CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

The offline and online effects of processing instruction

Nick Henry[ID]

Department of Germanic Studies, The University of Texas at Austin, Austin, TX, USA
Email: nhenry@austin.utexas.edu

Abstract
The present study investigated the effects of processing instruction (PI) on the acquisition of accusative case markers in German, focusing on whether PI influences learners' online processing behaviors. Third and fourth semester L1 English-L2 German learners were divided into two groups: a traditional instruction (TI) group and a PI group. Participants completed offline sentence interpretation and production tasks, as well as a self-paced reading (SPR) task, which provides a detailed investigation of how PI alters processing strategies. Results showed that the PI group outperformed the TI group on the sentence interpretation task and matched their performance on the production task. The SPR task revealed that, in some conditions, the PI group showed increased attention to and processing of accusative case markers after training, while the TI group did not. The results provide some support for the claim that PI is effective because it alters learners' processing strategies.

An important question within the field of second language (L2) acquisition concerns how learners connect morphosyntactic forms (cues) to their meanings and come to process them during online sentence comprehension. Insights from VanPatten's Input Processing model (e.g., VanPatten, 2015b) and its instructional application, processing instruction (PI) (e.g., VanPatten, 2004; Wong, 2004) suggest that the acquisition of morphosyntactic forms and the processing mechanisms needed to use these forms online is hindered by L2 learners' reliance on lexical-semantic or word order cues during comprehension (see also Ellis & Sagarra, 2010). While research on (PI) has shown that it helps learners process morphosyntactic forms more accurately, much previous research has relied primarily on offline measures, which may not tap implicit knowledge and do not provide a

CrossMark

moment-by-moment window into learners' processing behaviors. Recent research (e.g., Benati, 2020a, 2020c; Lee & Doherty, 2019; Wong & Ito, 2018; see Benati, 2020b) has begun to investigate the effects of PI on learners' processing system using online measures such as eye-tracking and self-paced reading (SPR), which is necessary to make strong claims about PI's impact on L2 learners' processing strategies. These studies have, however, used varying approaches and produced mixed results. The present study uses a combination of offline and an online measure (an SPR task) to investigate the effects of PI on the acquisition of accusative case markers in German, focusing on how PI changes learners' tendency to use these forms while processing sentences online.

Background and motivation

Input processing and L2 sentence processing

VanPatten's Input Processing model (VanPatten, 2015b) consists of a set of principles and their corollaries and provides a psycholinguistic account of the mechanisms that L2 learners use to process input. The backbone of the Input Processing model is the *Primacy of Meaning* and the *Availability of Resources* principles (see VanPatten, 2004, 2015a), which explain that learners primarily attend to meaning when processing input, and learners only process input for grammatical form if they have not depleted cognitive resources while processing it for meaning.

Given the primacy of meaning and limits on cognitive resources, the remaining principles of the Input Processing model specify the strategies used to filter input. One of the most important principles—and the focus of the present study—the *First Noun Principle* (FNP), states that "learners tend to interpret the first (pro)noun they encounter in an utterance as the subject" (VanPatten et al., 2013, p. 508; see Jackson, 2007; LoCoco, 1987; VanPatten, 1984).[1] For example, German word order is canonically subject-verb-object (SVO), but the use of case marking on articles or pronouns also allows for object-verb-subject (OVS). When processing OVS sentences, learners often fail to recognize that case-marked articles or pronouns signal OVS word order, and consequently, they often misinterpret OVS sentences like (2).

(1) *Der*   *Hund*   *hört*   *die*   *Katze.* (SVO)
     The$_{NOM}$ dog   hears   the$_{ACC}$ cat.
     "The dog hears the cat."

(2) *Den*   *Hund*   *hört*   *die*   *Katze.* (OVS)
     The$_{ACC}$ dog   hears   the$_{NOM}$ cat.
     "The cat hears the dog."

Additional subprinciples to the FNP state that L2 learners may also rely on lexical-semantic information like animacy, plausibility, and probability (Jackson, 2007), as well as contextual information (VanPatten & Houston, 1998, see VanPatten, 2015a).

It should be noted that VanPatten's Input Processing model is broadly consistent with research in L2 sentence processing and reflects the findings that L2 learners tend to favor the use of lexical-semantic information (e.g., Cameron, 2011;

Sagarra, 2007) and are slow to integrate morphosyntactic information (Hopp, 2006; Jackson, 2008; Keating, 2009), especially when cognitive resources are strained (e.g., Keating, 2009; McDonald, 2006; see Juffs & Harrington, 2011). However, the Input Processing model is specifically concerned with how (lower proficiency) learners *initially* connect linguistic forms to their meaning; thus, it does not describe processing among high proficiency learners, who have been shown to acquire (morpho) syntactic structures, compute full syntactic parses online, and build representations incrementally (e.g., Hopp, 2006, 2010; Jackson, 2008; Jackson & Dussias, 2009; Jackson & Roberts, 2010; Keating, 2009; Pliatsikas & Marinis, 2012; Rossi et al., 2006; Trenkic et al., 2013; Witzel & Witzel, 2012). One important question, however, is how learners transition from an overreliance on lexical-semantic information to processing (morpho)syntactic information in real time, and, relevant to the present study, whether instructional methods like VanPatten's (2004) PI can facilitate this sort of development.

Processing instruction

The instructional application of VanPatten's Input Processing model is known as PI. Using the principles laid out in the model, PI seeks to identify why a target cue is often not connected to its meaning (i.e., its "processing problem")—e.g., due to reliance on a first-noun strategy. In *Structured Input* (SI) activities, the input is manipulated so that learners must use the target cue to understand the input and complete a task (Wong, 2004). For example, in referential SI activities for the target form in the present study, German accusative case markers, participants typically listen to a mixture of SVO and OVS sentences and then select a picture that matches their interpretation of the sentence (see e.g., Henry et al., 2009). They then receive feedback telling them whether they processed the sentence correctly. Crucially, the input used in this picture-selection task does not contain any lexical-semantic cues (e.g., animacy), plausibility information, or context clues that learners can use to understand the sentences, and thus, they must attend to the target form. In this way, PI is a task-essential instructional intervention.

PI was first investigated by VanPatten and Cadierno (1993), who trained learners on clitic object pronouns and OVS word order in L2 Spanish and found that, though PI only trained comprehension of the target form, PI led to both increased comprehension and production of the target form. They therefore concluded that PI led to changes in learners' underlying competence. Since this seminal study, many studies have replicated this basic pattern of results, showing that PI leads to increases in both comprehension and production and is effective when compared with traditional output-based instruction (e.g,. Benati, 2001; Cadierno, 1995; Cheng, 2002; Collentine, 1998; Marsden & Chen, 2011) or a variety of other instructional methods (Benati, 2005; Farley, 2001; Kasprowicz & Marsden, 2017; Keating & Farley, 2008; Marsden, 2006; Qin, 2008; VanPatten et al., 2009). Further research has explored the role of explicit information (EI) in PI, finding that SI is sufficient to cause an increased knowledge of the target form (Benati, 2004; VanPatten & Oikkenon, 1996), but that EI may be beneficial in some circumstances (Fernández, 2008; Henry et al., 2009; VanPatten et al., 2013). See Lee (2015) for a complete review of PI research over the past twenty years.

Interpreting the effects of PI

Within the PI literature detailed above, positive effects for PI or SI activities are typically interpreted following VanPatten and Cadierno (1993), who argued that PI alters the way learners process and extract information from the input, providing intake for the developing system. This research has relied primarily on offline pretest and posttest measures. However, recent research has begun to use online measures to investigate this claim as these methods provide more detailed moment-by-moment information about learners' processing and performance on learning tasks. Additionally, these measures are less likely to be influenced by the use of conscious knowledge or EI (see Keating & Jegerski 2015; VanPatten, 2015a). Thus, these measures allow researchers to make stronger inferences about how PI alters processing strategies or attention to the target form.

Studies on PI and related instructional trainings that employ online processing measures are still relatively few (see Benati, 2020b), but several recent studies have used these methods given the detail they provide. As Lee et al. (2020) discuss, these studies have investigated a diverse range of questions, including the effects of PI (or related trainings) versus other types of instruction or no-training controls (Benati, 2020c, 2020a; Chiuchiù & Benati, 2020; Issa & Morgan-Short, 2019; Lee et al., 2020; McManus & Marsden, 2017, 2018; Wong & Ito, 2018); EI and feedback (Dracos & Henry, 2018; Wong & Ito, 2018), input enhancement (Issa & Morgan-Short, 2019), the modality of training (Ito & Wong, 2019), and the role of individual differences such as working memory (Dracos & Henry, 2021; Issa, 2019; Lee & Doherty, 2020).

