

USING L1 EXPLICIT INSTRUCTION TO REDUCE CROSSLINGUISTIC EFFECTS IN L2 GRAMMAR LEARNING

ABSTRACT

This study advances previous research about the effects of explicit instruction on second language (L2) development by examining learners' use of verbal morphology following different types of explicit information (EI) and comprehension practice. We investigated the extent to which EI about L1 and L2 can reduce the effects of crosslinguistic influence in oral production. Sixty-nine English-speaking learners of L2 French undertook either: (a) a 'core' treatment of EI about the L2 with L2 comprehension practice, (b) the same L2 core + L1 comprehension practice, (c) the same L2 core + L1 practice + EI about L1, or (d) the tests only. Results showed that providing additional EI about the L1 benefitted the accuracy of oral production immediately after the instruction and then 6 weeks later. These results suggest that tailoring instruction to include information about the L1 could help reduce the effects of crosslinguistic influence and facilitate L2 learning.

Keywords: crosslinguistic influence; foreign language learning; French; instruction; oral production; first language; grammar

A major focus of second language acquisition (SLA) research to date has sought to understand the competition and relationships between a learner's different languages (Calabria, Costa, Green, & Abutalebi, 2018). This research has repeatedly shown that use of a single language activates a speaker's other known languages (Marian & Spivey, 2003; Wu & Thierry, 2010), that prior language experience use can influence second language (L2) use (e.g., selective attention to linguistic cues, Ellis, 2018; MacWhinney, 2012), and that cross-linguistic differences can influence the route and rate of L2 morphosyntactic development and L2 processing (Author; Isabelli, 2008; Murakami, 2016; Tolentino & Tokowicz, 2014). However, despite major advances in what we know about the cognitive effects and mechanisms of learning an additional language, little research has systematically examined the next step in this program: how can this understanding about competition and relationships between a learner's different languages be used to facilitate language learning?

Although explicit instruction remains a dominant approach in classrooms for reducing crosslinguistic influence during L2 learning (Ranta & Lyster, 2017), the extent to which it can actually benefit L2 morphosyntactic development constitutes a long-standing debate (for reviews, see DeKeyser, 2017; VanPatten, 2017). One line of research has compared practice with and without explicit information (EI) about the L2 target feature. While some research has shown that practice with EI provides few if any learning benefits compared to practice in making connections between forms and their meanings (Author; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996), others have found that EI about the L2 appears to play an important role by drawing learners' attention to specific (learnability) aspects of the target feature, thus enhancing the effectiveness of the practice (Henry, Jackson, & Dimidio, 2017; VanPatten, Borst, Collopy, Qualin, & Price, 2013).

These lines of investigation have helped us understand the effectiveness of EI about the L2. A notable consistency in this previous research, however, is that they have largely adopted presence/absence designs, examining broad effects of practice with and without EI (e.g., Andringa, de Glopper, & Hacquebord, 2011; Sanz & Morgan-Short, 2004; Tolentino & Tokowicz, 2014) or with and without comprehension (or production) practice (DeKeyser & Sokalski, 1996). Less research has manipulated the nature of one broad type of instructional component. One exception is Author, who manipulated the nature of the L2 input-based practice while holding the type and amount of EI constant across conditions. To our knowledge, however, little research has manipulated the nature of EI or practice to address specific learnability problems, including crosslinguistic influence. The current study addresses this gap by introducing L1 EI and L1 practice into L2 instruction.

We compared three types of EI and comprehension practice designed to improve English-speakers' use of the *Imparfait* (IMP) in L2 French, a target feature well-documented to be late-acquired due to complex L1-L2 form-meaning mapping differences (Author, Bartning, 1997, 2009; Howard, 2005; Kihlstedt, 2015): one group received EI about the L2 plus extensive L2 comprehension practice; a second group received the same L2 EI, L2 practice, plus additional L1 comprehension practice; and a third group received the same L2 EI, L2 practice, L1 practice, plus additional L1 EI. This design allowed us to compare (a) EI only about L2 form-meaning mappings with EI about both L2 and L1 form-meaning mappings and (b) comprehension practice only of L2 sentences with comprehension practice of both L2 and L1 sentences. Of particular interest was the extent to which explicit instruction about the L1 can address learning difficulties resulting from crosslinguistic influence at the level of form-meaning mappings.

We begin by reviewing SLA research about crosslinguistic influence in L2 grammar learning and follow this with an overview of research that has investigated EI about the L2 to address crosslinguistic influence. The extent to which EI about the L1 may be able to improve L2 learning is then briefly reviewed.

BACKGROUND

Crosslinguistic Influence in L2 Grammatical Learning

Research to date has repeatedly shown that a speaker's prior linguistic knowledge/experience can influence L2 grammatical learning in two specific ways. First, the same linguistic cues (or form-meaning mappings) can vary cross-linguistically (MacWhinney, 2005, 2012, 2018). Second, prior language knowledge/experience can influence attention to cues (Ellis, 2006, 2008; Wulff & Ellis 2018).

The Unified Competition Model (MacWhinney, 2005, 2012, 2018) proposes that crosslinguistic influence can be at least partly determined by differences in L1 and L2 cue 'availability' and 'reliability'. Cues can vary in type (morphological, syntactic, prosodic, semantic, and pragmatic), availability (how frequently cues are present), reliability (how often cues lead to the same interpretation), and validity (the joint product of availability and reliability). The Unified Competition Model predicts crosslinguistic influence when the validity of the same cue differs crosslinguistically. For example, in terms of viewpoint aspect (Smith, 1997), the availability and, to some extent, the type of viewpoint aspect cues in English and French are similar because these languages use verbal inflections to express perfectivity (the meaning of 'completeness' e.g., he walked to school yesterday) and imperfectivity (the meaning of ongoingness, e.g., he was walking to school yesterday; the meaning of habituality, e.g., he

walked to school everyday). However, the reliability and type of these English and French cues differ (see Table 1):

- Past perfectivity and past habituality (one sub-type of imperfectivity) can be cued by the same verbal inflection in English (*Simple Past*, Comrie, 1976; Tagliamonte & Lawrence, 2000), but not in French (*Passé Composé* for perfectivity; IMP for past habituality, Hoffmann, 1995).
- Past habituality and past ongoingness can be cued by the same verbal inflection in French (IMP, Hoffmann, 1995), but not in English (*Simple Past*, *would*, and *used to* for habituality; *Past Progressive* for ongoing).

