

Processing instruction effects regardless of input modality and developmental processing constraints? A school lab classroom study on the morphosyntactic acquisition of L2-English

Tanja Angelovska and Dietmar Roehm

Abstract

We investigated whether Austrian L2-English learners would benefit more from written or auditory processing instruction (PI) on the third person singular -(e) s tense form. The instruction and all three tests (pre-test, immediate post-test and delayed post-test two weeks after the instruction) were conducted in school lab classrooms. Using accuracy scores and reaction times for the interpretation tasks and accuracy scores for a gap-filling task, this study extrapolates the instructional effects according to modality type. We aimed to find out whether (a) the modality of the PI affects its efficacy regardless of the imposed developmental processing constraints of the target feature and how (i.e. resulting in faster or slower and more or less accurate processing); (b) whether any differences in gains could be attributed to the modality of the assessment task; and (c) whether any gained positive effects of instruction (regardless of modality) are maintained over time.

KEYWORDS: L2 PROCESSING; PROCESSING INSTRUCTION; INPUT PROCESSING; MORPHOSYNTAX; MODALITY

Affiliations

Tanja Angelovska: University of Salzburg, Austria.

email: tanja.angelovska@sbg.ac.at

Dietmar Roehm: University of Salzburg, Austria.

email: dietmar.roehm@sbg.ac.at

Introduction

Current rapid technological developments mean that foreign language learners have more opportunities and access to a vast amount of auditory and written input that can significantly enhance the learning process beyond the classroom (Reinders and Benson 2017). Online and e-learning modules, courses and autonomous learning packages now offer learners greater exposure to listening components and auditory authentic material than the currently available traditional schoolbooks do. Nevertheless, we still do not know much about which input (auditory or written) modality is more beneficial for the acquisition of L2 morphosyntactic features.

The present study examines the effects of auditory and written modality types of processing instruction (PI) on attainment and retainment of the third person singular *-(e)s* tense form (the target feature on which learners received instruction) regarding both comprehension and production. Processing instruction is an input-based pedagogical intervention, which helps learners to correctly map form and meaning through modifying the input in such a way that it corrects L2 learners' default non-optimal strategies. It has proved effective (Benati 2001; Cadierno 1995; VanPatten and Cadierno 1993) for a variety of features affected by the principles of input processing (VanPatten 2004, 2015) in a number of target languages. Only two PI studies have dealt with the effects of PI in relation to the modality of the PI training so far. Thus, we tested (a) whether the modality of the PI training affects its efficacy regardless of the imposed developmental processing constraints of the target feature and how (i.e. resulting in faster or slower and more or less accurate processing); (b) whether any differences in gains could be attributed to both the modality of instruction and of the assessment task; and (c) whether any gained positive effects of instruction are maintained over time.

Background and motivation

The processability theory and the third person singular *-(e)s*

The acquisition of the third person singular *-(e)s* tense form is known to be one of the most challenging morphosyntactic features in second language acquisition (henceforth, SLA), and as such it has also been addressed within processability theory (Pienemann 1998), which relates processing effects to grammatical development in SLA. For example, according to this theory, in order to acquire *-(e)s*, L2 learners must be able to exchange grammatical information across the noun and verb phrase in a sentence, which poses

acquisitional difficulties due to the exchange happening between two different types of sentential phrases. In line with the processability theory, a concrete hypothesis referring to the aspect of teachability was formulated, stating ‘the effect of teaching intervention is constrained by the learner’s current state of development’ (Pienemann 2015:137). This idea originated from older research accounts supporting the so-called non-interface position in SLA, according to which language acquisition is impossible if the teaching is not in line with learners’ natural sequences of acquisition (Hahn and Felix 1985; Wode 1981).

Even though certain linguistic structures and forms underlying the IP principles have been attested as marking features of L2 learners’ developmental stages (Pienemann 1998), there is only one study so far that has tested directly learners’ readiness in L2-Spanish (see Farley and McCollam 2004). They found that learners did indeed interpret and produce the form (subjunctive) for which they were classified as developmentally unready. Consequently, learners need to be developmentally ready for acquisition (i.e. they need to have reached stage 5). For these reasons, we chose the third person *-(e)s* tense form as a primary target feature for this study, in order to find out whether PI independent of modality (written or auditory) would override the imposed constraints on learners who are not developmentally ‘ready’ (Pienemann 1989:61) and would help to accelerate the acquisition process.

Processing instruction and the third person singular *-(e)s*

Processing instruction is an input-based instructional intervention aimed at instilling the appropriate processing strategies through sufficient practice with Structured Input Activities (henceforth, SIAs), so that L2 learners can successfully map form and meaning without having to focus on (mechanical) production. The SIAs are created by taking into consideration learners’ default and non-optimal processing strategies. This kind of practice helps L2 learners to process morphosyntactic grammatical forms in the input more efficiently than training that is more traditional. Processing instruction is informed by VanPatten’s model of input processing (henceforth, IP) (VanPatten 1996, 2004, 2014, 2015). This model explains what L2 learners do when they are processing input. With two main principles and several sub-principles, it describes the mapping of form and meaning during real-time comprehension or ‘the moment-by-moment computation of sentence structure’ (VanPatten 2015:92). For the current study, the first principle (the Primacy of Meaning Principle:

learners process input for meaning before they process it for form) and four of its sub-principles are relevant.

- Principle 1a – the Primacy of Content Words Principle: learners process content words in the input before anything else.
- Principle 1b – the Lexical Preference Principle: learners tend to rely on lexical items as opposed to grammatical form to get meaning when both encode the same semantic information.
- Principle 1c – the Preference for Non-redundancy Principle: learners are more likely to process non-redundant meaningful forms before they process redundant meaningful forms.
- Principle 1d – the Meaning-Before-Non-meaning Principle: learners are more likely to process meaningful grammatical forms before non-meaningful forms irrespective of redundancy.

Taken literally, in a present simple tense English sentence (*John walks his dog every day*), the comprehender will comprehend the form-meaning connection through the available content words (e.g. the adverbial phrase *every day*) before processing the grammatical morpheme *-(e)s*. Also, the third person singular *-(e)s* tense form in English semantically conveys the meaning that the agent is neither the person speaking nor the persons spoken to, in which case there is also an agreement in number between the (pro)noun and the verb. Hence, this agreement helps to figure out the agent (meaningful) and the grammatical information (non-meaningful and redundant). Redundancy here affects only the *-(e)s* and not the agent (e.g. *John* is the person).