Most of these studies have found that PI and similar trainings change online processing behaviors (e.g., Chiuchiù & Benati, 2020; Dracos & Henry, 2018; Issa & Morgan-Short, 2019; Lee & Doherty, 2019; Lee et al., 2020; Wong & Ito, 2018). For example, an eye-tracking study by Wong and Ito (2018) found that PI led learners to reject a first-noun interpretation of causative *faire* constructions in French. Similarly, Benati (2020c) found that, after PI, learners' eye movements showed smaller first noun bias and faster picture selection while they processed English causative passive sentences. Lee and Doherty (2019) found that PI directed learners' attention to grammatical information on Spanish verbs marked for the passive voice and noted that learners' processing became more native-like after instruction. However, other studies have found mixed results or shown that training had no effect on online processing (Dracos & Henry, 2021; Ito & Wong, 2019; McManus & Marsden, 2017). For example, Dracos & Henry (2021) found that a PI-like training did not increase sensitivity to grammatical violations for Spanish subject and tense morphology. McManus and Marsden (2017) found no changes in online processing behavior for the French *Imparfait* when learners received training in the L2 only.

It is also important to note that these studies have used a variety of methodological approaches. For example, many studies employ eye-tracking (e.g., Issa & Morgan-Short, 2019; Ito & Wong, 2019; Wong & Ito, 2018), but some use SPR (Dracos & Henry, 2021; Lee et al., 2020; Malovrh et al., 2020; McManus & Marsden, 2017). Further, some studies measure online processing during the training task itself (e.g., Benati, 2020c; Issa et al., 2015; Wong & Ito, 2018), meaning that reported changes in processing could be related directly to either the task itself or the

provision of feedback during these tasks (or both). Similarly, some studies use online tasks that mirror the task completed during training. For example, Wong and Ito (2018) used a forced choice task in both the PI training and in the eye-tracking task, although the eye-tracking task used a different modality (graphic rather than text-based) and did not provide feedback. Some studies, on the contrary, have tested sensitivity to the target form after training using tasks that are quite different from the training, such as SPR (Dracos & Henry, 2021; McManus & Marsden, 2017, 2018). While the former approaches show how PI affects processing in these specific tasks, the latter approach could provide critical evidence about how the effects of PI generalize to other types of processing tasks.

As this brief overview makes clear, this research lacks a common approach and common findings, making further study necessary. In particular, the literature would benefit from studies that allow researchers to make conclusions about how PI changes sensitivity to the target form, how training generalizes to new task types, and how learners acquire new processing routines over time.

The present study

The goal of the present study is to investigate the effects of PI on the acquisition of accusative case markers in German, with particular focus on whether PI influences learners' online processing performance. This study compares the effects of PI and traditional instruction (TI) using traditional offline tasks (sentence interpretation and picture description) to assess learning gains and online measures (a SPR task) to assess moment-by-moment changes to processing behavior. As noted in the previous section, these methods provide stronger evidence about any such changes than do offline tasks; they also have the potential to reveal processing behavior that is not captured by accuracy measures. This study therefore contributes to the relatively small number studies on PI that use such methods, which to date have shown mixed results. Notably, the present study also diverges from much of the prior research in that it uses an online assessment task, which does not closely resemble the training task. Thus, this study also addresses how PI facilitates changes in processing that are generalizable to different task types. The research questions are as follows:

RQ1a: To what extent does Processing Instruction (PI) lead to more accurate interpretation of accusative case markers in German than Traditional Instruction (TI)?

RQ1b: To what extent does Processing Instruction (PI) lead to more accurate production of accusative case markers in German than Traditional Instruction (TI)?

RQ2: To what extent do PI and TI lead learners to process German accusative case markers more readily when comprehending sentences online, as measured by a self-paced reading task?

Research Questions 1a and 1b focus on the offline interpretation and production assessments. In line with previous research, it is hypothesized that (a) only the PI

group will show increased accuracy on both the interpretation and the production task, demonstrating increased knowledge of the underlying form and (b) the TI group will improve only on the trained skill (i.e., the production task). Research Question 2 is the primary focus of the present study and addresses online processing behavior. It is hypothesized that only the PI group will demonstrate greater sensitivity to case markers during the SPR task.

Methodology

Participants

Participants were drawn from third and fourth semester (low intermediate to intermediate level) German courses at a university in the United States ($N = 159$). Eligible participants met the following requirements: (1) they were native speakers of English, (2) they listed German as their primary and most proficient second language, and (3) they completed all tasks in the experiment. Only those participants who scored below 50% on OVS items in a pretest interpretation task (see below)—and thus did not demonstrate knowledge of the target form—were selected to complete the study. A final total of 51 participants were included in this study (27 male; 24 female). Participants were randomly divided into two treatment groups: The TI group ($n = 26$), and the PI group ($n = 25$).

To ensure comparability between the groups, participants completed a language background questionnaire and a written, multiple-choice proficiency test in German (University of Wisconsin Testing and Evaluation, 2006). Participants also completed a verbal working memory test based on Waters and Caplan (1996). Descriptive statistics for the questionnaire and for these tests are provided in the supplemental materials (Supplement 2). Statistical analyses showed that the two groups were similar with respect to their age, years of German instruction, and time spent in a German speaking country (all $t < 1$). There were no differences between the groups' scores on the proficiency task ($t(49) = 0.288$, $p = .775$) or the working memory task ($t(49) = 1.535$, $p = .131$).[2]

The target form

The target form was the German masculine article *den*, which indicates the object of the sentence, as seen in (1) and (2). As discussed previously, L2 learners use a subject-first strategy to process input and often misinterpret OVS sentences (see LoCoco, 1987 and pretest scores from Culman et al., 2009; Henry et al., 2009; Henry et al., 2017; Kasprowicz & Marsden, 2017). This strategy likely stems not just from the FNP, but also the prevalence of SVO sentences in German (between 80%–95% of transitive sentences with full NPs; Kempe & MacWhinney, 1998; Schlesewsky et al., 2000). Furthermore, overlap between nominative and accusative forms for feminine and neuter genders reinforces the importance of word order as a cue. Indeed, native (L1) speakers also show a subject-first bias. Studies employing SPR tasks demonstrate that L1 speakers incur processing costs in OVS sentences and show higher reading times (RTs) when they encounter disambiguating case information (e.g., Schriefers et al., 1995). This suggests that they process case

information incrementally. However, similar research has shown that less proficient L2 speakers typically wait until the end of the sentence to integrate case markers into the parse (Hopp, 2006; Jackson, 2008).

Materials

TI treatment

All of the materials used in the TI treatment are found in Supplement 1 (Training Materials) and are available in the IRIS database (http://www.iris-database.org; Marsden et al., 2016). TI treatment was adapted from grammar units found in two popular college-level textbooks for beginning German (Lovik et al., 2010; Moeller et al., 2009) and one website popular among German instructors (Thuleen, 1999). These materials are not only used in L2 classrooms themselves; they are similar to other textbooks and online materials used in classrooms and recommended to students. Thus, the training was "traditional" in that it was representative of grammar training used in textbook-based curricula in many current college-level German courses (see the supplemental materials, Supplement 1).

The training consisted of EI, computer training with feedback, and production tasks. As is common for German textbooks, the EI focused on an explanation of direct objects, and it introduced the nominative- and accusative-marked definite and indefinite articles for each gender in German. Consistent with the EI in the source textbooks, there was no information about OVS word order and participants were never explicitly told to avoid a subject-first interpretation strategy. The EI was presented on a computer via E-Prime (Schneider et al., 2012) across several slides such that participants could read each individual slide as long as they wanted but could not review slides once they had moved on.

After receiving the EI, participants completed two computer activities. In the first activity, participants read target sentences and identified the direct object (or indicated that there was no direct object in the sentence). The second activity was a multiple-choice fill-in-the-blank exercise, in which participants supplied nominative and accusative definite articles. In both activities, participants received simple one-word feedback immediately after answering each question. Neither activity contained OVS sentences, as is typical for college-level grammar activities in German.
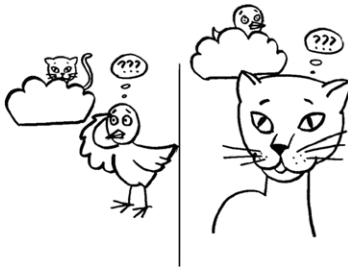
Finally, participants completed two pencil-and-paper activities in which they produced the accusative case. In the first activity, the participants saw a series of items with price tags and wrote four sentences to explain which items they would buy for their dorm rooms without exceeding their budget. In the second activity, participants checked boxes to indicate what items they and a friend own, and they wrote five sentences comparing their possessions.