Since the validity (the product of availability and reliability) of cues for imperfectivity is different in each language, the Unified Competition Model predicts that this cue validity difference can give rise to crosslinguistic influence, and that L2 learners approximate use of L2 cues only gradually over time (Gass, 1987; MacWhinney, 1987). More specifically, the Unified Competition Model predicts that expressing past habituality would be more difficult than ongoingness for L1 English learners of L2 French because of the validity of cues for habituality in English (see also Andersen, 1984; Slobin, 1973).

<TABLE 1 HERE>

Table 1
Viewpoint Aspect Meanings in French Sentences with English Glosses

Viewpoint meaning	French sentence with English gloss
Past habituality	<i>Elle jouait au foot (e.g., tous les jours)</i> 'She played / would play / used to play football (everyday)'
Past ongoingness	<i>Elle jouait au foot (e.g., quand + le Passé Composé)</i> 'She was playing football (when...)'
Past perfectivity	<i>Elle a joué au foot (hier)</i> 'She played football (yesterday)'

Patterns of difficulty that reflect these learning challenges are borne out by research into the SLA of French IMP. First, persistent restriction in the use and comprehension of IMP to the state verbs *avoir* and *être* has been observed (Ayoun 2004, 2013; Bartning, 2009; Kihlstedt, 2015), suggesting challenges for learners to generalize the IMP inflectional system to other verb types. Research also shows difficulties associated with IMP's different viewpoint aspect meanings that can be attributed to L1-L2 cue validity differences (Author; Ayoun, 2004, 2013; Howard, 2005; Kihlstedt, 2002). For example, while IMP is used to express both past ongoingness and past habituality, research indicates (a) that these viewpoint aspect meanings are acquired together and (b) that the order of acquisition of these meanings appears to be influenced by L1 background: Ongoingness acquired before habituality for English-speaking learners (Howard, 2005), but habituality acquired before ongoingness for Swedish-speaking learners (Kihlstedt, 2002). These observations suggest that the nature of L1 form-meaning mappings could play an important role in explaining IMP acquisition (see also Andersen, 1984; Ayoun, 2013, MacWhinney, 2005; Salaberry, 2008). For example, English expresses ongoingness with one form (one-to-one form-meaning mapping), whereas habituality can be expressed by a variety of forms (many-to-one form-meaning mapping). As a result, English speakers could be less

sensitive to habituality than ongoingness because habituality is dispersed across a wider variety of forms (Author; Andersen, 1984, Athanasopoulos & Bylund, 2013; Slobin, 1973). Focusing on habituality, English speakers have also been shown to initially use *Passé Composé* to express habituality (e.g., “*parfois je suis allée visiter mes amis à Paris le weekend*” from Howard, 2005, p. 188) while using IMP to express ongoingness (Author; Howard, 2005; see also Ayoun, 2004; Starren 2001). This usage reflects English’s form-meaning mappings and arguably an English speaker’s drive to grammatically distinguish one meaning from the other, as is done in by their L1: one form for ongoingness (Past Progressive in L1 English and IMP in L2 French) and a different form for habituality (Simple Past in L1 English and PC in L2 French).

An important question informed these lines of research is the extent to which instruction tailored to the nature of the learning problem (e.g., increasing learners’ sensitivity to the concept of habituality and its form-meaning mappings) can facilitate learning in cases of persistent crosslinguistic influence brought about by cue validity differences. In the following section, we review research designed to address these types of learning problems. We focus on two main approaches to reducing crosslinguistic influence in L2 grammatical learning: (a) explicit instruction about L2, (b) explicit instruction about L2 and L1.

Explicit Instruction about the L2

An important body of work informed by theoretical and empirical research about persistent crosslinguistic influence effects in L2 learning, especially for polyfunctional forms such as French IMP, has examined the extent to which instruction that addresses the cause of crosslinguistic influence can improve L2 learning (e.g., Cintrón-Valentín & Ellis, 2016; Ellis & Sagarra, 2011; VanPatten, 2015). One approach to this has been to provide explicit information

about language processing strategies (i.e. information about cues, what cues to attend to) followed by practice in order to develop more appropriate L2 processing behaviours (e.g., Author; Henry, Jackson, & DiMidio, 2017; Wong, Zhao, & MacWhinney, 2018; Tolentino & Tokowicz, 2014; Zhao and MacWhinney, 2018).

Based on evidence that the extensive prior use of the L1 tunes how speakers attend to language and subsequently biases which cues get noticed and processed (Ellis, 2006, 2008; Wulff & Ellis 2018), Ellis and colleagues manipulated attention during L2 processing to cues that might be missed due to L1-L2 cue validity differences (Cintrón-Valentín & Ellis, 2016; Ellis & Sagarra, 2011; Ellis et al., 2014; see also Dracos & Henry, 2018). Results at immediate posttest indicated that instruction targeting attention during processing can improve L2 grammatical learning by increasing attention to L2 cues that would have been missed due to crosslinguistic influence (resulting from a lifetime of prior L1 use). Ellis and Sagarra's (2011) meta-analysis of this body of research additionally indicated a graded effect of L1-L2 cue validity differences: Chinese speakers (no tense morphology) were found to be less able than speakers of Spanish and Russian (rich tense morphology) to learn inflectional cues in an inflectionally rich language (Latin).