The first PI study by VanPatten and Cadierno (1993) has motivated a growing body of research demonstrating the benefits of PI. In the field of PI, two studies (Benati, Lee and Houghton 2008; Mavrontoni and Benati 2013) have investigated the effects of PI on the acquisition of the English third person singular *-(e)s* form, either as a primary or as a secondary target feature. Mavrontoni and Benati (2013) compared the effects of PI and traditional instruction (TI) on the processing of *-(e)s* (primary target feature). They recruited school-age children with Greek as their L1 from two different age groups (aged 8–10 and aged 15–17). In total, thirty-four participants were assigned randomly to the PI and the TI within their age groups. Twenty students were included in the first age group (8–10 years), and fourteen students were included in the second age group (15–17 years). A pre-/post-test design was used, in which the pre-test was administered one week before the intervention. Both groups received the same explicit information and the same number of tasks. The PI group received referential

and affective SIAs, and the TI group received mechanical output-based activities. The instructional period lasted for a total of three hours on two consecutive days, followed only by an immediate post-test, consisting of an interpretation and a production task. The sentence-level comprehension task included twenty items (ten of which were target items). The subjects listened to the sentences and ticked whether the action described in the sentence was performed by one agent (third person singular) or by multiple agents (third person plural). The written production task focusing on describing the daily routine of a person was a gap-filling task. Participants receiving the PI treatment scored significantly higher in the interpretation task than the participants receiving TI. The PI subjects scored equally well on the production task as the subjects from the TI group. In addition, it was found that PI is effective for both age groups. Results are clearly in line with other PI studies (Angelovska and Benati 2013; Benati and Angelovska 2015; Benati 2001; Benati, Lee and Houghton 2008; VanPatten and Cadierno 1993; VanPatten and Wong 2004).

Benati, Lee and Houghton (2008) compared the effects of PI and TI on the processing of the English past tense (*-ed*), and the English third person singular *-(e)s* served as a secondary target item. The aim of the research was twofold. First, to examine the primary effects of PI and TI on the acquisition of the English past tense plus the transfer-of-training effects of the PI and TI treatment on the acquisition of *-(e)s*. These authors recruited eighty-one students (14 years old) with Korean as their L1 studying English in beginning-level classes. This original data pool was reduced to twenty-six, as the excluded participants scored higher than the selection criterion of 50% in the pre-test battery. The selected participants were assigned randomly to either the PI (structured-input activities) or the TI (output-based and production-oriented tasks) group. The number of tasks was equal in both groups and the instructional period lasted four hours. Both groups received the same treatment. The PI group was given further information about useful processing strategies to connect form and meaning. The TI group did not receive any additional information. The tests consisted of an interpretation and a production task for both the primary and the secondary target item. The results showed that the PI group significantly outperformed the TI group in the interpretation task. In the production task, both groups improved equally. With regard to the transfer-of-training effects, the PI group did indeed reveal a significant improvement in the interpretation, as well as in the production task. Benati, Lee and Houghton (2008) interpreted this effect as a result of the PI training, which apparently not only alters the non-optimal default processing strategies of the target

feature being trained, but also affects similar features affected by the same processing principle.

However, neither of the two existing PI studies measuring the instructional effects on the acquisition of the third person singular *-(e)s* (Mavrontoni and Benati 2013 – on primary effects; Benati, Lee and Houghton 2008 – on secondary effects) extrapolates the improvement effects from the type of modality. Moreover, both studies included rather small intervention groups. Hence, as suggested by Mavrontoni and Benati (2013), replication with a larger sample is needed, in order to gain robust results. The current study focuses on the effects of PI in the processing of the English third person singular present simple tense as a primary target feature, and attempts to address the existing gaps with regard to size and the measurement of long-term retention effects. In addition, it extrapolates these effects from the type of modality in which the intervention was delivered (auditory or written).

Effects of modality in L2 acquisition

The interest in input modality springs from existing research into cognitive psychology dealing with modality effects on attention and the impact modality has on memory and on the organisation of information. For example, Dennis (1977) found that the recall of verbal stimuli in concurrent tasks was easier for participants than it was for auditory stimuli, and was thus in support of Treisman's model of attention (1969). According to Treisman, two concurrent tasks in the same modality might cause perceptual interference, as both tasks require access to the same analyser system (1969:288). Thus, attention is limited if participants must use one shared analyser for two different input modalities – as has been the case in classical PI designs, where, for example, learners listen to a sentence and at the same time respond by making a choice presented in another modality (written sentences or pictures). When investigating the participants' organisation and report of words, Penney (1980) found that recall by the modality of presentation outperformed serial recall significantly, and auditory stimuli had a higher recall than visually presented stimuli. This auditory superiority effect in a serial recall is extremely robust, as it has been confirmed in a variety of studies on memory and language (Beaman 2002; Cowan, Saults and Brown 2004; Frankish 1985; Frick 1989; Penney 1980).

The number of SLA studies investigating whether the modality of the input influences the processing of language is rather limited. The earliest study is VanPatten's study (1990) using only auditory mode and examining whether learners can consciously attend to meaning and form. Participants

were randomly assigned to one of four groups listening for (1) content, (2) a keyword and content, (3) the definite article *la* and (4) the third person plural morpheme *-n*. The subjects were asked to note down every instance of the linguistic target item they noticed and to write free written recalls in their L1 (English). The results showed an overall decrease in comprehension by group 3 and group 4. Thus, when subjects were required to attend to linguistic forms with no communicative value (article *la*, morpheme *-n*), it was difficult for them to simultaneously focus on the meaning of the content. VanPatten concluded that the conscious attention to grammatical forms with no communicative value negatively affects the comprehension of input (1990:294).

A replication of VanPatten (1990) with a changed input mode (changed from auditory to written) confirmed the original results, showing that learners' comprehension of written input was negatively affected if they were required to simultaneously attend to specific linguistic items with low communicative value (Greenslade, Bouden and Sanz 1999). Wong (2001) achieved a partial replication of VanPatten (1990), directly comparing the auditory and written modes within the same pool of participants, while in VanPatten (1990) participants were only presented with auditory input. Wong's (2001) findings regarding the auditory mode paralleled those found in VanPatten's original study (1990), but they contradict Greenslade, Bouden and Sanz (1999). With regard to the written mode, no significant differences between the conditions (attending to the meaning or to the form) could be found. Wong concluded that 'learners' limited attentional capacity is not constrained in the same way during input processing in the auditory and written modes' (2001:358).