PI treatment

Materials used in the PI treatment are found in Supplement 1 and are available in the IRIS database. The PI training consisted of EI, a computer-delivered referential SI activity, and two affective SI activities (see the supplemental materials, Supplement 1). The EI was derived from VanPatten et al. (2013) and was designed

A. SVO Sentence   B. OVS Sentence

Der   Vogel sucht   die Katze.   Den   Hund hört die   Katze.
The$_{NOM}$ bird   looks for the$_{ACC}$ cat.   The$_{ACC}$ dog   hears the$_{NOM}$ cat.
"The bird looks for the cat."   "The cat hears the dog"
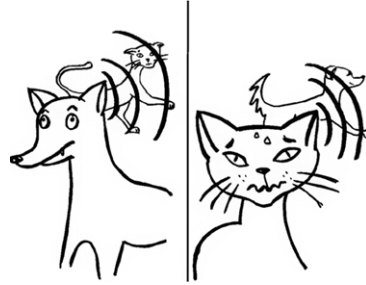

Figure 1. Example items from referential activity (picture-selection task) in PI training.

to emphasize the definite articles' function as a cue to grammatical roles. As the feminine, neuter, and plural definite articles are the same in the nominative and accusative cases, training focused on the masculine nominative and accusative articles *der* and *den*. The EI also highlighted why articles are important for determining grammatical roles and explained that OVS word order is used to emphasize the object of the sentence, often in response to a direct request for information. At the end of the EI, participants received a warning that the first noun is not always the subject of the sentence. The EI was presented via E-Prime over several slides similarly to the TI group.

After the presentation of the EI, participants in the PI group completed the referential activity, which was adapted from VanPatten et al. (2013) and presented via E-Prime (Schneider et al., 2012). In this task, participants heard an SVO or OVS sentence while viewing two pictures: one representing the SVO interpretation of the sentence and the other representing the OVS interpretation (see Figure 1). They used the keyboard to choose the picture they thought best represented the meaning of the target sentence and then received one-word feedback on their answer. The training consisted of 38 OVS sentences and 12 SVO sentences recorded by a female native speaker of German. These items appeared in a repeating sequence of three OVS items, followed by one SVO item to ensure even distribution of the SVO distractor items.[3] Each sentence was a simple transitive construction consisting of a NP-V-NP sequence in which one NP was masculine and therefore unambiguously identified its case and grammatical role. The sentences contained only animate nouns, and both SVO and OVS interpretations were equally plausible. Further, sentences were presented without context or intonational cues.[4] Consequently, the case markings were the only cue that participants could use to comprehend the sentences.

Finally, participants completed two affective activities adapted from Farley (2004) and used by Henry et al. (2017). In these activities, participants read a series of OVS sentences relating to interpersonal relationships and marked their opinions

about these statements. These activities did not have correct answers, and thus acted primarily as an input flood, exposing the participants to more OVS sentences in a different communicative context (see also Marsden & Chen, 2011 for a full discussion of affective activities and their function within PI training).

It should be emphasized that the TI and PI trainings were identical in that they both used EI, computer-based activities, and open-ended activities, and the quality and quantity of corrective feedback was identical.[5] Thus, the differences between the TI and PI trainings lie solely in the principles of PI (Lee & VanPatten, 2003): the PI training concentrated on one form (the masculine accusative), it used both aural and written input, and attention to the meaning of the target form was required for successful completion of the tasks. Most importantly, the PI activities were designed to compel the participants to process the target form instead of relying on a first-noun strategy. The TI training, on the other hand, presented multiple forms at once, focused on accurate identification and production of the target forms, and included reliable animacy, plausibility, context, and word order cues (by only including SVO sentences) that mediated the interpretation of the input.

Assessment measures

The full set of assessment measures used in this study are found in Supplement 3 (Assessment Materials) and available in the IRIS database. The primary means of assessing the offline effects of training was a written pretest/posttest consisting of a sentence interpretation task and a production task. There were two versions of the test, which were counterbalanced among participants.

The sentence interpretation task consisted of four SVO and four OVS sentences mixed with 12 distractor items.[6] Each target item was followed by a Yes/No comprehension question targeting successful interpretation of grammatical roles, as seen in (3). Questions were presented in English so that participants could not simply match the articles between the sentence and the comprehension question.

(3)   *Die Oma   überrascht   der   Opa   während der Party.*
       The$_{ACC}$ grandma surprises   the$_{NOM}$ grandpa during the party.
       "The grandpa surprises the grandma during the party."
       Is the grandpa surprising the grandma?   Yes   No

The production task consisted of four picture-series—two target series and two distractor series. As in Figure 2, each series contained three pictures, a question prompt, and vocabulary relating to the pictures. Participants were instructed to use the pictures to respond to the prompt and write at least one sentence per picture to tell a story. The target items depicted a person interacting with a masculine person or object in each of the three pictures. Thus, an appropriate response to the picture specifically elicited the masculine articles *der* and *den*. Participants were not limited to three sentences, nor were they required to use the verbs presented to them.

The online effects of training were investigated through a SPR task administered before and after training. This task tested the sensitivity to case marking through a comparison of RTs on SVO and OVS sentences. Given that OVS sentences are far

What is the boy doing with the double bass (der Bass)?
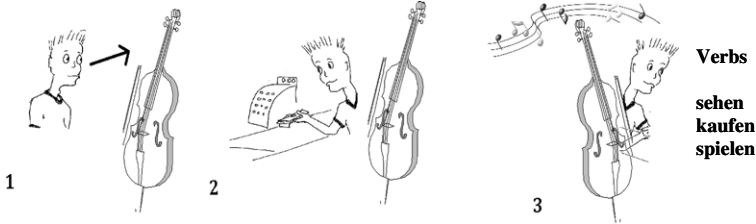

Verbs

sehen
kaufen
spielen

Figure 2. Example item from production task.

less frequent in German and require more complex structure building than SVO sentences, effortful processing and longer RTs are expected on the disambiguating segments (i.e., masculine nouns) of OVS sentences. That is, if the accusative *den* is read in NP1, or if the nominative *der* is read in NP2, surprisal effects should lead to increased RTs. Thus, if participants were to show increased RTs on OVS relative to SVO sentences, as has been seen in research with L1 speakers (Schriefers et al., 1995), this would provide evidence that L2 German learners are sensitive to case-marking information and use it to assign grammatical roles.

The SPR task was presented in a phrase-by-phrase noncumulative moving window design (Just et al., 1982) using E-Prime (Schneider et al., 2012). In this task, participants were first presented with a fixation cross to mark the beginning of the sentence. When they pressed the spacebar on the keyboard, they saw the first phrase in the sentence followed by dashes representing the position of the words in the remainder of the sentence. Using the spacebar, they advanced through the phrases until they reached the end of the sentence. Only one phrase was visible to them at a time. Following each sentence, participants answered a comprehension question in English to ensure that they read the sentence for meaning.

In total, the SPR task consisted of 72 items and comprehension questions, including 24 experimental sentences and 48 fillers. The 24 experimental sentences consisted of an NP-V-NP sequence followed by one or two prepositional phrases. Each sentence included one masculine NP and one NP that was either feminine or neuter. There were four versions of each sentence, created by crossing two variables: word order (SVO or OVS) and placement of the masculine nouns (i.e., the disambiguating noun). The quadruplets were divided into five to seven regions for presentation in the SPR task.

For each sentence, the masculine noun is the point of disambiguation, and thus was used as a critical region for the analyses. The two regions that followed the critical regions were spillover regions. In (4), bold-faced type indicates critical regions, italicized type represents the spillover regions, and slashes indicate the division of the sentences.

(4a) SVO, Masculine-First
     Der Opa   /*überrascht* /die Oma   /*während* /der Party.
     The$_{NOM}$ grandpa /surprises   /the$_{ACC}$ grandma /during   /the party.
     "The grandpa surprises the grandma during the party."

(4b)   OVS, Masculine-First
    Den Opa   /*überrascht* /die Oma   /während /der Party.
    The$_{ACC}$ grandpa /surprises   /the$_{NOM}$ grandma /during   /the party.
    "The grandma surprises the grandpa during the party."

(4c)   SVO, Masculine-Second
    Die Oma   /*überrascht* **/den Opa**   /*während* /der Party.
    The$_{NOM}$ grandma /surprises   /the$_{ACC}$ grandpa /during   /the party.
    "The grandma surprises the grandpa during the party."