Taken together, this body of research indicates that the training of language processing skills targeting (a) competing cues and (b) learned attention resulting from prior language use can improve L2 grammatical processing.

Explicit Instruction about the L1

In a recent review of language pedagogy research, Ellis and Shintani (2014) note that “there is almost no research that has investigated the actual effects of the classroom use of the L1

on L2 learning” (p. 247). Albeit a very small body of research, some studies have investigated EI about the L1 to address learning difficulties arising from crosslinguistic influence, specifically for lexis (Laufer & Girsai, 2008; White & Horst, 2012) and grammar (Horst, White & Bell, 2010; Kupferberg, 1999; Spada, Lightbown & White, 2005). These studies have compared interventions consisting of explicit, contrastive information about L1 and L2 with interventions of explicit information about L2. For example, to improve French-speaking learners’ use of possessive determiners in L2 English, Spada et al. (2005) provided EI about L1 and L2 highlighting that in French a possessive determiner agrees with the grammatical gender of the noun, but in English it agrees with the natural gender of the possessor. Learners were provided with ‘rule of thumb’ EI: “Ask “Whose is it?” If it belongs to a man or a boy, use his. If it belongs to a woman or girl, use her”. This EI was followed by classroom-based, communicative oral practice. For example, learners “played a game in which they had to describe their classmates without using their names: his hair is short and his t-shirt is yellow [...]” (p. 211). Pretest-Posttest results showed increased accuracy of possessive determiner use in writing and speaking tests and better verbalization of rules about when and how to use English possessive determiners. Similar benefits were reported by Kupferberg (1999) for Hebrew-speaking learners’ use of grammatical aspect forms in L2 English. Instruction required learners to translate Hebrew sentences into English, which was followed metalinguistic contrastive EI about structural and functional L1-L2 differences. Written production results showed that EI about L1-L2 structural and functional differences improved learners’ production of grammatical aspect forms, especially past perfect (for similar results, see also Kupferborg & Olshtain, 1996).

Taken together, these lines of research indicate benefits for providing EI about L1 and L2 combined with output practice. However, this research agenda still has some gaps. First, since no

delayed posttests were used and tests (largely) elicited language of a more controlled nature (rather than under time and oral communicative pressure), the durability and generalizability of learning gains remains unclear. Second, these studies have not addressed more complex cases of crosslinguistic influence at the level of form-meaning mappings arising from L1-L2 cue validity differences, like the case of IMP in L1 English learners of L2 French.

In addition, we observe that such research has only investigated the benefits of L2 practice. For example, although Spada et al. (2005) provided EI about L1 and L2, the practice was only in the L2. Thus, we do not yet know the extent to which practice involving interpretation (or production) of L1 *and* L2 sentences can reduce crosslinguistic influence in L2 learning. For example, following the tenets of Skill Acquisition theory (DeKeyser, 2017), EI about L1 followed by practice in interpreting the L1 may help consolidate declarative knowledge about the L1, and make L1 processing explicit in a way that serves more accurate L2 processing.

To address these gaps, Author provided EI about L2 and L1 form-meaning mappings for aspect in French (L2) and English (L1) and comprehension practice of both L2 and L1 sentences (unlike any of the aforementioned studies) to investigate their effects on L2 online and offline processing of aspect in L2 French. Author's instruction lasted 3.5 hours, delivered over four weeks. Results showed that EI about L1 and L2 processing routines followed by comprehension practice of French (L2) and English (L1) sentences improved learners' speed (online self-paced reading test) and accuracy (offline sentence judgement test in reading and listening) of aspectual interpretation (IMP, *Passé Composé*, *Présent*) four days after instruction (Immediate Posttest) and six weeks later (Delayed Posttest).

Although Author's evidence suggested that L1 EI (combined with L2 EI and L2 practice) benefited L2 online and offline performance in comprehension tests, we do not yet know the

extent to which it benefitted other skills, such as oral production. In line with calls to better understand the type of language knowledge and skills resulting from instruction, examining performance in oral production tests following comprehension practice would allow us to understand the extent to which practice can develop different types of language use (e.g. does comprehension practice only benefit performance on comprehension tests). Evaluating instructional effectiveness on tests that are not similar to the instruction itself and using more than one test is frequently recommended (e.g., Larsen-Freeman, 2015; Lightbown, 2008; Norris & Ortega, 2000), and it can be useful for both pedagogical (e.g., can teaching help language use in a range of contexts) and theoretical reasons (e.g., to inform our understanding of transfer appropriate processing, implicit/explicit knowledge accounts, skill specificity, and linguistic and cognitive models that foreground the role of input and input processing for learning). Thus, the present study set out to examine the extent to which instruction under a particular condition (i.e. comprehension) benefitted language use in a different condition (i.e. oral production). A few studies with L2 EI and L2 comprehension practice have assessed learning on semi-spontaneous oral production tests (e.g., Author; Sanz, 2004). However, to our knowledge, no previous research has investigated the extent to which EI about *L1* and comprehension practice of *L1* sentences can affect the accuracy of performance in L2 oral production.

In sum, the present study systematically investigated how differences in the type of EI (about only L2 form-meaning mappings vs. both L2 and L1 form-meaning mappings) and comprehension practice (of only L2 sentences vs. both L2 and L1 sentences) benefited the accuracy of L2 oral production. This extended our previous research showing that comprehension practice benefitted online and offline performance in comprehension tests.

THE PRESENT STUDY

The present study examined whether providing L2 learners with different types of EI (about L2 only vs. about L2 + L1) plus comprehension practice (interpreting L2 sentences only vs. interpreting L2 + L1 sentences) benefited the accuracy of IMP use in oral production outcome measures immediately after instruction and six weeks later, and whether the type of EI and/or comprehension practice moderated performance. We sought to address the following research questions:

1. To what extent can providing comprehension-based instruction (EI plus comprehension practice) improve the accuracy of IMP use in L2 oral production immediately after instruction (Posttest) and then six weeks later (Delayed Posttest)?
2. Compared to L2-only EI plus practice, to what extent are accuracy changes over time different for additional L1 practice with and without L1 EI?