This contradiction pushed for further research, adding important findings to the role of modality without any specific focus on a morphosyntactic feature (Hama and Leow, 2010; Kerz, Wiechmann and Riedel 2017; Lund 1991; Park 2004; Williams 2005). These SLA studies confirmed the assumption that the auditory presentation of stimuli has a more challenging nature than the written presentation. For example, Lund (1991) compared the listening and reading comprehension of L2-German learners at three different levels. In all three levels, one section of the subjects received the text in the auditory mode, and the other section of the subjects received the text in the written mode. The results showed that participants working in the written condition recalled more specific information and details than those working in the auditory condition. The subjects experiencing the auditory condition, however, remembered more main ideas than those experiencing the written condition. Lund also found that participants with lower proficiency benefited less from listening than from reading. Similarly,

Park (2004) compared L2 listening and reading comprehension and found that participants performed better on the reading than on the listening comprehension. He concluded that the written mode is less demanding on the students' attentional capacities. However, these studies used offline measures or measures that do not tap into learners' processing capacities moment by moment, that is, in real-time.

Two recent PI studies (Wong and Ito 2018; Ito and Wong 2019) employed an eye-tracking experiment to investigate whether learners benefit from the consistency of input modality. Wong and Ito (2018) compared the effects of PI and TI on L2 online processing of the French causative construction and examined the role that explicit information might play in two experiments. In experiment 1, neither group received explicit information, and in experiment 2 both experimental groups received explicit information. The input modality was not consistent, and participants read sentences on a screen in the training phase and listened to sentences through headphones in the testing phase. Their results are discussed in relation to accurate picture selection and changes in gaze, which were expressed in logit, or the log of the ratio between the numbers of fixations on the two areas of interest. For our study, only the results from experiment 1 are relevant. They showed that the PI group had significantly higher accuracy scores than the TI group, with an observed change in the eye-movement pattern after training for the PI, but not for the TI group. Ito and Wong (2019) replicated Wong and Ito (2018). Ito and Wong (2019) changed the input modality, so that auditory PI training was consistent in the practice and the testing stages. The experiment consisted of pre-training eye-tracking, auditory PI training and a post-training eye-tracking block (Ito and Wong 2019:10). In the eye-tracking block, participants were required to listen to the full sentence and then select the picture that matched the sentence. Pictures were presented, and after 1.5 seconds a sentence was played. Selection accuracy and fixations on correct and incorrect pictures were used for the data analysis. Although their results indicate a significant improvement in the picture selection accuracy after the auditory PI training, the effects did not differ from those of the written PI training. However, the increase in fixations on the correct picture was slower for the PI auditory group than for the PI written group, possibly because subjects trained with auditory PI needed longer to overcome the non-optimal default processing strategies than those trained using written PI – in line with previous findings (Hama and Leow 2010; Kerz, Wiechmann and Riedel 2017; Lund 1991; Park 2004).

Research questions and hypotheses

Using accuracy, reaction times (RTs) and production data, we aimed to find out whether the modality of PI training affects its efficacy and how (i.e. resulting in faster or slower and more or less accurate processing). We further attempted to investigate whether any differences in gains could be attributed to the modality of the assessment task, and whether any gained positive effects of instruction are maintained over time. Based on Ito and Wong (2019), who found that subjects trained using auditory PI needed longer to overcome the non-optimal default processing strategies than those trained using written PI, we proposed the following four hypotheses concerning modality.

- H1: the A-group will show reliably lower accuracy scores on both interpretation tasks (auditory and written) than the W-group.
- H2: the A-group will display longer RTs on both interpretation tasks (auditory and written) than the W-group.
- H3: the W-group will show higher accuracy scores on the production task than the A-group.
- H4: both W-group and A-group will retain the positive effects of PI on interpreting and producing the target feature *-(e)s*.

Experiment

Procedure

Various Austrian schools in Salzburg and its region equipped with technically adequate school labs were contacted and sent a call for participation. Headmasters and English as a foreign language (EFL) teachers were informed about the aims of the study. An agreement was made with the EFL teachers willing to recruit pupils from their classes that the target feature under consideration should not be instructed during the project duration, in order to eliminate possible side effects from regular EFL instruction influencing the effects of the intervention. Parents were informed about the aim and the procedure, and written consent forms for the participation of their children were obtained. The required permissions from the Educational Directorate of Salzburg and from the Ethical Committee of the University were issued.

A pre-test/post-test design was used. The pre-test was administered five days before the intervention. The intervention lasted seventy minutes on two consecutive days. The immediate post-test was administered

immediately after the intervention. The delayed post-test was administered two weeks after the intervention. The instruction and all three tests (pre-test, immediate post-test and delayed post-test two weeks after the instruction) were conducted in school lab classrooms equipped with computers meeting adequate technical requirements and installed software to assess learners' reaction times in parallel with accuracy data. Two researchers familiar with PI conducted the instruction. To guarantee the anonymity of the participants, the data were anonymised.

Method

Participants

L2 beginning learners of English attending first and second grades at Austrian lower secondary schools were recruited for the study. The original pool of 202 participants was reduced to eighty-nine students after employing the selection criterion of 60%. To be included in the final data pool, participants had to score 60% or lower on the pre-test battery. They were then randomly assigned to either the W-group, receiving written SIAs only ($n = 35$), the A-group, receiving auditory SIAs ($n = 34$), and the C-group, control group ($n = 20$), receiving no instruction.

Materials

Intervention package

The instructional materials designed for the A-group and the W-group were the same. The only difference between the groups was the type of modality used during the intervention. Throughout the experiment, participants from both groups received no explicit information on the target third person singular *-(e)s* tense form or on the distractor regular past simple tense *-ed* structure. They were only given brief information about the processing strategy or what they should not do ('Do not pay attention to the temporal adverbial to derive meaning'). The sentences for the PI auditory training were spoken by a female native speaker of American English and recorded at a sampling rate of 44.1 kHz. The sound files were normalised and cut using the software Praat (Boersma and Weenink 2014 [1992]). The material was designed by following the guidelines suggested in Lee (2014) and Wong (2004).