(4d)   OVS, Masculine-Second
    Die Oma   /*überrascht* **/der Opa**   /*während* /der Party.
    The$_{ACC}$ grandma /surprises   /the$_{NOM}$ grandpa /during   /the party.
    "The grandpa surprises the grandma during the party."

Half of the comprehension questions specifically tested the assignment of grammatical roles like the questions in the offline sentence interpretation task; the other half of the comprehension questions inquired about information found in the prepositional phrase. While it is less preferable to have comprehension questions that draw attention to the target form (Keating & Jegerski, 2015; Leeser et al., 2011), it was important to have an independent measure of sentence interpretation during this online task in order to draw some comparison to the offline task. Only 12 of the 72 sentences in the SPR focused on grammatical roles, and participants had to pay attention to all parts of the sentence in order to answer all comprehension questions.

The experimental stimuli were divided into four experimental lists. These lists were controlled so that there were no differences in word length in the regions of interest, and so that the SVO and OVS interpretations of the sentences were equally plausible. Rather than controlling for the raw frequency of the words used in the SPRT, the stimuli were controlled for familiarity by selecting the words from the textbook used in the participants' German classes. However, a pilot experiment showed that many participants were not as familiar with the words as expected; therefore, participants were trained on the words before both the pretest and the posttest. The lists were distributed so that participants only saw one version of each sentence, and, thus, participants saw six sentences from each of the conditions described above.

Procedure

The experiment was conducted in three sessions and followed the same procedure regardless of the participants' group. Session one was conducted in the participants' classrooms and included the language background questionnaire, proficiency test, and the pretest (i.e., the offline sentence interpretation and production tasks). Sessions two and three were held in a quiet laboratory, and participants were tested individually. Session two was completed between 1 and 3 weeks after Session one. Session two began with the vocabulary training designed to teach participants the primary nouns and verbs used in the SPR task. In the vocabulary training, participants saw and listened to each word three times while viewing their English

translation. Participants were then tested on their knowledge of the words, and they repeated words they could not identify correctly (see the supplemental materials, Supplement 1). After the vocabulary training, participants completed a working memory task to ensure comparability between participants[7] and the pretest SPR task. Session three was completed between 4 and 6 weeks after Session two. In Session three, participants first completed a reduced version of the vocabulary training, in which they saw and listened to the words twice. This was done to ensure that participants retained the vocabulary and had seen the words in the SPR task recently, similar to the pretest. They then completed the PI or TI training, the posttest SPR task, the posttest, and a vocabulary test to ensure that they had retained the vocabulary throughout the experiment.[8]

Data scoring

Sentence interpretation
For the sentence interpretation task, target SVO and OVS items were scored separately; one point was given for correct Yes/No responses, and zero points were given for incorrect responses. There was a maximum score of four points, one for each target sentence.

Written production
For the production task, separate scores were computed for production of the nominative and accusative cases. The score was computed as a ratio of the number of accurate uses to the number of obligatory occasions in each response. Thus, the maximum score was one (100%). For the purpose of this measure, both the nominative or accusative markers *der* or *den,* as well as their corresponding pronouns *er* and *ihn*, were considered to be correct if they accurately described the action depicted in the picture. An obligatory occasion was defined as a point in the sentence in which the masculine articles were necessary to complete the phrase grammatically. Participants who did not produce any obligatory occasions (e.g., if they named the characters or used alternate vocabulary to describe the pictures) could not be given a score for this task and were treated as missing data in the analyses.

SPR comprehension
In the SPR task, responses to the comprehension questions for the experimental sentences were recorded by E-Prime, along with RTs for each region. For comprehension questions that tested grammatical role assignment, scores were computed separately for SVO and OVS items and reflect a percentage of correct answers.

SPR reading times
RT data were trimmed by first removing all RTs below 200 ms and above 4,000 ms from the data; in addition, any RT outside a range of +/− 3 standard deviations from a participant's mean RT for the experimental sentences was defined as an outlier and excluded from analyses. These trimming procedures resulted in a loss of 3.5% of the data. In addition, data were only analyzed for sentences that were correctly understood, resulting in a loss of an additional 27.6% of the data. Thus, to ensure that participants

completed the tasks correctly and contributed enough data to the sample, participants were removed from the final data set if they had an overall comprehension rate less than 60% for the entire task or less than 33% for one of the experimental conditions. Four participants from the PI group were excluded from the analyses, meaning that the PI group consisted of 21 participants for the SPR task.

Data analysis

Analyses were performed using mixed-effects models, which allow flexibility for different data distributions (e.g., binomial variables) and are robust against homoscedasticity and sphericity (McManus & Marsden, 2019). Models were fit using the *mixed()* function in the R package *afex*[9] (Singmann et al., 2021). For all analyses, the maximal model structure was attempted (Barr et al., 2013). When the maximal model did not converge, it was systematically reduced using the procedure laid out by Singmann (2021). The structure of the final model is noted in the results for each analysis. To explore significant main effects and interactions, post hoc contrasts were performed using the package *emmeans* (Lenth, 2020) and a Bonferroni adjustment for multiple comparisons.

Statistical analyses for the interpretation task and accuracy on the SPR comprehension questions focus on OVS sentences as these are the foci of the research questions in the present study. Analyses for the production task focus on the production of the accusative case.[10] These analyses were all carried out using Generalized Linear Mixed Models (GLMM) with a logit link binomial error distribution and using likelihood ratio tests. Likelihood ratio tests were used to compute p values for fixed effects. The maximal model included fixed effects of Time (Pretest vs. Posttest) and group (PI vs. TI), and the Time × Group interaction, with by-participant random slopes and intercepts for Time plus the correlation between slopes and intercepts.

Analysis of RTs for the SPR task was conducted using linear mixed effects models of log-transformed RTs, which compensated for the non-normality of these data. Separate analyses were carried out for each segment. Masculine-First and Masculine-Second sentences were analyzed separately, as the critical regions differed between them. The maximal model included the fixed effects of Time (Pretest vs. Posttest), Group (PI vs. TI), and Word Order (SVO vs. OVS) and the interactions between them. The random effects included by-subject and by-item random intercepts and random slopes for Time, Word Order, and the Time x Word Order interaction plus correlations among slopes and intercepts.

Results

Sentence interpretation

The descriptive statistics for the interpretation task are presented in Table 1. The maximal GLMM model did not converge. Results of the final[11] model yielded an effect for Group ($\chi^2(1) = 17.48$, p < .001), Time ($\chi^2(1) = 55.18$, p = < .001), and the Group x Time interaction ($\chi^2(1) = 46.72$, p = < .001). Follow-up comparisons for the Group x Time interaction confirmed that there were no differences between the groups at pretest

Table 1. Descriptive statistics for sentence interpretation task (maximum score of four)

	SVO Sentences				OVS Sentences			
	M(SD)	95% CI	Mdn	IQR	M(SD)	95% CI	Mdn	IQR
TI								
Pretest	3.50 (0.71)	3.21, 3.79	4.00	1.00	0.73 (0.72)	0.44, 1.02	1.00	1.00
Posttest	3.81 (0.49)	3.61, 4.01	4.00	0.00	0.81 (1.44)	0.22, 1.39	0.00	2.00
PI								
Pretest	3.68 (0.48)	3.21, 3.87	4.00	1.00	0.64 (0.63)	0.42, 0.93	1.00	1.00
Posttest	3.68 (0.69)	3.40, 3.96	4.00	0.50	3.40 (0.91)	3.02, 3.78	4.00	1.00

Note: Values for the 95% confidence interval are bootstrapped using one thousand samples

Table 2. Descriptive statistics for written production task (ratio of correct to obligatory occasions)

	Nominative Forms				Accusative Forms			
	M(SD)	95% CI	Mdn	IQR	M(SD)	95% CI	Mdn	IQR
TI								
Pretest	0.95 (0.12)	0.91, 1.00	1	0	0.56 (0.4)	0.40, 0.72	0.55	0.83
Posttest	1.00 (0.00)	1.00, 1.00	1	0	0.75 (0.38)	0.60, 0.91	1.00	0.50
PI								
Pretest	0.95 (0.12)	0.90, 1.00	1	0	0.46 (0.4)	0.30, 0.63	0.50	0.88
Posttest	0.99 (0.05)	0.97, 1.01	1	0	0.85 (0.32)	0.72, 0.98	1.00	0.00

($OR = 0.891$, $SE = 0.464$, z-ratio $= -0.221$, $p = .825$), but the PI group outperformed the TI group at posttest ($OR = 45.834$, $SE = 25.815$, z-ratio $= 6.791$; $p < .001$). Further comparisons showed that the PI group improved from pretest to posttest ($OR = 0.0167$, $SE = 0.008$, z-ratio $= -8.318$, $p < .001$), but the TI group did not ($OR = 0.860$, $SE = 0.334$, z-ratio $= -0.387$, $p = .699$).