METHODOLOGY

Participants

Participants were 69 university learners of French as a foreign language in semester two of a four-year Bachelor of Arts Honours degree program in French at a British university. All participants were L1 (British) English speakers, aged 18-21, had completed A2-level French (English school leaving qualification, equivalent to CEFR level B2, typically after 700-800 hours of instruction), and had not spent more than six weeks abroad in a French-speaking country (mean 3.3 weeks, SD=6.1). Mean years learning French was 10.3 (SD=2.7).

Target Feature

The target feature was French IMP inflectional verbal morphology, a past tense form used to express past habituality and ongoingness (e.g., *il jouait au foot* - ‘he was playing/used to play football’), selected because SLA research has repeatedly shown this form to be late-acquired due to functional complexity (Ayoun, 2004, 2013), including complex L1-L2 form-meaning differences for viewpoint aspect (see previous discussion and Table 1, Author; Howard, 2005; Kihlstedt, 2015). All exemplars of IMP were third-person singular: 25 regular (e.g., *marcher* ‘walk’) and 23 irregular (e.g., *courir* ‘run’) verb types balanced across 48 lexical verb types: twelve states (e.g., be happy), twelve activities (e.g., run in the park), twelve accomplishments (e.g., walk to the shop) and twelve achievements (e.g., find a letter). We balanced verb type frequency across these four lexical semantic classes using Lonsdale and Le Bras’s (2009) frequency dictionary of French.

Study Design

The study included three testing points (Pretest in week 1, Posttest in week 5, Delayed Posttest in week 12) and four groups (L2+L1, L2+L1prac, L2-only, Control). All treatments were administered via laptops using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2012). Participants were assigned to a group using matched randomization based on Pretest performance, resulting in 16 in the Control group, 17 in the L2-only group, and 17 in the L2+L1 group. 19 participants were in the L2+L1prac group. Treatments were delivered in four 45-minute sessions over three weeks: two sessions in week one, and one session each in weeks two and three. Each session had a different instructional focus: present vs. past ongoingness (session 1), present vs. past habituality (session 2), past ongoingness vs. past habituality (session 3), and past ongoingness vs. past habituality vs. past perfectivity (session 4). The Control group only

completed the Pretest, Posttest, and Delayed Posttests and received no treatment. Participants received no explicit French grammar instruction as part of their university program during the study, corroborated by interviews with university tutors. The whole study was piloted on a condensed timescale with ten comparable learners.

Instructional Treatments

All three instructional treatments included an identical core of L2 EI and L2 practice (see Appendix for example). For the full materials for these treatments see Author, IRIS [\[link\]](#), and Open Science Framework [\[link\]](#). This common core is briefly presented before describing the L1 treatment components uniquely received by the L2+L1 and L2+L1prac groups.

L2 EI. EI about the L2 was pre-practice, provided for approximately five minutes at the start of each session, and during-practice following incorrect answers (see Appendix for pre-practice EI used in Session 1). The pre-practice EI depicted conceptual information via a short video and images. For example, in Session 1, the concept of ongoingness was depicted using a short video of a man eating an apple bite by bite, but the apple never gets fully eaten. After seeing the video, participants were asked to think about (but not verbalize) how they might express in French what they just saw in video. Two possibilities were provided: *il mange une pomme* ('he is eating an apple') and *il mangeait une pomme* ('he was eating an apple'). Recommendations to aid processing were then provided. For example, attend to the verb ending to distinguish present from past ongoingness (-e vs. -ait in writing, mǎʒ vs. mǎʒɛ in speech [the EI used audio recordings for speech, not IPA]).

L2 comprehension practice. Pre-practice EI was followed by form-meaning mapping comprehension practice of French sentences, in equal amounts of listening and reading, that

required learners to attend to the meanings expressed by IMP, *Présent* and *Passé Composé* to complete the task (i.e., verbal inflections were ‘task-essential’, see Loschky & Bley-Vroman, 1993). For example, Session 1’s aim was for learners to interpret IMP and Present inflections to distinguish between present ongoingness vs. past ongoingness, so learners first read or heard a French sentence (e.g., *il joue au foot* ‘he plays/is playing football’) and then had to select the stimulus’s meaning from fixed options (e.g., ‘right now’ vs. ‘in the past’) (see Table 2 for examples of the L2 and L1 practice sentences).

<TABLE 2 HERE>

TABLE 2

Examples of L2 and L1 Practice Used in Session 1 (English glosses provided for illustration)

Target meaning	Present ongoing	Past ongoing
French stimulus used in L2 practice (received by all treatment groups)	<i>Elle...</i> ‘She’	<i>Elle...</i> ‘She’
	<i>joue au foot</i> ‘is playing football’	<i>jouait au foot</i> ‘was playing football’
	<i>porte une cravate</i> ‘is wearing a tie’	<i>portait une cravate</i> ‘was wearing a tie’
English stimulus used in L1 practice (received by L2+L1 and L2+L1prac groups)	<i>He...</i>	<i>He...</i>
	<i>is drinking a glass of wine</i>	<i>was drinking a glass of wine</i>
	<i>is knocking at the door</i>	<i>was knocking at the door</i>
Response options	<i>Maintenant</i> [X] ‘Now’	<i>Maintenant</i> ‘Now’
	<i>Dans le passé</i> ‘In the past’	<i>Dans le passé</i> [X] ‘In the past’

The L2 practice included 552 exemplars (384 in IMP [192 ongoing, 192 habitual], 96 in *Présent*, 72 in *Passé Composé*), balanced across reading and listening. Aural stimuli were

recorded by two native French speakers. The French sentences were verified for authenticity and comprehensibility by 26 native French speakers.