Subjects received two types of SIAs, namely referential and affective activities. In the referential SIAs, the subjects read or heard ten sentences – five target sentences with *-(e)s* and five distractors with *(-ed)* for each activity – and they had to decide whether the action described in the sentences

takes place regularly (e.g. *every day*) or if it took place in the past (e.g. *yesterday*). The sentences did not include any time adverbial. After each activity, the students were provided with a solution sheet to check their answers. The affective SIA consisted of ten sentences including the target form *-(e)s*. Subjects were supplied with binary options to give their personal response (e.g. *Me, too/I don't do that* or *interesting/boring*). Overall, the students from both groups completed six referential and three affective SIAs. The instructional and assessment package was adapted from Mavrontoni and Benati (2013).

Assessment tests and scoring

The performance was assessed using three different tests. Each test had the same structure and consisted of one written interpretation task, one auditory interpretation task and one written production task. The interpretation tasks were programmed with the software PsychoPy 3.0.3 (Peirce *et al.* 2019), which enabled us to receive accuracy scores and reaction times for them. The written and auditory interpretation tasks were sentence-level interpretation tasks. All the sentences followed the same pattern: 'name' + 'a transitive verb' + 'an indefinite article' + 'a noun' (e.g. *Peter drinks a lemonade*). The target sentences only included verbs ending in the phoneme /s/, so that only verbs ending in the voiceless plosives /k/, /p/ or /t/ were included (e.g. *drink, stop, get*). As only the unvoiced /s/ sound is common in the German language, this enabled us to avoid any negative effects on the participants' reaction times due to unfamiliarity with other sounds (e.g. voiced /z/ and the /tʃ/ sound). We used only familiar words, common in the schoolbooks they were using. All instructions were provided in German to ensure that students could follow them and complete the tasks properly. The instructions were displayed at the top of the screen, written in white letters on a grey background. Participants had to decide when the action took place by pressing one of the three keyboard buttons labelled *regularly*, *in the past* and *I don't know*. The experiment started with a test trial including three sentences. After the test trial, the actual experiment, including ten target sentences and ten distractors, started. To avoid any negative presentation effects, all stimuli were fully randomised. Participants used colour-coded (one colour representing one of the three answer options) arrow keys, which were operated with only one hand. Thus, any negative effects due to right- or left-handedness could be avoided. The sentences were not timed. As soon as the participants pressed one of the colour-coded arrow keys, the next sentence appeared on the screen. Since only the target sentences were scored, the maximum score for this task was ten points. Each correct answer equalled one point, and each incorrect answer

equalled zero points. The *I don't know* answers were not included in the analysis.

In the auditory interpretation task, participants listened only once to twenty sentences presented one by one. The sentences were recorded in a sound studio and spoken by the same female native speaker as the sentences in the training material and under the same conditions as the written interpretation task. Each stimulus was preceded by a pause of 100 milliseconds (ms) and followed by a pause of 300 ms. This can be regarded as an artefact of the experimental design, in which the stimulus and the three answer options appeared on the screen at the same time only for the written assessment task, whereas for the auditory assessment task the stimulus was played (while an empty grey background was shown) and then another screen window with the three answer options appeared. Because reaction times were always calculated relative to the onset of the answer cues, the auditory assessment task showed systematic shorter reaction times. Ten target sentences describing regular actions and ten distractors (*-ed*) with past events were randomly played. The students listened to the sentences and decided if the action described in the sentences takes place regularly. Participants were required to choose one of the following three possible answers (*yes, no, I don't know*) to respond to the question (*Does the action take place regularly?*). The scoring procedure was the same as for the written interpretation task.

The written production task was an open cloze text consisting of ten target items and ten distractors. The students were asked to fill in the gaps with the correct verb form. As in the study conducted by Angelovska (2015), the written production task did not include any information on the verb form that should be used to avoid framing the students' answers in any way. The students had no time pressure when filling in the blanks. The number of proper names and the number of personal pronouns used in the target sentences were distributed equally, in order to avoid any negative effects due to the corresponding nouns or pronouns. The participants could attain a maximum of ten points in the written production task, as only the target forms were scored. For each correct target form (*-(e)s*), one point was given, for each incorrect answer or for no answer, zero points were given. No partial points were awarded.

Data analysis

Accuracy data scores for the written and the auditory interpretation task (twenty items each, of which ten were targets) and the written production task (twenty gaps, of which ten were targets) from the three groups (W,

A, C) were included in the ANOVA analysis. Data were collected at three different test points: pre-test (T1), immediate post-test (T2) and delayed post-test (T3). In addition, RTs data from the written interpretation task and the auditory interpretation task were analysed. Reaction times were trimmed to minimise the effects of outliers. Outliers that were more than 2.5 standard deviations (*SDs*) above the mean were replaced with the value 2.5 *SD* above the mean RTs per condition per subject, which affected 2.65% of all data.

In total, we included data from eighty-nine subjects (n for A = 34, n for W = 35, n for C = 20). Data from eleven subjects had to be excluded from the ANOVA analysis due to missing test points. Therefore, for the final analysis we had seventy-eight subjects in the following group constellations: A = 30, W = 31, C = 17.

The reported analyses were conducted using the R software (R Development Core Team, 2018). Mean accuracy scores and RTs were each subjected to a repeated-measures ANOVA, including the within-subject factors TEST (T1, T2, T3), MODE (written/auditory task) and the between-subject factor INT (W-, A-, C-group). All statistical analyses were carried out hierarchically; that is, only significant interactions ($p < 0.05$) were resolved and only significant effects have been reported. For post hoc planned comparisons, the probability level was adjusted according to the modified Bonferroni procedure (cf. Keppel 1991).

Results

Accuracy for interpretation tasks

An overview of the mean accuracy scores and *SDs* for all groups and all tests on the interpretation tasks (written and aural) is provided in Table 1.

Table 1: Descriptive statistics for accuracy scores on the interpretation tasks (written and auditory) for all three groups and tests.