Written production

The descriptive statistics for the production task are shown in Table 2. The maximal GLMM model did not converge. Results of this final model[12] yielded no effect for Group ($\chi^2(1) = 0.15$, $p = 0.695$), but there was a significant effect for Time ($\chi^2(1) = 94.22$, $p = < .001$), which was qualified by a Group x Time interaction ($\chi^2(1) = 13.93$, p $= < .001$). Follow-up comparisons for the Group x Time interaction confirmed that both groups improved from pretest to posttest (PI: $OR = 0.017$, $SE = 0.011$, z-ratio $= -6.350$, $p < .001$; TI: $OR = 0.199$, $SE = 0.074$, z-ratio $= -4.319$, $p < .001$), but there were no differences between the groups at pretest ($OR = 0.417$, $SE = 0.385$, z-ratio $= -0.946$, $p = .343$) or posttest ($OR = 4.836$, $SE = 4.863$,

Table 3. Descriptive statistics for SPR comprehension questions (proportion of correct answers)

Time	Pretest				Posttest			
	M(SD)	*95% CI*	*Mdn*	*IQR*	*M(SD)*	*95% CI*	*Mdn*	*IQR*
TI								
Total	0.72 (0.05)	0.70, 0.75	0.73	0.07	0.74 (0.07)	0.71, 0.76	0.75	0.10
SVO	0.77 (0.21)	0.69, 0.85	0.83	0.33	0.79 (0.15)	0.73, 0.85	0.83	0.17
OVS	0.25 (0.23)	0.16, 0.34	0.17	0.38	0.29 (0.19)	0.21, 0.36	0.25	0.33
PI								
Total	0.73 (0.05)	0.71, 0.75	0.73	0.07	0.76 (0.05)	0.74, 0.78	0.77	0.07
SVO	0.78 (0.21)	0.68, 0.87	0.83	0.42	0.8 (0.15)	0.74, 0.87	0.83	0.25
OVS	0.26 (0.20)	0.17, 0.35	0.17	0.25	0.44 (0.24)	0.33, 0.55	0.50	0.33

Note: Values for the 95% confidence interval are bootstrapped using one thousand samples; values for SVO and OVS sentences reflect only those items in which grammatical role assignment was tested in the comprehension question.

Table 4. Proportion of participants scoring at or above chance on OVS sentences on the SPR task

	Pretest	Posttest
Traditional Instruction	6/26 (23.1%)	8/26 (30.7%)
Processing Instruction	6/25 (24%)	15/25 (60%)

z-ratio $= 1.567$, $p = .117$). The interaction thus appears to be driven by the comparatively larger gains made by the PI group relative to the TI group.

SPR comprehension questions

The descriptive statistics for SPR comprehension accuracy are given in Table 3. Results of the maximal model[13] yielded a significant effect for Time ($\chi^2(1) = 6.63$, $p = .01$), but no effect for Group ($\chi^2(1) = 2.53$, $p = .112$) or the Group x Time interaction ($\chi^2(1) = 1.77$, $p = .184$). Follow-up pairwise comparisons for Time confirmed that the effect stemmed from higher accuracy scores on the posttest ($OR = 0.556$, $SE = 0.127$, z-ratio $= -2.570$, $p = .010$).

Given that these results differed from those of the sentence interpretation task, a post hoc analysis was conducted to disentangle changes in individual participants from overall group means and evaluate the degree to which they had maintained or abandoned a strong first noun strategy. For this analysis, the proportion of participants who scored at or above chance on the OVS sentences was calculated for each group at each test time. As can be seen in Table 4, about 24% of the participants in both groups scored at or above chance on the pretest. On the posttest, 60% of the participants in the PI group scored at or above chance, while only 30.7% of the TI group did.

Table 5. Mean reading times (SDs) by group and condition for SPR task, Masculine-First items

Segment	NP1	V	NP2	Preposition	Final
Traditional Instruction					
SVO, Pretest	909 (295)	807 (358)	909 (357)	540 (153)	935 (383)
OVS, Pretest	1014 (390)	982 (487)	915 (388)	579 (167)	962 (336)
SVO, Posttest	831 (308)	800 (372)	881 (356)	519 (121)	786 (241)
OVS, Posttest	779 (285)	784 (378)	828 (259)	500 (101)	862 (444)
Processing Instruction					
SVO, Pretest	1062 (535)	898 (436)	957 (371)	554 (181)	918 (377)
OVS, Pretest	1172 (632)	930 (508)	1022 (473)	540 (210)	888 (385)
SVO, Posttest	1126 (682)	825 (402)	1055 (509)	517 (156)	861 (387)
OVS, Posttest	1253 (658)	872 (399)	1234 (662)	500 (110)	872 (343)

Table 6. Model results for Segment 1, Masculine-First sentences

Effect	Df	F	p
Time	1, 38.05	3.89	.056+
Word Order	1, 13.35	5.72	.032*
Group	1, 44.89	3.38	.072+
Time × Word Order	1, 42.87	0.29	.595
Time × Group	1, 38.83	8.11	.007**
Word Order × Group	1, 45.03	2.44	.125
Time × Word Order × Group	1, 44.70	4.41	.042*

SPR reading times

The analyses presented here focus on the disambiguating segments (Segments 1 and 3 for Masculine-First and Masculine-Second sentences, respectively). Analyses for spillover regions are presented in the supplemental materials (Supplement 2) alongside a post hoc analysis of overall RTs.

RTs for Masculine-First sentences

The mean RTs and standard deviations for Masculine-First sentences at each segment are displayed by group and condition in Table 5.

For the critical segment, Segment 1 (the first NP), the maximal model did not converge. Results of the final model[14] (Table 6, Figure 3) yielded a significant main effect for Word Order and a marginally significant main effect for Group, which were qualified by significant Time x Group and Time x Word Order x Group interactions. Pairwise comparisons for the SVO-OVS contrast (Table 7) indicated that

Table 7. Results of the pairwise comparisons for the SVO-OVS contrast in Masculine-First items

Group	Time	Estimate	SE	df	Lower CL	Upper CL	t	p
TI	Pre	−0.04	0.02	75.07	−0.11	0.02	−1.83	0.28
TI	Post	0.02	0.02	69.60	−0.04	0.08	0.76	1.00
PI	Pre	−0.03	0.03	80.21	−0.10	0.03	−1.26	0.85
PI	Post	−0.07	0.03	68.10	−0.14	−0.01	−2.78	0.03*

Figure 3. Model results for Segment 1, Masculine-First sentences.

the PI and TI groups had similar RTs on SVO and OVS sentences on the pretest. In contrast, the PI group—but not the TI group—had significantly higher RTs on OVS sentences in the posttest.

RTs for Masculine-Second sentences
The mean RTs and standard deviations for Masculine-Second sentences at each segment are displayed by group and condition in Table 8.

For the critical segment, Segment 3 (the second NP), the maximal model did not converge. Results of the final model[15] (Table 9, Figure 4), yielded no significant effects, although a marginally significant effect for Group arose because the PI group had higher RTs overall.

