in *IRAL*.

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

Cheng, An-Chung. 2004. Processing instruction and Spanish *ser* and *estar*: Forms with semantic-aspectual value. In Bill VanPatten (ed.). *Processing instruction: Theory, research, and commentary*, 119-141. Mahwah, NJ: Erlbaum.

Farley, Andrew. P. 2001. The effects of processing instruction and meaning-based output instruction. *Spanish Applied Linguistics* 5. 57-94.

Farley, Andrew. 2004. The relative effects of processing instruction and meaning-based output instruction. In Bill VanPatten (ed.). *Processing instruction: Theory, research, and commentary*, 143-168. Mahwah, NJ: Erlbaum.

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

Gely, Anne. 2005. Output-Based Instruction versus Processing Instruction on the acquisition of the French imperfect tense. Unpublished Master's thesis, University of Greenwich, London.

Keating, Greg & Andrew P. 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, Rachel. 2013. The effects of processing instruction with and without output: Acquisition of the Spanish subjunctive in three conjunctional phrases. *Hispania*, 96, 153-169.

Lee, James & Bill VanPatten. 1995. *Making communicative language teaching happen.* New York: McGraw-Hill.

Lee, James F. & Alessandro Benati 2007a. *Delivering processing instruction in classrooms and virtual contexts: Research and practice*. Equinox, London.

Lee, James F. & Alessandro Benati. 2007b. *Second language processing: An analysis of theory, problems and possible solutions*. Continuum: London.

Lee, James F. & Alessandro Benati. 2009. *Research and perspectives on processing instruction*. Berlin: Mouton de Gruyter.

Lee, James F. & Alessandro Benati. 2013. *Individual differences and processing instruction*. Equinox, London.

Morgan-Short, Kara & Harriet 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, Anne. 2011. The effects of a combined output and input-oriented approach in teaching reported speech. *Research in Language, 9.2*.

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

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

VanPatten, Bill. 2002. Processing instruction: An update. *Language Learning, 52,* 755-803

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

Van Patten, Bill. 2015a. Input processing in adult SLA. In B.VanPatten & J. Williams, J. (Eds.), *Theories in second language acquisition2nd edition* (pp.113-135). New York: Routledge.

VanPatten, Bill. 2015b. Foundations of processing instruction. *IRAL*, 53, 91-109.

VanPatten, Bill & Teresa Cadierno. 1993. Explicit instruction and input processing. *Studies in Second Language Acquisition* 15. 225-243.

VanPatten, Bill & Soile Oikennon. 1996. Explanation vs. structured input in processing instruction. *Studies in Second Language Acquisition* 18. 495-510.

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

VanPatten, Bill, Jeffrey Farmer & Caleb Clardy. 2009. Processing instruction and meaning-based output instruction: A response to Keating and Farley (2008). *Hispania* 92. 116-126.

VanPatten, Bill & Alessandro Benati. 2015. *Key Terms in SLA* (2nd edition). London: Bloomsbury.

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

Pre-tests	Study 1 Sentence-level (Interpretation and Production) Discourse Level (interpretation) 1week before instruction	Study 2 Discourse-level (Interpretation and Production) 1week before instruction
Instructional treatments	SI SO SI+SO Control group (Study 1) 2 consecutive days (3 hours)	

Post-tests	Sentence-level (Interpretation and Production) Discourse Level (interpretation) Immediate after instruction	Discourse-level (Interpretation and Production) Immediate after instruction
Delayed Post-tests	Sentence-level (Interpretation and Production) Discourse Level (interpretation) Three weeks after instruction	Discourse-level (Interpretation and Production) Three weeks after instruction

Figure 1. Overview of the two studies

Listen to the sentences and answer the questions. Pay careful attention to the structure of each sentence in order to understand **who** is actually performing the action:

1)	Who mended the dress?	a) Jane	b) someone else
2)	Who repaired the car	a) Penny	b) someone else
3)	Who delivered the flowers?	a) Emma	b) someone else
4)	Who got dirty with pink paint?	a) Mary	b) someone else
5)	Who took a test?	a) Tom	b) someone else