Accuracy (%)	Pre-test		Post-test		Delayed post-test	
	Aural	Written	Aural	Written	Aural	Written
Control	52.46 (24.30)	63.31 (28.37)	66.13 (28.82)	64.94 (33.56)	57.88 (37.15)	71.57 (35.53)
Written SIA	59.54 (22.98)	49.96 (33.28)	78.46 (21.07)	75.63 (29.68)	75.44 (25.29)	78.75 (28.95)
Aural SIA	51.80 (20.04)	45.81 (20.97)	64.13 (22.13)	66.72 (30.63)	64.16 (23.13)	66.99 (29.08)

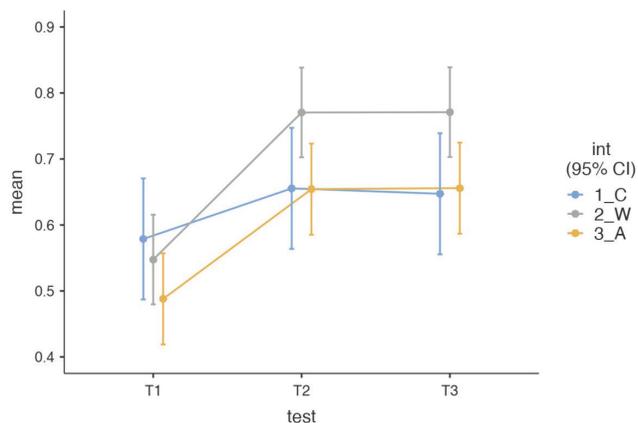


Figure 1: Group differences in accuracy according to test points.

The $3 \times 2 \times 3$ ANOVA for accuracy revealed a main effect only for TEST $F(2, 150) = 18.87, p < 0.001$. No other effects were significant. For an illustration, see Figure 1.

To resolve the main effect of TEST, a post hoc test was conducted. It revealed a significant difference between T1 and T2 ($F(1, 77) = 27.24, p < 0.001, \eta^2_p = 0.231, \text{Cohen's } f = 0.549$), as well as between T1 and T3 ($F(1, 77) = 23.12, p < 0.001, \eta^2_p = 0.252, \text{Cohen's } f = 0.580$), but not between T2 and T3 ($F < 1, \eta^2_p = 0.000, \text{Cohen's } f = 0.004$).

RTs for interpretation tasks

An overview of mean reaction times and standard deviations for all groups and all tests on the interpretation tasks (written and aural) is provided in Table 2.

Table 2: Descriptive statistics for RTs on the interpretation tasks (written and auditory) for all three groups and tests.

RT (seconds)	Pre-test		Post-test		Delayed post-test	
	Aural	Written	Aural	Written	Aural	Written
Control	2.24 (1.44)	4.29 (1.61)	1.84 (1.11)	2.72 (1.02)	1.46 (1.17)	1.54 (0.82)
Written SIA	3.00 (1.91)	6.03 (1.95)	1.34 (1.00)	2.57 (1.83)	1.37 (0.93)	1.68 (1.37)
Aural SIA	1.97 (1.07)	5.52 (1.90)	1.26 (0.83)	2.77 (1.43)	1.36 (1.15)	2.17 (1.54)

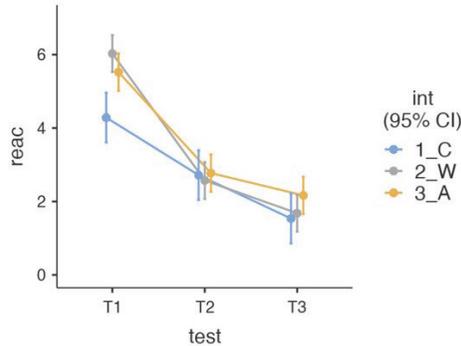


Figure 2: RTs per group according to test point for the written task.

Because the presentation mode (written and auditory) for the two tasks differed, we analysed RTs for each task separately. The 3×3 ANOVA for RTs in the written task revealed a main effect for TEST ($F(2, 150) = 558.82, p < 0.001$). In addition, the results showed a significant interaction of INT*TEST $F(4, 150) = 24.76, p < 0.05$. The resolution of the interaction INT*TEST by TEST revealed only a significant main effect of INT for T1 ($F(2, 75) = 4.81, p < 0.05$), but not for T2 ($F < 1$) and T3 ($F(2, 75) = 1.52, p = 0.224$).

Post hoc tests for T1 revealed significant main effects of INT for the C-group versus the W-group ($F(1, 46) = 9.80, p < 0.01$) and the C-group versus the A-group ($F(1, 45) = 5.07, p < 0.05$), but no effect for the W-group versus the A-group ($F(1, 59) = 1.05, p = 0.31$), with the fastest RTs for the C-group and the slowest reaction times for the W-group and the A-group (see Figure 2).

The 3×3 ANOVA for RTs in the auditory task revealed a main effect for TEST ($F(2, 150) = 21.70, p < 0.001$). In addition, the results showed a significant interaction of INT*TEST $F(4, 150) = 2.73, p < 0.05$. The resolution of the interaction INT*TEST by TEST revealed only a significant main effect of INT for T1 ($F(2, 75) = 3.59, p < 0.05$), but not for T2 ($F(2, 75) = 2.12, p = 0.127$) and T3 ($F < 1$).

Post hoc tests for T1 revealed only a significant main effect of INT for the W-group versus the A-group ($F(1, 59) = 6.63, p < 0.05$), but no effect for the C-group versus the W-group ($F(1, 46) = 2.02, p = 0.162$) and the C-group versus the A-group ($F < 1$), with the fastest RTs for the C-group and the A-group, and the slowest reaction times for the W-group (see Figure 3).

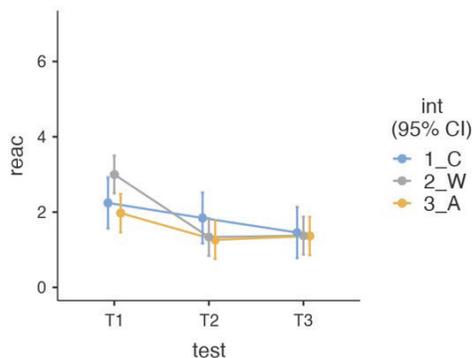


Figure 3: RTs per group according to test point for the auditory task.

Accuracy scores data for the production task

An overview of mean accuracy scores and standard deviations for all groups and all tests on the production task is provided in Table 3.

The 3×3 ANOVA for the production task revealed a main effect for TEST ($F(2, 164) = 6.16, p < 0.01$) and a significant interaction of INT*TEST ($F(4, 164) = 2.96, p < 0.05$).