L2+L1 treatment. In addition to the same L2 EI and L2 practice, the L2+L1 treatment included brief EI about English form-meaning mappings for viewpoint aspect, lasting approximately 3 minutes, which followed the same design as the L2 EI (see Appendix for L1 EI used in session 1). For instance, in Session 1, about present vs. past ongoingness, learners saw the same man-eating-the-apple video, were asked to think about how they might express in English what they just saw in the video. Two possibilities were provided: *he is eating an apple* and *he was eating an apple*. Recommendations to aid processing were then provided. For example, attend to the verb auxiliary (*is* vs. *was*) to distinguish present from past ongoingness.

The L1 practice followed the same design features as described for the L2 practice, but with fewer sentences: 160 English sentences (56 in Past Progressive [ongoing], 56 in Past Simple [habitual], 16 in Present Simple [habitual], 16 in Present Progressive [ongoing], 16 in Past Simple [Complete], equally balanced across reading, listening, and lexical aspect type). See Table 2 for examples of the L1 practice.

L2+L1prac treatment. This was very similar as the L2+L1 treatment, except that participants received no EI about English, neither before nor during the practice. Participants completed the exact same L1 practice as in the L2+L1 treatment.

Oral Production Outcome Measures

To examine the extent to which instruction under a particular condition (i.e. comprehension) benefitted language use in a different condition (i.e. oral production), two

different oral production tests were used to assess performance following comprehension practice. See IRIS and OSF for the full tests.

Picture-Based Oral Narrative (to elicit habitual IMP). Two picture-based narrative stories, the cat story and the sister story, as used in previous French L2 research (Author) and adapted from Dominguez, Tracy-Ventura, Arche, Mitchell, and Myles (2013), were used. Both stories were structurally similar and set in the past, involving unambiguous perfective contexts (for *Passé Composé* use) and habitual contexts (for IMP use). The stories contrasted the protagonists' long-standing daily routines (i.e. past habitual events) with a one-time event (perfective). For the cat story, pictures show the daily routines of a girl and her pet cat (habitual events), followed by one specific day when the cat escaped (one-time, perfective events). For the sister story, two adult sisters talk about recurrent childhood events (habitual events), followed by the events from one specific day of their holiday in Spain (one-time events). Short instructions in English were provided for completing the stories, a series of French lexical prompts to structure the stories, and a list of five French vocabulary items (two nouns and five verbs) for use when retelling the story. Participants were given two minutes to look through the pictures before telling the story. Both stories were piloted for equivalency with ten L2 learners and ten French native speakers.

Activity Description Oral Production Test (to elicit ongoing IMP). This test was designed to elicit descriptions of ongoing/interrupted events in the past. Learners were first shown an event in progress (e.g., a car driving down a road), and then shown the same event but with an interruption (e.g., a policeman stopping the car). The learner was asked to say in French what was happening (a context for IMP) before the intervening event happened, as shown in Figure 1.

Short instructions were provided at the start of the test. Participants did not look through the images before beginning. Two versions were created, each with 28 stimuli, 16 of which depicted ongoing events, equally balanced across the four lexical aspect classes. The remaining twelve events were distractors. Both versions were piloted for equivalency with ten L2 French learners and ten L1 French speakers.

The two versions of each test were administered in a split-block design to reduce test familiarity effects between consecutive test points (e.g., test version A at Pretest and Delayed Posttest, and test version B at Posttest).

<FIGURE 1 HERE>

FIGURE 1

Example Test Item From the Activity Description Oral Production Test



Data Coding and Analysis

All data were digitally recorded and then orthographically transcribed, by an expert user of French, using CHAT from CHILDES (MacWhinney, 2000) and protocols designed and tested for French SLA (see Author and [WEBSITE-LINK-REMOVED](#)). All transcripts were double checked for accuracy by the first author and one other expert user of French. CHAT transcripts were first automatically tagged for part-of-speech information using the French MOR program,

followed by automatic and manual disambiguation of initial part-of-speech taggings using the French POST program (Parisse & Le Normand, 2000). We used the %VCX program (Dominguez et al., 2013) to manually tag all verbal morphology for aspectual information (IMP, *Passé Composé*, *Présent*, Other), appropriateness of use (Appropriate, Inappropriate), and context (Habitual, Ongoing, Perfective). This tagging enabled automatic analysis of aspectual information. The CLAN command COMBO was used to automatically compute frequency counts for all combinations of form, (in)appropriateness of use, and context. The first author and a research assistant each coded the same 113 transcripts from each outcome test (25% of the total data) using %VCX, compared their codings, and discussed any differences. The first author coded the remaining files. Cronbach's alpha inter-rater reliability coefficients from these codings were .80 for the Picture-Based Narrative and .91 for the Activity Description Test.¹

Our analysis of IMP production used the 'target-like use' (TLU) metric (Pica, 1983; Ellis, 1994), which analyses a morpheme's distribution in both appropriate and inappropriate contexts (rather than just in appropriate contexts, as with 'suppliance in obligatory contexts'). TLU was computed using the frequency counts automatically generated by CLAN, as follows: $N \text{ of appropriate uses} / (\text{Total } N \text{ of appropriate contexts} + N \text{ of uses in inappropriate contexts})$. Because the stative verbs *avoir* (have) and *être* (be) are well-documented to be overused and rote-learned (Howard, 2005; Kihlstedt, 2015), they were excluded from our TLU analyses. Appropriacy of IMP use was determined as use of IMP to describe habitual and ongoing events. Habitual and ongoing contexts were determined according to the obligatory contexts elicited by the tests, as previously described. For instance, the use of *Présent* to describe a past habitual event was coded as inappropriate, whereas use of IMP to describe the same event was coded as appropriate.