Table 8. Mean reading times (SDs) by group and condition for SPR task, Masculine-Second items

Segment	NP1	V	NP2	Preposition	Final
Traditional Instruction					
SVO, Pretest	1054 (420)	917 (321)	800 (247)	536 (104)	885 (324)
OVS, Pretest	1000 (320)	936 (449)	739 (205)	464 (105)	1002 (427)
SVO, Posttest	835 (265)	802 (413)	713 (218)	498 (103)	831 (307)
OVS, Posttest	810 (348)	774 (323)	698 (299)	478 (77)	877 (354)
Processing Instruction					
SVO, Pretest	1153 (484)	988 (552)	959 (412)	541 (171)	895 (333)
OVS, Pretest	1220 (588)	977 (592)	922 (502)	522 (166)	932 (404)
SVO, Posttest	1117 (584)	813 (403)	1028 (484)	548 (199)	790 (297)
OVS, Posttest	1103 (534)	778 (421)	968 (670)	511 (189)	901 (446)

Table 9. Model results for Segment 3, Masculine-Second sentences

Effect	Df	F	p
Time	1, 33.00	2.25	.143
Word Order	1, 18.45	2.10	.164
Group	1, 45.01	3.45	.070+
Time × Word Order	1, 19.82	0.37	.547
Time × Group	1, 44.90	2.16	.149
Word Order × Group	1, 65.90	0.01	.922
Time × Word Order × Group	1, 43.28	0.41	.527

Discussion

The goal of the present study was to investigate the effects of PI on the acquisition and processing of the accusative case in German, specifically testing whether PI or TI leads to significant changes in performance accuracy in comprehension and production (RQ1a and 1b, respectively) and in learners' processing strategies (RQ2). The results show that the PI group improved their comprehension of OVS sentences in both the offline comprehension task and the SPR task. They also improved accuracy with accusative case markers on the written production task. The TI group, on the other hand, only improved on the written production tasks. Thus, as hypothesized, only the PI group demonstrated increased knowledge of the underlying form. On the SPR task, only the PI group had increased RTs for OVS sentences, though this effect was only observed when the masculine noun (i.e., the disambiguating noun) was read first. Thus, hypothesis for RQ2 was only partially confirmed. Together, these results suggest that only the PI training pushed participants to
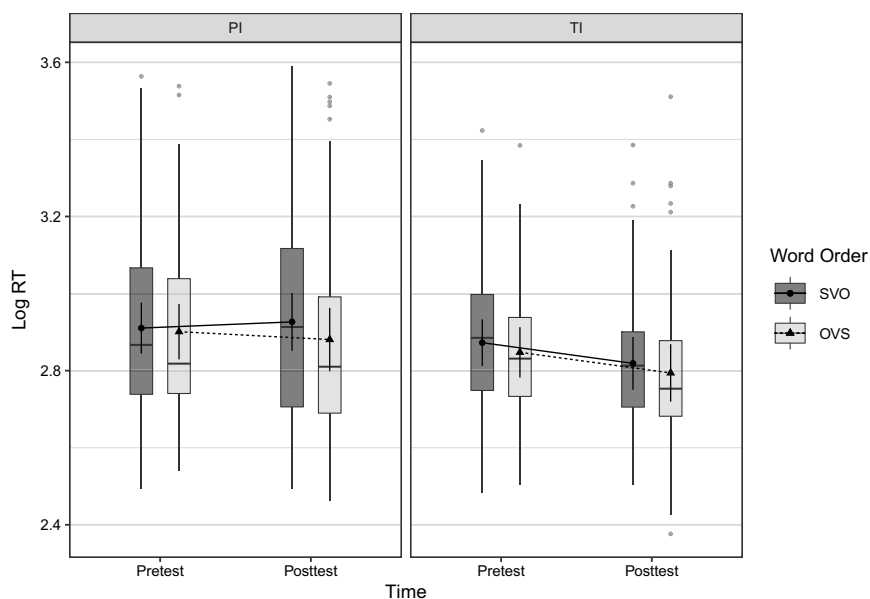
Figure 4. Model results for Segment 3, Masculine-Second sentences.

abandon a strict first-noun strategy and process case-marking information more readily. These results are now discussed with reference to the role of PI in the development of form-meaning connections and its impact on the ability to process case cues online.

The development of form-meaning connections

Like many previous studies, the present study replicated the findings of VanPatten and Cadierno (1993), showing that the PI group improved on both comprehension and production measures, while the TI group improved only on production. In line with previous PI research, results from the present study seem to indicate that the PI training led these learners to realize the relationship between case-marking articles and grammatical roles and to create or strengthen form-meaning connections in the developing system, while the TI training did not.

The TI group's improvement in the production task can be explained by the development of explicit knowledge about the target forms, especially because the production-focused practice did not require the learners to connect articles to their grammatical roles at all. Rather, the TI group simply needed to learn to produce *den* post-verbally in both the training and in the production assessments. Indeed, the TI training did not provide the learners with any evidence that word order is not a reliable cue to grammatical role assignment in German. On the other hand, the learners could not have completed PI training successfully without realizing the connection between the articles and grammatical roles. Presumably, this would have helped them develop the ability to process word order and case-marking

morphology as separate (i.e., non-converging) cues to grammatical role and gain more detailed and more robust knowledge in the underlying system. This could explain why the PI group outperformed the TI group on the posttest comprehension measure and also improved in the production measure, despite not practicing production at any point in the training.

The use of case markings online

Together, the analyses on the SPR task showed that, at pretest, neither group used case markers to understand the comprehension questions or processed case-marking information online. On the posttest, the TI group did not significantly improve their comprehension of OVS sentences in the comprehension questions, nor did they reduce their reliance on a first-noun strategy. The analysis of RTs indicated that there were no differences in how they read SVO and OVS sentences, and thus they did not process case markers online. This is expected because the comprehension tasks indicated that they lacked the appropriate form-meaning connections needed to disambiguate between the word orders.

On the other hand, the PI group improved on the SPR comprehension questions, and fewer participants in the PI group used a strict first-noun strategy to interpret sentences on the posttest. The RTs showed that the PI group did process case marking information, but that they did so on a limited basis: participants had elevated RTs on OVS sentences in Masculine-First sentences, but not in Masculine-Second sentences. Thus, while these results clearly suggest that PI can promote incremental processing, its effects may be limited by aspects of the training (e.g., the length), salience of the target form in sentence context, or perhaps cognitive factors (see below). In addition to these primary results from the SPR RTs, it is also noteworthy that the post hoc analysis of RTs (see Supplement 2) showed that the PI group had higher RTs than the TI group across all conditions, but only in Segments 1 and 3 (the two NPs) on the posttest. This is significant given that the PI training aimed to teach learners to focus on articles as relevant cues. As there were no RT differences at other regions, it seems likely that this difference stems directly from the PI training itself and not from task effects or other external factors. Notably, this effect would be predicted if PI influences how learners direct processing resources to articles, whether they are masculine or not. That is, this could indicate that participants deliberately checked the first and second NPs for case-marking information.

The results of the SPR task, then, indicate that PI affected the processing system in two ways: (1) it helped participants identify and process the relevant cues in the input and (2) participants were better able to integrate case-marking information into their representation of the input. Thus, these results replicate effects observed by previous studies on PI (Benati, 2020c; Chiuchiù & Benati, 2020; Issa & Morgan-Short, 2019; Wong & Ito, 2018), showing that PI helps learners develop form-meaning connections and identify, attend to, and process cues in the input. However, given the null effects in the Masculine-Second sentences, it seems that PI may have a limited impact on the ability to integrate information rapidly and recover from misparses in some types of sentences. This suggests that processing routines emerge in two stages, as learners first learn to allocate attention to useful cues and are later able to integrate this information into the parse more fully.

It is unclear why the learners showed no sensitivity to word order in Masculine-Second sentences. One possibility is that this study was simply unable to capture these effects due to low statistical power, especially given a relatively small sample size and the elimination of a portion of the data. Further, it is possible that the participants treated ambiguous (non-masculine) NPs in the first segment as nominative and either (a) did not attend to case marking in the second NP or (b) were unable to fully recover from their initial SVO interpretation when they later encountered the nominative *der*. As hinted above, the ability to attend to morphological cues and recover from mis-parses may need time to develop, as these skills are influenced by the interaction of various internal and external factors such as working memory capacity (e.g., Dracos & Henry, 2021; Juffs & Harrington, 2011; Keating, 2009), the speed of lexical access (e.g., Miller, 2011, 2014), and task type (e.g., Leeser et al., 2011). It is also possible that learners would be better able to integrate case information online if it were supported, for example, by implicit prosody (Dekydtspotter et al., 2008; Harley et al., 1995; Hwang, 2007; Ying, 1996). Future research will need to employ larger sample sizes and employ methods that can detect smaller effects; such research could shed light on how these factors—and other aspects of PI trainings—mediate the availability of native-like processing behaviors after PI. It may be particularly interesting to explore whether repeated PI interventions utilizing increasingly complex discourse can increase the efficacy of training by promoting the ability to use the forms in a wider variety of sentence types and in more complex tasks. Similarly, future research could investigate the effects of massed and distributed practice for both offline performance and online processing (see Bird, 2010; Miles, 2014).

In sum, the present study supports a narrow interpretation of the claim that PI changes processing behaviors (see e.g., VanPatten & Cadierno, 1993) because PI compelled learners to direct attentional resources to and process the relevant morphosyntactic cues, but it did not lead to a target-like incremental processing pattern in all conditions. Despite this limitation, the changes that were observed are consistent with the goals of the PI training, which aimed to teach participants to attend to articles instead of simply relying on a first-noun strategy. It is also important to recognize that PI influenced these learners' processing strategies in a task (the SPR task) that is very different from the SI activity. To the extent that this task is generalizable to a non-laboratory setting, PI would therefore be useful, not only because it aids in the development of form-meaning connections but also because changes in processing behavior should allow learners to extract meaningful information from input beyond the instructional context, creating better intake for the developing system.