Resolving the interaction INT*TEST by INT revealed a significant effect of TEST for the W-group ($F(2, 64) = 8.96, p < 0.001$), a marginal significant effect for the C-group ($F(2, 36) = 2.94, p = 0.065$), but no effect for the A-group ($F(2, 64) = 1.35, p = 0.267$). A post hoc test to resolve the main effect of TEST for the C-group revealed no significant difference between T1 and T2 ($F < 1$) and between T1 and T3 ($F(1, 18) = 2.54, p = 0.129, \eta_p^2 = 0.124$, Cohen's $f = 0.375$), but between T2 and T3 ($F(1, 18) = 6.83, p < 0.05, \eta_p^2 = 0.275$, Cohen's $f = 0.616$). In contrast, a post hoc test to resolve the main effect of TEST for the W-group revealed significant differences between T1 and T2 ($F(1, 32) = 20.15, p < 0.001, \eta_p^2 = 0.386$, Cohen's $f = 0.794$) and between T2 and T3 ($F(1, 32) = 5.24, p < 0.05, \eta_p^2 = 0.141$, Cohen's

Table 3: Descriptive statistics for mean accuracy scores on the production task for all three groups and tests.

Written production	Pre-test	Post-test	Delayed post-test
Control	6.79 (2.95)	7.10 (3.38)	5.42 (3.64)
Written SIA	6.45 (2.80)	8.14 (2.98)	7.30 (3.25)
Aural SIA	5.82 (3.61)	6.51 (4.18)	6.54 (3.89)

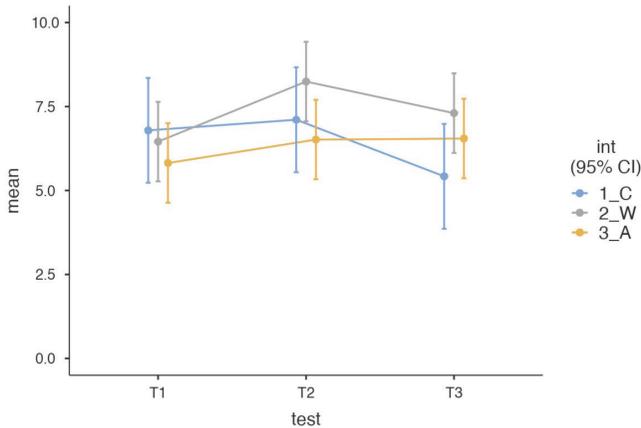


Figure 4: Accuracy scores for production: group vs test points for *-(e)s*.

$f = 0.405$), and a marginal significant effect between T1 and T3 ($F(1, 32) = 3.45$, $p = 0.072$, $\eta^2_p = 0.097$, Cohen's $f = 0.328$).

With respect to H3, we compared the production accuracy on the post-tests between the two treatment groups. The analysis revealed a marginal significant difference between the W-group and the A-group at T2 ($F(1, 64) = 3.73$, $p = 0.057$, $\eta^2_p = 0.055$, Cohen's $f = 0.241$), but not at T1 and T3 ($F_s < 1$); see Figure 4.

Discussion and conclusion

The present study investigated the effects of PI in two different modalities (written and auditory) on the processing of the third person singular *-(e)s* tense form by L2 beginning learners of English measured through accuracy scores (for interpretation and production) and RTs (for interpretation).

We hypothesised that *the A-group would show reliably lower accuracy scores on both interpretation tasks (auditory and written) than the W-group (H1)*. The ANOVA results revealed that all groups improved over time, and the effects were retained up to the second post-test, with no significant difference between the immediate post-test and the delayed post-test. We did not find any PI treatment modality effect. The fact that there were no group differences needs to be carefully interpreted because all three groups showed higher scores after the pre-test (i.e. over time). This finding is not entirely in line with previous studies. For example, Ito and Wong (2019) found an instructional effect, which was stronger for the written than for the auditory PI.

We further hypothesised that *the A-group would display longer RTs on both interpretation tasks (auditory and written) than the W-group (H2)*. This hypothesis could not be confirmed in our study, as we did not find any significant group differences for the two post-tests. We could not confirm that L2 learners' cognitive processing is constrained differently when input is presented in auditory or written modality. The only significant difference that we found referred to group differences at the pre-test, with results for the auditory task showing that the W-group was slower than the A-group and the C-group. For the written task, we found that the W-group and the A-group were slower than the C-group. These differences disappeared for both post-tests, which could be attributed to treatment-improvement effects. However, it could also be simply a practice effect related to speed. However, the reason why the RTs of the C-group were the shortest at the pre-test could also be a result of regular EFL lessons and possible longer exposure to authentic input within these lessons, which could have speeded up their initial performance (i.e. pre-test). To conclude, in contrast to previous findings, which lacked RTs data but demonstrated higher performance under the written presentation modality (Hama and Leow 2010; Kerz, Wiechmann and Riedel 2017; Lund 1991; Park 2004; Williams 2005), we found that assessment modality was not a predictor.

We further hypothesised that *the W-group would show higher accuracy scores on the production task than the A-group (H3)*. This hypothesis was partly confirmed because the W-group showed higher production accuracy at the immediate post-test than the A-group. But this difference disappeared at the delayed post-test. Hence, subjects trained with written PI overcame the non-optimal default processing strategies more successfully, which is in line with previous findings (Hama and Leow 2010; Kerz, Wiechmann and Riedel 2017; Lund 1991; Park 2004).

With regard to retention effects, we hypothesised that *both W-group and A-group would retain the positive effects of PI on interpreting and producing the target feature -(e)s (H4)*. With respect to interpretation ability, our results show that all groups retained the effects with no significant difference between the immediate post-test and the delayed post-test. However, the same result was traced in the case of the C-group, which did not undergo any intervention. Thus, there is no clear indication that the two treatment groups profited from the instructional effects in their interpretation ability. Most probably, in order for PI to commence and lead to correct form-meaning mapping at the comprehension level, learners need to be developmentally ready to acquire such features that require an exchange of information between two different phrases. In contrast, there is a clear indication of intervention improvement effects for production

accuracy for the W-group. Only this group displayed significant improvement from the pre-test to the immediate post-test, and a marginal significant improvement from the pre-test to the delayed post-test, whereas the other two groups did not. Obviously, the written delivered input did not drain the attentional capacities, as was the case with the auditory delivered intervention – a finding, which is in line with Park (2004) and Wong (2001). With regard to retention, the H4 cannot be fully confirmed because the intervention effects for the production ability of the W-group diminished over time. Please note that with respect to production accuracy, there was also a marginal significant decrease of production accuracy from the immediate to the delayed post-test for the C-group. This is difficult to interpret, because this group did not receive any PI treatment.