Examination of descriptive statistics and graphics showed that the data were neither normally distributed nor had equal variances (according to box plots, Q-Q plots, and Shapiro-Wilks tests). We therefore present the results of 4 x 3 robust repeated measures (RM) ANOVAs with bootstrapped procedures (Larson-Hall, 2014), with Group as the between-subjects factor (L2+L1, L2+L1prac, L2-only, Control) and test point as the within-subjects factor (Pretest, Posttest, Delayed Posttest). We set the alpha level at .05. Although Mauchly's Test of Sphericity was not statistically significant ($p > .05$), the residual SSCP matrix showed deviations from Sphericity, so a Greenhouse-Geisser correction factor was used. No important deviations from normality and homogeneity of variances for the residuals were discovered. If, according to a robust RM-ANOVA, a statistically significant effect was found, pairwise comparisons with Bonferroni correction were used for the posthoc tests using the Games-Howell test for separate covariance matrices. Partial eta squared (η_p^2) is reported for all omnibus tests.

We used Cohen's d effect sizes (ES) and 95% confidence intervals (CIs) for d to interpret magnitudes of change for all between- and within-subjects paired comparisons (instead of p -values, Larson-Hall & Plonsky, 2015). Within-subject ES at Posttest were calculated using the mean and standard deviation of the Pretest as a baseline, and at Delayed Posttest using the Posttest as baseline. CIs for d that passed through zero were considered unreliable indicators of change (Field, 2013). We also calculated between-group ES changes with effects adjusted for baseline differences, similar to gains scores (Author, see supplementary materials). We draw on Plonsky and Oswald's (2014) Cohen's d field-specific benchmarks for interpreting our d values (within-subjects: 0.60 (small), 1.00 (medium), 1.40 (large); between-subjects: .40 (small), .70 (medium), 1.00 (large)), as well as effect sizes for relevant interventions found by relevant meta-analyses and (Shintani, Li & Ellis, 2013) individual studies (Author).

RESULTS

Habitual IMP in the Picture-Based Oral Narrative

In the picture-based oral narrative, there was a statistically significant two-way interaction between Time and Group ($F(5, 112) = 7.662, p = .000, \eta_p^2 = .264$), indicating between-group differences for appropriate IMP use over time. There were also statistically significant main effects for Time ($F(1.8, 112.1) = 43.705, p = .000, \eta_p^2 = .406$) and Group ($F(3,64) = 16.522, p = .000, \eta_p^2 = .436$).

Between-Group Differences in Habitual IMP Use. Group scores were compared at Pretest, Posttest, and Delayed Posttest (see Table 3).

At Pretest, comparisons confirmed no between-group differences (all CIs for d passed through zero, see Table 3). Appropriate IMP use for habitual events ranged from 31%-36% across all groups (see Table 1). Other forms inappropriately used in these past habitual contexts included PC (35%, examples 1-3) and, to a lesser extent, PRES (18%, examples 4-6).

1. *pendant sa jeunesse chaque soir (erm) Alex (erm) a fait erm ses devoirs (participant 214)*

‘during her youth, every evening Alex (erm) did^{-PAST PERFECTIVE} (erm) her homework’

2. *donc chaque matin Nathalie a lu son livre préféré à ses poupées (participant 219)*

‘so every morning Natahalie read^{-PAST PERFECTIVE} her favourite book to her dolls’

3. *pendant sa jeunesse chaque soir Alex elle a écrit beaucoup (participant 228)*

‘during her youth, every evening Alex wrote^{-PAST PERFECTIVE} a lot’

4. *chaque soir pendant sa jeunesse Alex fait des choses très calme (participant 212)*

‘every evening during her youth Alex does^{-PRESENT} things very calmly’

5. *chaque matin Nathalie peint un image et construit un maison des boîtes (participant 224)*

‘every morning Nathalie paints^{-PRESENT} a picture and builds^{-PRESENT} a house out of boxes’

6. *pour Pompon le chat (erm) chaque matin il dort (participant 242)*

‘for Pompon the cat (erm) every morning he sleeps^{-PRESENT}

Following training at Posttest, comparisons with Control showed large differences because of more appropriate IMP use in the treatment groups. At Delayed Posttest, only the L2+L1 group’s use of IMP to express past habituality was more appropriate than Control (large ES). We found no differences between (a) Control and L2+L1prac (negligible ES) and (b) Control and L2-only (negligible ES).

At Posttest, two of the between-treatment-group comparisons showed small but unreliable and negligible differences: L2+L1’s use of IMP was slightly more appropriate than L2+L1prac (small but unreliable ES because CIs for d passed through zero); L2+L1 and L2-only performed similarly (negligible ES). IMP scores in the L2-only group were higher than L2+L1prac (medium ES).

At Delayed Posttest, L2+L1’s scores were higher than both L2+L1prac (large ES) and L2-only (large ES). There were no Delayed Posttest differences between L2-only and L2+L1prac (negligible ES).

< TABLE 3 HERE>

< TABLE 4 HERE>

TABLE 3
Means (and Standard Deviations) for Habitual IMP (%TLU) in the Picture-Based Oral Narrative

	Pretest	Posttest	Delayed Posttest
L2+L1 ($n=17$)	31.18 (21.13)	80.51 (14.46)	76.10 (13.12)
L2+L1prac ($n=19$)	36.55 (22.75)	73.15 (7.58)	46.57 (24.92)
L2-only ($n=17$)	36.58 (21.61)	82.29 (11.8)	43.83 (22.19)
Control ($n=16$)	35.33 (23.24)	36.63 (23.54)	40.30 (24.52)

TABLE 4
Between-Group Comparisons for Habitual IMP in the Picture-Based Oral Narrative at Each Test Point (Mean Difference, Mean Standard Error (SE), p , and Cohen's d ES [with CIs for d])