Conclusions

The present study is an important contribution to PI research in that it provides evidence that PI affects learners' online processing strategies and generalizes to different task types. It also provides evidence that learners develop native-like processing routines by first learning to attend to relevant cues and later learning to integrate this information rapidly and efficiently. More generally, this study highlights the

benefits of developing psycholinguistically motivated instructional methods that are then evaluated using psycholinguistic (i.e., processing) tasks. Studies that use such an approach can not only lead to improvements in instructional materials but also add precision to the theoretical models and constructs used to develop those materials. Consequently, a combination of classroom and psycholinguistic approaches has much to offer both fields and can facilitate an informative dialogue between the disciplines.

Notes

1 1. Although the *First Noun Principle* is stated as a universal, there is conflicting evidence as to whether this is indeed universal or if learners transfer word order strategies from their L1 (Isabelli, 2008; Seibert-Hanson & Carlson, 2014). See VanPatten (2015a) for an in-depth discussion of this issue.

2 See Table S1 in the supplemental materials for full background information.

3 This study was part of a larger study on PI, which compared multiple PI trainings and used the commonly cited *trials to criterion* method to analyze differences between the groups (see Fernández, 2008). The distribution and sequencing of items was necessary to meet the goals of that study and followed other PI studies that use the trials to criterion method and utilize PI to emphasize correct processing of the target items (here, OVS sentences) rather than distractors (SVO sentences) (e.g., VanPatten et al., 2013). None of the participants reported that they were aware of this sequence in a debriefing questionnaire.

4 The sentences were the same as the *Monotone Prosody* sentences used in Henry et al. (2017). Please refer to that paper for information on how intonation was controlled.

5 While the time-on-task was similar for both the PI and TI groups, the use of audio in the computer-guided PI activity vs. written text in the TI activities resulted in greater variability in the TI group.

6 As this was part of a larger study, eight of these distractor sentences tested for adverb-verb mismatches. Four contained a transitive verb that took an inanimate object.

7 As the evaluation of WM was not a goal of this study, these data will not be analyzed further. See table S1 in Supplement 2 for group means and SDs.

8 No participant scored below 87% on either this test, indicating that every participant had retained the vocabulary and gender information. See table S1 in the Supplemental Materials for group means and SDs.

9 The *afex* package acts as a wrapper for the package *lme4* (Bates et al., 2015).

10 Descriptive results for SVO sentences are presented, though not included in the statistical analyses. All learners were highly accurate with SVO sentences and the nominative case throughout the experiment.

11 Score ~ Group * Time + (1 | Subject).

12 Score ~ Group * Time + (1 | Subject).

13 Score ~ Group * Time + (Time_F | Subject).

14 The final model was: LogRT ~Time * Word Order * Group + (Time * Word Order || Subject) + (Word Order | Item). This includes the full fixed effects structure with by-subject and by-item random effects. For the random effects structure, "single bar notation" (e.g., |) indicates that the random effects included the correlation between random intercepts and slopes as was done for "Items" in this model. "Double bar notation" (i.e., ||) indicates that the correlation was excluded as was done for "Subjects" in this model. See Singmann and Kellen (2019) for a guide on reading mixed model notation.

15 LogRT ~ Time * Word Order * Group + (Time * Word Order | Subject) + (Time * Word Order || Item).

Conflict of Interest. I have no known competing interests to declare.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1–48. doi: 10.18637/jss.v067.i01.

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*, **5**(2), 95–127. http://doi.org/10.1177/136216880100500202

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

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

Benati, A. (2020a). An eye-tracking study on the effects of structured input and traditional instruction on the acquisition of English passive forms. *Instructed Second Language Acquisition*, **4**(2), 158–179.

Benati, A. (2020b). Online methods and tasks in research measuring the effects of processing instruction. *Instructed Second Language Acquisition*, **4**(2), 111–123.

Benati, A. (2020c). The effects of structured input and traditional instruction on the acquisition of the English causative passive forms: An eye-tracking study measuring accuracy in responses and processing patterns. Advance online publication. *Language Teaching Research*. https://doi.org/10.1177/1362168820928577

Bird, S. (2010). Effects of distributed practice on the acquisition of second language English syntax. *Applied Psycholinguistics*, **31**, 635–650. https://doi.org/10.1017/S0142716410000172

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

Cameron, R. D. (2011). *Native and nonnative processing of modality and mood in Spanish.* [Unpublished Doctoral Dissertation]. Florida State University.

Cheng, A. C. (2002). The effects of Processing Instruction on the acquisition of ser and estar. *Hispania*, **85**(2), 308–323.

Chiuchiù, G., & Benati, A. (2020). A self-paced-reading study on the effects of structured input and textual enhancement on the acquisition of the Italian subjunctive of doubt. *Instructed Second Language Acquisition*, **4**(2), 235–257.

Collentine, J. (1998). Processing Instruction and the Subjunctive. *Hispania*, **81**(3), 576–587.

Culman, H., Henry, N., & VanPatten, B. (2009). The role of explicit information in instructed SLA: An online study with Processing Instruction and German accusative case inflections. *Die Unterrichtspraxis/Teaching German*, **42**(1), 19–31. http://doi.org/10.1017/S0272263109990027

Dekydtspotter, L., Donaldson, B., Edomnds, A. C., Liljestrand Fultz, A., & Petrush, R. A. (2008). Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners. *Studies in Second Language Acquisition*, **30**(4), 453–480.

Dracos, M., & Henry, N. (2018). The effects of task-essential training on L2 processing strategies and the development of Spanish verbal morphology. *Foreign Language Annals*, **51**(2), 344–368.

Dracos, M., & Henry, N. (2021). The role of task-essential training and working memory in offline and online morphological processing. *Languages*, **6**(1), 24.

Ellis, N. C., & Sagarra, N. (2010). The bounds of adult language acquisition. *Studies in Second Language Acquisition*, **32**, 553–580. http://doi.org/10.1017/S0272263110000264

Farley, A. P. (2001). Authentic Processing Instruction and the Spanish subjunctive. *Hispania*, **84**(2), 289–299.

Farley, A. P. (2004). *Structured input: Grammar istruction for the acquisition oriented classroom.* McGraw Hill.

Fernández, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition*, **30**, 277–305. http://doi.org/10.1017/S0272263108080467

Harley, B., Howard, J., & Hart, D. (1995). Second language processing at different ages: Do younger learn-ers pay more attention to prosodic cues to sentence structure? *Language Learning*, **45**(1), 43–71.

Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA. *Studies in Second Language Acquisition*, **31**(04), 559. http://doi.org/10.1017/S0272263109990027

Henry, N., Jackson, C. N., & DiMidio, J. (2017). The role of explicit instruction and prosodic cues in Processing Instruction. *The Modern Language Journal*, **101**, 294–314. https://doi.org/10.1111/modl.12397

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, **22**(3), 369–397.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, **120**(4), 901–931.

Hwang, H. (2007). *Prosodic phrasing in sentence comprehension: Evidence from native English speakers and native Korean-speaking second language learners of English*. [Unpublished Doctoral Dissertation]. University of Hawai'i.

Isabelli, C. A. (2008). First noun principle or L1 transfer principle in SLA? *Hispania*, **91**(2), 465–478.

Issa, B. I. (2019). Examining the Relationships Among Attentional Allocation, Working Memory, and Second Language Development. In R. P. Leow (ed.). *The Routledge Handbook of Second Language Research in Classroom Learning* (pp. 464–479). New York: Routledge.

Issa, B. I., & Morgan-Short, K. (2019). Effects of external and internal attentional manipulations on second language grammar development. *Studies in Second Language Acquisition*, **41**(2), 389–417.

Issa, B. I., Morgan-Short, K., Villegas, B., & Raney, G. (2015). An eye-tracking study on the role of atten-tion and its relationship with motivation. *EUROSLA Yearbook*, **15**, 114–142.

Ito, K., & 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), 443–468.

Jackson, C. N. (2007). The use and non-use of semantic information, word order, and case markings during comprehension by L2 learners of German. *The Modern Language Journal*, **91**(3), 418–432.

Jackson, C. N. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, **58**(4), 875–909.