In our study, we disassembled modality from instructional effects. Future interventional designs should contrast the effects of a single mode delivered PI in relation to PI composed of both auditory and written structured input activities. Another limitation of our study is that we only have the RTs as indirect measurements of cognitive processing. Thus, future studies should consider employing more sophisticated online measures (cf. Roehm and Angelovska 2019) that tap into the PI effects. Moreover, future PI studies should focus on determining the exact amount of interventional exposure needed before the PI effects commence, which most likely depends on the processing constraints of the target feature. It would be reasonable to carry out replication studies targeting more complex morphosyntactic features with an increased intervention time. This could contribute to understanding whether the effect of the intervention would be likewise constrained in relation to the amount of treatment time and complexity of the target feature. In our study, it was only possible to conduct the intervention for a limited time because of given regulated school arrangements and a fixed school schedule. More research is needed to find out whether a more intense interventional exposure would indeed accelerate the acquisition process for beginning EFL learners and help override the imposed processing constraints – a result of developmental non-readiness for late acquired morphosyntactic features that future research endeavours need to embrace when investigating the effectiveness of grammar interventions.

About the author

Tanja Angelovska is an Associate Professor for English Linguistics and Language Teaching at the University of Salzburg. Her research areas include psycholinguistics and L2/ L3 acquisition. She has led and been involved in several projects investigating the acquisition, processing and use of English as a second and third language across various age groups and language combinations. She is associate member of the Centre

for Research and Enterprise in Language (University of Greenwich), the Centre for Applied Research and Innovation in Language Sciences and Education (University of Portsmouth) and the Language Acquisition, Multilingualism, & Cognition (LAM-C) Laboratory (Wilfrid Laurier University), *Brain & Cognition Lab (Texas A&M University, USA)*, and editorial board member of *Journal of Multilingual Theories and Practices*, *Instructed Second Language Acquisition* and *Journal of Monolingual and Bilingual Speech*. She has published in *Language Awareness*, *International Review of Applied Linguistics in Language Teaching* and *Applied Linguistics Review*, and she is co-author of *Second Language Acquisition: A Theoretical Introduction to Real World Applications* (2016, Bloomsbury) and co-edited *L3 Syntactic Transfer* (2017, John Benjamins).

Dietmar Roehm acquired extensive expertise in the areas of signal analysis and modelling of EEG data, eye-tracking, language processing and architecture of languages as a Senior Research Fellow at the MPI for Human Cognitive and Brain Sciences (Leipzig) and as a visiting researcher at the UMass Amherst. His research focuses on language-related brain oscillations, language comprehension and on the role of prediction in (sign) language processing and disorders of language. More recently, he also has a focus on L2/L3 learning and processing (with English/German as a second/third language) across various age groups. Since 2014 he is Full Professor for Psycho-/Neuro- and Clinical Linguistics at the Department of Linguistics, principle investigator in the Centre for Cognitive Neuroscience (University of Salzburg) and head of the research group Neurobiology of Language. In addition, he is an Associate Member of the Language Acquisition, Multilingualism, & Cognition (LAM-C) Laboratory, Wilfrid Laurier University, Waterloo, Canada.

Acknowledgements

We would like to acknowledge the valuable contribution in recruitment and data collection by Alexandra Unterberger and Sabrina Klaffenböck. We are grateful for the feedback and support we received from the *Instructed Second Language Acquisition* editors and reviewers. Any remaining deficiencies are our own.

Funding: This research was supported by a research grant from the Austrian Agency for International Cooperation in Education and Research (OeAD) (Project ‘Processing instruction for L3 English: Differences between balanced and unbalanced bilinguals? (PI-BI-L3)’, grant number MK 01/2018).

Ethics statement: All participants gave written informed consent in accordance with the Declaration of Helsinki and the American Psychological Association’s Ethical Principles of Psychologists and Code of Conduct. The protocol was approved by the Institutional Ethics Review Board of the University of Salzburg, Austria (EK-GZ 07/2018) and the Educational Directorate of Salzburg (510003/0014-PA-ZV-IKT/2019).

References

- Angelovska, T. and Benati, A. (2013) Processing instruction and the age factor: can adults and school-age native speakers of German process English simple past tense correctly? In: A. Benati and J. Lee (eds) *Individual Differences and Processing Instruction* 131–153. London: Equinox.
- Beaman, C. P. (2002) Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology* 55(2): 371–89. <https://doi.org/10.1080/02724980143000307>
- Benati, A. (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* 52: 95–127. <https://doi.org/10.1177/136216880100500202>; <https://doi.org/10.1191/136216801678766877>
- Benati, A. and Angelovska, T. (2015) The effects of processing instruction on the acquisition of English past simple tense: age and cognitive tasks demands. *International Review of Applied Linguistics in Language Teaching* 53(2): 249–69. <https://doi.org/10.1515/iral-2015-0012>
- Benati, A., Lee, J. F. and Houghton, S. D. (2008) From processing instruction on the acquisition of English past tense to secondary transfer-of-training effects on English third person singular present tense. In A. G. Benati and J. F. Lee (eds) *Grammar Acquisition and Processing Instruction: Secondary and Cumulative Effects* 88–120. Clevedon: Multilingual Matters. <https://doi.org/10.21832/9781847691057-006>
- Boersma, P. and Weenink, D. (2014 [1992]) *Praat: Doing Phonetics by Computer*. Retrieved 5 January 2020 from <http://www.fon.hum.uva.nl/praat>
- Cadierno, T. (1995) Formal instruction from a processing perspective: an investigation into the Spanish past tense. *Modern Language Journal* 79(2): 179–93. <https://doi.org/10.1111/j.1540-4781.1995.tb05430.x>
- Cowan, N., Saults, J. S. and Brown, G. D. A. (2004) On the auditory modality superiority effect in serial recall: separating input and output factors. *Journal of Experimental Psychology* 30(3): 639–44. <https://doi.org/10.1037/0278-7393.30.3.639>
- Dennis, I. (1977) Component problems in dichotic listening. *Quarterly Journal of Experimental Psychology* 29(3): 437–50. <https://doi.org/10.1080/14640747708400620>
- Farley, A. P. and McCollam, K. (2004) Learner readiness and L2 production in Spanish: processability theory on trial. *Estudios de Linguística Aplicada* 22(4): 47–69.
- Frankish, C. (1985) Modality-specific grouping effects in short-term memory. *Journal of Memory and Language* 24: 200–9. [https://doi.org/10.1016/0749-596X\(85\)90024-5](https://doi.org/10.1016/0749-596X(85)90024-5)
- Frick, R. W. (1989) Recency and the modality effect in immediate ordered recall. *Canadian Journal of Psychology* 43: 494–511.
- Greenslade, T., Bouden, L. and Sanz, C. (1999) Attending to form and content in processing L2 reading texts. *Spanish Applied Linguistics* 3: 65–90.
- Hahn, A. and Felix, W. S. (1985) Natural processes in classroom second-language learning. *Applied Linguistics* 6(3): 223–38. <https://doi.org/10.1093/applin/6.3.223>
- Hama, M. and Leow, R. P. (2010) Learning without awareness revisited: extending Williams (2005). *Studies in Second Language Acquisition* 32(3): 465–91. <https://doi.org/10.1017/S0272263110000045>
- Ito, K. and Wong, W. (2019) Processing instruction and the effects of input modality and voice familiarity on the acquisition of the French causative construction. *Studies in Second Language Acquisition* 41(2): 443–68. <https://doi.org/10.1017/S0272263118000281>
- Keppel, G. (1991) *Design and Analysis: A Researcher's Handbook* (3rd edn). Englewood