6)	Who had a loyal dog?	a) Jim	b) someone else
7)	Who invited all the people to the party?	a) Peter	b) someone else
8)	Who tuned the guitar?	a) Bill	b) someone else
9)	Who had played the game before?	a) John	b) someone else
10)	Who served breakfast?	a) Alex	b) someone else

Instructor's script

1) Jane had her dress mended last Monday
2) Penny repaired the car herself
3) Emma had flowers sent to her boyfriend
4) Mary had the fence painted pink
5) Tom took a driving test
6) Jim had a very loyal dog when he was younger
7) Peter invited everyone to the party
8) Bill had his guitar tuned before the concert
9) John played the game often so he knew the rules well
10) Alex had breakfast served for him and his girlfriend

Figure 2. Sample of an aural SI activity

The ideal flat mate.

Step 1. Write at least five house works that you like doing and you always volunteer to do.

"I always take the dog for a walk. I never miss a chance for a walk at the park"

Step 2. Write at least five house works that you hate doing and you would never miss a chance to have them done by someone else. Follow the structure given in the example below.

"I always have the dishes done by my sister. I hate this chore"

Step 3. Now discuss your sentences with a partner. Your partner will subsequently present his/her preferences to you. Compare and contrast similarities and differences. Are you a compatible pair? Could you be ideal flat mates?

Step 4. Make a grid on the board, a volunteer should fill it out based on what was said (things I do myself vs. things I ask someone else to do for me). Now it's highly possible that you can all find the ideal flat mate in our classroom. Present him/her to the class and explain why you think he/she would be an ideal flat mate for you.

Figure 3. Sample of an SO activity

	Pre-test	Post-test	Delayed

Groups	*n*	Mean	SD	Mean	SD	Mean	SD
SI	13	.9231	.7595	6.153	.6887	6.076	.6405
SO	15	.9333	.5936	1.000	.5345	.6000	.6324
SI+SO	16	.5000	. 6324	5.500	.7303	5.187	.6551
C	10	.7000	.6749	.4000	.5164	.3000	.4830

Table 1. Study 1- Interpretation sentence-level task (descriptive statistics)

		Pre-test		Post-test		Delayed	
Groups	*n*	Mean	SD	Mean	SD	Mean	SD
SI	13	1.000	.7071	4.153	.80063.969		.5991
SO	15	.6333	.5936	5.533	.8338	4.866	.7432
SI+SO	16	.7500	. 7746	5.475	.8062	5.100	.6311
C	10	.2000	.4216	.3000	.4830	.4000	.6762

Table 2. Study 1- Production sentence-level task (descriptive statistics)

		Pre-test		Post-test		Delayed	
Groups	*n*	Mean	SD	Mean	SD	Mean	SD
SI	13	.5385	.5188	4.769	.83204.538	.7762	
SO	15	.7333	.5676	.600	.6324	.3333	.4879
SI+SO	16	.4347	. 5123	5.375	.6191	5.000	.6324
C	10	.3000	.4830	.2000	.4216	.4000	.6992

Table 3. Study 1- Interpretation discourse-level task (descriptive statistics)

		Pre-test		Post-test		Delayed	
Groups	*n*	Mean	SD	Mean	SD	Mean	SD
SI	10	0.750	0.700	7.000	1.314	6.400	.6263
SO	10	0.745	0.631	1.087	0.450	0.850	.5555
SI+SO	10	0.735	0.340	7.200	1.341	6.500	.3562

Table 4. Study 2 - Interpretation discourse-level task (descriptive statistics)

		Pre-test		Post-test		Delayed	
Groups	*n*	Mean	SD	Mean	SD	Mean	SD
SI	10	.1500	0.232	2.700	0.653	2.300	0.525
SO	10	.1000	0.222	.2255	0.452	.0155	0.443
SI+SO	10	.3333	0.300	2.500	0.341	2.400	0.500

.

Table 5. Study 2 – Production discourse-level task (descriptive statistics)