	Pretest		Posttest		Delayed Posttest	
	Mean difference (SE)	p , d [CIs]	Mean difference (SE)	p , d [CIs]	p , d [CIs]	p , d [CIs]
L2+L1 vs. L2+L1prac	-5.37 (7.31)	.883, -.24 [-.90, .42]	7.36 (3.91)	.264, .65 [-.04, 1.30]	29.54 (6.68)	.001, 1.46 [.69, 2.16]
L2+L1 vs. L2-only	-5.40 (7.33)	.882, -.25 [-.92, .43]	-1.79 (4.53)	.979, -.13 [-.80, .54]	32.28 (6.25)	.000, 1.77 [.94, 2.52]
L2+L1 vs. Control	-4.15 (7.75)	.950, -.19 [-.87, .50]	43.87 (6.85)	.000, 2.26 [1.34, 3.07]	35.81 (6.91)	.000, 1.84 [.98, 2.60]
L2-only vs. L2+L1prac	-.03 (7.39)	1.00, .00 [-.66, .65]	-9.14 (3.35)	.051, .93 [1.60, .22]	2.74 (7.97)	.986, -.12 [-.77, .54]
L2+L1prac vs. Control	1.22 (7.81)	.999, .05 [-.61, .72]	36.52 (6.14)	.000, 2.17 [1.29, 2.95]	6.27 (8.49)	.881, .25 [-.42, .92]
L2-only vs. Control	1.25 (7.82)	.999, .06 [-.63, .74]	45.66 (6.54)	.000, 2.48 [1.52, 3.31]	3.53 (8.16)	.972, .15 [-.54, .83]

Note. Grey shading indicates reliable and meaningful ES because CIs for d do not pass through zero.

Within-Group Changes in Habitual IMP Use. We compared performance between the three test points (see Table 5). In the Control group, no reliable changes were found over time (negligible ES). All treatment groups improved between Pretest and Posttest (large ES). However, between Posttest and Delayed Posttest, appropriate IMP use decreased majorly for both L2+L1prac (large ES) and L2-only (large ES), to the extent that Pretest-Delayed scores were not different (negligible ES). In contrast, we found no differences between L2+L1's Posttest and Delayed Posttest scores (negligible ES), indicating that their Pretest-Posttest improvement was maintained. Parallel coordinate plots (see Figure 2) show these trajectories in detail (each line indicates an individual learner).

These results suggest that all three interventions improved learners' appropriate IMP use in semi-spontaneous oral production immediately after instruction (i.e., at Posttest). However, these gains were maintained six weeks later only for learners who had received L1 EI (i.e., the L2+L1 group).

<TABLE 5 HERE>

< FIGURE 2 HERE>

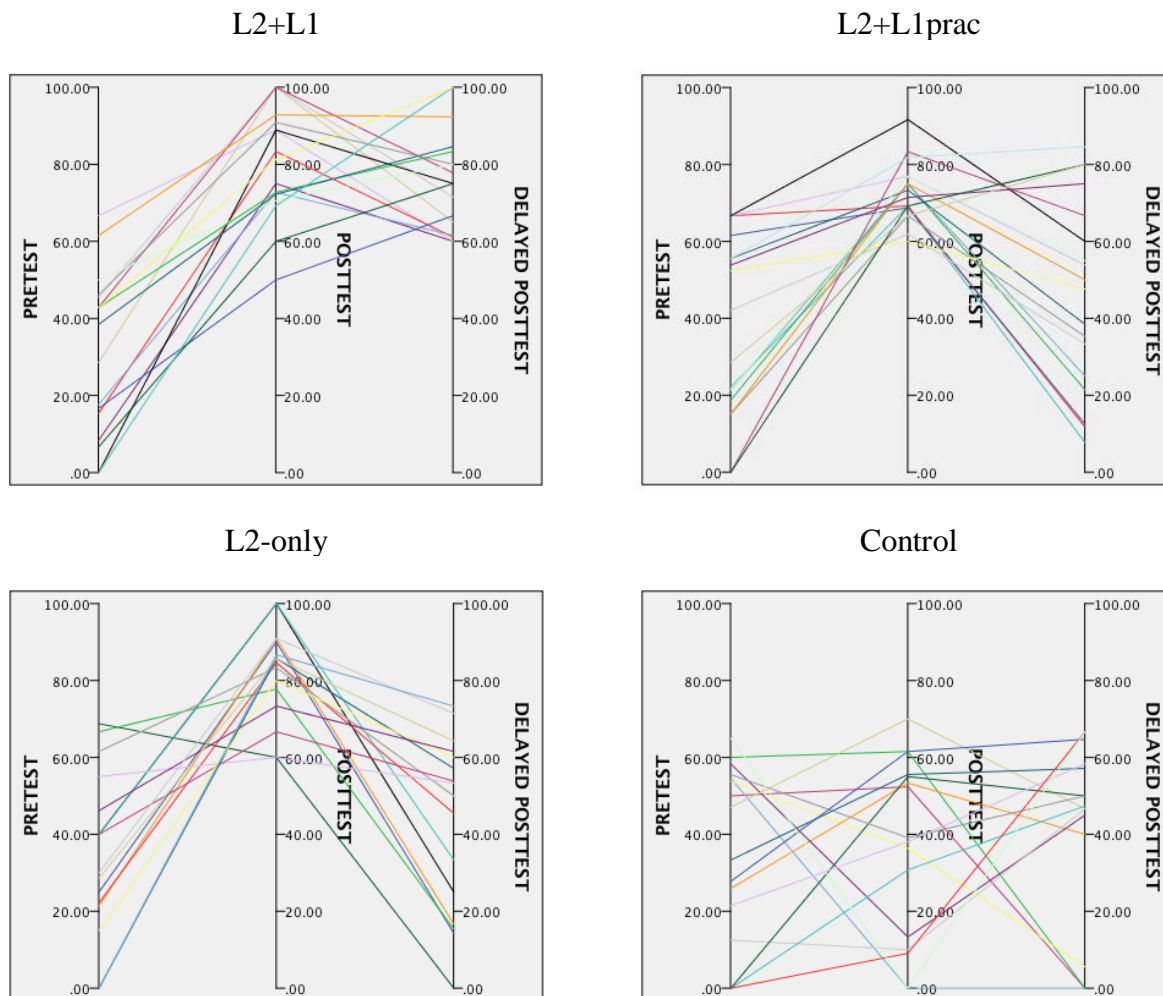
TABLE 5

Within-Group Comparisons for Habitual IMP in the Picture-Based Oral Narrative (Mean Difference, Mean Standard Error (SE), p , and Cohen's d ES with CIs for d)