- Cliffs, NJ: Prentice-Hall.
- Kerz, E., Wiechmann, D. and Riedel, F. B. (2017) Implicit learning in the crowd: investigating the role of awareness in the acquisition of L2 knowledge. *Studies in Second Language Acquisition* 39(4): 711–34. <https://doi.org/10.1017/S027226311700002X>
- Lee, J. F. (2014) Processing instruction: where research meets practice. In A. Benati, C. Laval and M. Arche (eds) *The Grammar Dimension in Instructed Second Language Learning* 58–81. London: Bloomsbury Academic.
- Lund, R. J. (1991) A comparison of second language listening and reading comprehension. *Modern Language Journal* 75(2): 196–204. <https://doi.org/10.1111/j.1540-4781.1991.tb05350.x>
- Mavrontoni, M. and Benati, A. (2013) The effects of processing instruction and traditional instruction on two different school-age learners: the case of English present simple tense, third person singular. In J. F. Lee and A. Benati (eds) *Individual Differences and Processing Instruction* 185–210. Sheffield, UK: Equinox.
- Park, G.-P. (2004) Comparison of L2 listening and reading comprehension by university students learning English in Korea. *Foreign Language Annals* 37(3): 448–58. <https://doi.org/10.1111/j.1944-9720.2004.tb02702.x>
- Pearce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E. and Lindeløv, J. K. (2019, February) *PsychoPy3: Experiments in Behavior Made Easy*. Retrieved 5 May 2019 from <https://link.springer.com/article/10.3758/s13428-018-01193-y>
- Penney, C. G. (1980) Order of report in bisensory verbal short-term memory. *Canadian Journal of Experimental Psychology* 34(2): 190–5. <https://doi.org/10.1037/h0081038>
- Pienemann, M. (1989) Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics* 10(1): 52–79. <https://doi.org/10.1093/applin/10.1.52>
- Pienemann, M. (1998) *Language Processing and Second Language Development – Processability Theory*. Amsterdam: John Benjamins. <https://doi.org/10.1075/sibil.15>
- Pienemann, M. (2015) An outline of processability theory and its relationship to other approaches to SLA. *Language Learning* 65(1): 123–51. <https://doi.org/10.1111/lang.12095>
- Reinders, H. and Benson, P. (2017) Research agenda: language learning beyond the classroom. *Language Teaching* 50(4): 561–78. <https://doi.org/10.1017/S0261444817000192>
- Roehm, D. and Angelovska, T. (2019) Neurolinguistic accounts of L2 learning and teaching. In J. I. Liontas (ed.), *TESOL Encyclopedia of English Language Teaching Updates*. Hoboken, NJ: Wiley.
- Treisman, A. M. (1969) Strategies and models of selective attention. *Psychological Review* 76(3): 282–99. <https://doi.org/10.1037/h0027242>
- VanPatten, B. (1990) Attending to form and content in the input: an experiment in consciousness. *Studies in Second Language Acquisition* 12(3): 287–301. <https://doi.org/10.1017/S0272263100009177>
- VanPatten, B. (1996) *Input Processing and Grammar Instruction: Theory and Research*. Norwood, NJ: Ablex.
- VanPatten, B. (2004) Input processing in second language acquisition. In B. VanPatten (ed.) *Processing Instruction: Theory, Research, and Commentary* 5–31. Mahwah, NJ: Erlbaum. <https://doi.org/10.4324/9781410610195>
- VanPatten, B. (2014) Input processing in adult second language acquisition. In B. VanPatten and J. Williams (eds) *Theories in Second Language Acquisition* (2nd edn) 113–34. New York: Routledge. <https://doi.org/10.4324/9780203628942>

- VanPatten, B. (2015) Foundations of processing instruction. *International Review of Applied Linguistics in Language Teaching* 53(2): 91–109. <https://doi.org/10.1515/iral-2015-0005>
- VanPatten, B. and Cadierno, T. (1993) Explicit instruction and input processing. *Studies in Second Language Acquisition* 15(2): 225–43. <https://doi.org/10.1017/S0272263100011979>
- VanPatten, B. and Wong, W. (2004) Processing instruction and the French causative: another replication. In B. VanPatten (ed.) *Processing Instruction: Theory, Research, and Commentary* 97–118. New York: Routledge. <https://doi.org/10.4324/9781410610195>
- Williams, J. N. (2005) Learning without awareness. *Studies in Second Language Acquisition* 27(2): 269–304. <https://doi.org/10.1017/S0272263105050138>
- Wode, H. (1981) *Learning a Second Language: An Integrated View of Language Acquisition*. Tübingen: Gunter Narr.
- Wong, W. (2001) Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition* 23(3): 345–68. <https://doi.org/10.1017/S0272263101003023>
- Wong, W. (2004) The nature of processing instruction. In B. VanPatten (ed.) *Processing Instruction: Theory, Research, and Commentary* 33–65. New York: Routledge.
- Wong, W. and Ito, K. (2018) The effects of processing instruction and traditional instruction on L2 online processing of the causative construction in French: an eye-tracking study. *Studies in Second Language Acquisition* 40(2): 241–68. <https://doi.org/10.1017/S0272263117000274>