Jackson, C. N., & Dussias, P. E. (2009). Cross-linguistic differences and their impact on L2 sentence proc-essing. *Bilingualism: Language and Cognition*, **12**(1), 65–82.

Jackson, C. N., & Roberts, L. (2010). Animacy affects the processing of subject–object ambiguities in the second language: Evidence from self-paced reading with German second language learners of Dutch. *Applied Psycholinguistics*, **31**(04), 671–691.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, **44**(02), 137–166. http://doi.org/10.1017/S0261444810000509

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, **111**(2), 228–38.

Kasprowicz, R., & Marsden, E. (2017). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, **39**(6), 886–911. https://doi.org/10.1093/applin/amw051

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, **59**(3), 503–535. http://doi.org/10.1111/j.1467-9922.2009.00516.x

Keating, G. D., & Farley, A. P. (2008). Processing instruction, meaning-based output instruction, and meaning-based drills: Impacts on classroom L2 acquisition of Spanish object pronouns. *Hispania*, **91**(3), 639–650.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, **37**(1), 1–32.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in Second Language Acquisition*, **20**(04), 543–587. http://doi.org/10.1017/S0272263198004045

Lee, J. F. (2015). The milestones in twenty years of processing instruction research. *International Review of Applied Linguistics in Language Teaching*, **53**(2), 111–126.

Lee, J. F., & Doherty, S. (2019). Native and nonnative processing of active and passive sentences: The effects of processing instruction on the allocation of visual attention. *Studies in Second Language Acquisition*, **41**(4), 853–879.

Lee, J. F., & Doherty, S. (2020). Prior knowledge and other individual differences in the development of accuracy over the time-course of a pre-test/treatment/post-test study of instructed second language acquisition. *Instructed Second Language Acquisition*, **4**(2), 124–157.

Lee, J. F., Malovrh, P. A., Doherty, S., & Nichols, A. (2020). A self-paced reading (SPR) study of the effects of processing instruction on the L2 processing of active and passive sentences. *Language Teaching Research.* Advance online publication. https://doi.org/10.1177/1362168820914025

Lee, J. F., & VanPatten, B. (2003). *Making Communicative Language Teaching Happen* (2nd ed.). New York: McGraw Hill.

Leeser, M. J., Brandl, A., & Weissglass, C. (2011). Task effects in second language sentence processing research. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research: Insights from psycholinguistics* (pp. 179–198). Amsterdam, NL: John Benjamins.

Lenth, R. V. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.3. https://CRAN.R-project.org/package=emmeans

LoCoco, V. (1987). Learner comprehension of oral and written sentences in German and Spanish: The importance of word order. In B. VanPatten, T. R. Dvorak, & J. F. Lee (eds.) *Foreign language learning: A research perspective* (pp. 119–129). Rowley, MA: Newbury House.

Lovik, T., Guy, D., & Chavez, M. (2010). *Vorsprung (Enhanced E)*. Stamford, CT: Cengage Heinle.

Malovrh, P. A., Lee, J. F., Doherty, S., & Nichols, A. (2020). A self-paced reading study of language processing and retention comparing guided induction and deductive instruction. *Instructed Second Language Acquisition*, **4**(2), 203–234.

Marsden, E. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning*, **56**(3), 507–566. http://doi.org/10.1111/j.1467-9922.2006.00375.x

Marsden, E., & Chen, H. Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, **61**(4), 1058–1098. http://doi.org/10.1111/j.1467-9922.2011.00661.x

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In Mackey, A., & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1–21). New York: Routledge.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, **55**(3), 381–401. http://doi.org/10.1016/j.jml.2006.06.006

McManus, K., & Marsden, E. (2019). Signatures of automaticity during practice: Explicit instruction about L1 processing routines can improve L2 grammatical processing. *Applied Psycholinguistics*, **40**, 205–234.

McManus, K., & Marsden, M. (2017). L1 explicit instruction can improve L2 online and offline performance. *Studies in Second Language Acquisition*, **39**(3), 459–492. https://doi.org/10.1017/S027226311600022X

McManus, K., & Marsden, M. (2018). Online and offline effects of L1 practice in L2 grammar learning: A partial replication. *Studies in Second Language Acquisition*, **40**(2), 459–475. https://doi.org/10.1017/S0272263117000171

Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, **42**, 412–428. https://doi.org/10.1016/j.system.2014.01.014

Miller, A. K. (2011). Rethinking nonnative processing constraints: Evidence from L2 French. In L. Plonsky & M. Schierloh (Eds.), *Selected Proceedings of the 2009 Second Language Research Forum* (pp. 109–120). Somerville, MA: Cascadilla Proceedings Project.

Miller, A. K. (2014). *(Re)Activation of referents during recond language sentence processing: A self-paced reading study*. [Paper presentation]. The 2014 Second Language Research Forum, Columbia, SC.

Moeller, J., Adolf, W., Hoercherl-Alden, G., & Berger, S. (2009). *Deutsch Heute* (9th E.). Boston, MA: Cengage Heinle.

Pliatsikas, C., & Marinis, T. (2012). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, **16**(01), 167–182.

Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language Teaching Research*, **12**(1), 61–82. http://doi.org/10.1177/1362168807084494

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of cognitive neuroscience*, **18**(12), 2030–48.

Sagarra, N. (2007). Working memory and L2 processing of redundant grammatical forms. In Z. Han (Ed.), *Understanding Second Language Process* (pp. 133–147). Clevedon: Multilingual Matters.

Schlesewsky, M., Fanselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth & L. Konieczny (Eds.), *German Sentence Processing* (pp. 65–93). Dordrecht: Kluwer.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2012). *E-Prime v 2.0.10 [Computer Software]*. Pittsburgh, PA: Psychology Software Tools Inc.

Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, **34**, 499–520.

Seibert-Hanson, A., & Carlson, M. (2014). The roles of first language and proficiency in L2 processing of Spanish clitics: Global effects. *Language Learning*, **64**(2), 310–342.

Singmann, H. (2021). Mixed Model Reanalysis of RT data. Retrieved April 5, 2021, from https://cran.r-project.org/web/packages/afex/vignettes/afex_mixed_example.html

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Schachar, M. (2021). afex: Analysis of Factorial Experiments. Retrieved from https://cran.r-project.org/package=afex

Singmann, H., & Kellen, D. (2019). An Introduction to Mixed Models for Experimental Psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). London: Psychology Press.

Thuleen, N. (1999). nthuleen.com. Retrieved from http://nthuleen.com/teach/grammar.html

Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2013). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, **17**(2), 237–257.

University of Wisconsin Testing and Evaluation. (2006). *German placement test [Assessment instrument]*. Madison, WI.

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

VanPatten, B. (Ed.). (2004). *Processing Instruction: Theory, Research, and Commentary*. Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B. (2015a). Foundations of processing instruction. *International Review of Applied Linguistics in Language Teaching*, **53**(2), 91–109. http://doi.org/10.1515/iral-2015-0005

VanPatten, B. (2015b). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (2nd ed., pp. 113–134). Mahwah, NJ: Lawrence Erlbaum.

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

VanPatten, B., Collopy, E., Price, J. E., Borst, S., & Qualin, A. (2013). Explicit information, grammatical sensitivity, and the First-Noun Principle: A cross-linguistic study in Processing Instruction. *The Modern Language Journal*, 97(2), 506–527. http://doi.org/10.1111/j.1540-4781.2013.12007.x

VanPatten, B., & Houston, T. (1998). Contextual effects in processing L2 input sentences. *Spanish Applied Linguistics*, **2**(1), 53–70.

VanPatten, B., Inclezan, D., Salazar, H., & Farley, A. P. (2009). Processing instruction and dictogloss: A study on object pronouns and word order in spanish. *Foreign Language Annals*, **42**(3), 557–575. http://doi.org/10.1111/j.1944-9720.2009.01033.x

VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, **18**(4), 495–510. Retrieved from http://journals.cambridge.org/abstract_S0272263100015394

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology*, **49**(1), 51–79.

Witzel, J., & Witzel, N. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, **33**, 419–456.

Wong, W. (2004). The nature of Processing Instruction. In B. VanPatten (Ed.), *Processing Instruction: Theory, Research, and Commentary* (pp. 33–67). Mahwah, NJ: Lawrence Erlbaum.

Wong, W., & Ito, K. (2018). The effects of processing instruction and tranditional instruction on L2 online processing of the causative construction in French: An eyetracking study. *Studies in Second Language Acquisition*, **40**(2), 241–268.

Ying, H. G. (1996). Multiple constraints on processing ambiguous sentences: Evidence from adult L2 learners. *Language Learning*, **46**(4), 681–711.