	Pretest vs. Posttest		Pretest vs. Delayed Posttest		Posttest vs. Delayed Posttest	
	Mean difference (SE)	p, d [CIs]	Mean difference (SE)	p, d [CIs]	Mean difference (SE)	p, d [CIs]
L2+L1 ($n=17$)	-49.32 (4.75)	.000, 2.72 [1.74, 3.58]	-.44.92 (5.58)	.000, 2.55 [1.60, 3.39]	4.40 (4.83)	.376, -.32 [-.99, .36]
L2+L1prac ($n=19$)	-36.60 (5.42)	.000, 2.16 [1.32, 2.91]	-9.70 (8.49)	.270, .42 [-.23, 1.05]	26.33 (5.59)	.000, -1.44 [-2.12, -.070]
L2-only ($n=17$)	-45.71 (6.99)	.000, 2.63 [1.66, 3.47]	-7.24 (8.95)	.430, .33 [-.35, 1.00]	38.46 (6.06)	.000, -2.16 [-2.95, -1.27]
Control ($n=16$)	-1.30 (8.42)	.879, .06 [-.64, .75]	-4.97 (10.22)	.634, .21 [-.49, .90]	-3.67 (8.89)	.686, .15 [-.55, .84]

Note. Grey shading indicates reliable and meaningful ES because CIs for d do not pass through zero.

FIGURE 2
Parallel Coordinate Plots of Habitual IMP in the Picture-Based Oral Narrative



Ongoing IMP in the Activity Description Oral Production Test

In the Activity Description Oral Production Test, we found a statistically significant two-way interaction between Group and Time ($F(4, 97) = 9.285, p = .000, \eta_p^2 = .300$), indicating that ongoing IMP use varied between groups as a function of test point. There were also statistically significant main effects for Group ($F(3,65) = 33.957, p = .000, \eta_p^2 = .610$) and Time ($F(1.5, 97) = 83.680, p = .000, \eta_p^2 = .563$).

Between-Group Differences in Ongoing IMP Use. See Table 7 for all between-group comparisons. At Pretest, there were no meaningful between-group differences (all CIs for *d* passed through zero). Scores ranged from 35%-40% across all groups (see Table 6). Other forms inappropriately used in these past ongoing contexts included PRES (30%, examples 7-9) and, to a lesser extent, auxiliary + infinitive / present participle created forms (16%, examples 10-12).

7. *il quitte son travail (participant 219)*

'he leaves^{-PRESENT} his job

8. *il sonne la cloche (participant 206)*

'he rings^{-PRESENT} the bell'

9. *elle regarde un film (participant 250)*

'she watches^{-PRESENT} a film'

10. *il était écrivant un lettre (participant 228)*

'he was^{-AUXILIARY-PAST} writing^{-PRESENT PARTICIPLE} a letter

11. *il était sonner la cloche (participant 247)*

'he was^{-AUXILIARY-PAST} ring^{-INFINITIVE} the bell

12. *il était faisant le ski (participant 242)*

'he was^{-AUXILIARY-PAST} skiing^{-PRESENT PARTICIPLE} '

At both Posttest and Delayed Posttest, all treatment groups' IMP use was more appropriate than the Control group (large ES for all treatment group vs. control comparisons). These results contrast with our findings for habitual IMP, which showed no between-group differences at Delayed Posttest between (a) Control and L2+L1prac and (b) Control and L2-only.

Comparisons between the treatment groups showed no reliable differences at Posttest or Delayed Posttest. At Posttest, comparisons between L2+L1 versus L2+L1prac revealed a small but unreliable difference (CIs for d passed though zero) due to slightly higher scores in the L2+L1 group. No differences were found between L2+L1 and L2-only (negligible ES) and L2+L1prac and L2-only (negligible ES). At Delayed Posttest, no differences were found between L2+L1 versus L2+L1prac (negligible ES) and L2+L1 and L2-only (negligible ES). A small but unreliable difference (CIs for d passed though zero) was found between L2+L1prac and L2-only due to slightly higher scores in the L2-only group.

<TABLE 6 HERE>

<TABLE 7 HERE>

<TABLE 8 HERE>

TABLE 6
Means (and Standard Deviations) for Ongoing IMP in the Activity Description Oral Production Test

	Pretest <i>M(SD)</i>	Posttest <i>M(SD)</i>	Delayed Posttest <i>M(SD)</i>
L2+L1 ($n=17$)	35.95 (21.17)	80.66 (7.66)	77.15 (14.09)
L2+L1prac ($n=19$)	36.19 (22.32)	76.14 (9.26)	73.14 (9.14)
L2-only ($n=17$)	40.81 (17.74)	79.29 (9.08)	77.88 (10.33)
Control ($n=16$)	38.27 (21.59)	34.26 (18.93)	40.83 (19.68)

TABLE 7

Between-Group Comparisons for Ongoing IMP in Activity Description Oral Production Test at Each Test Point (Mean Difference, Mean Standard Error (SE), p , and Cohen's d ES with CIs for d)

	Pretest		Posttest		Delayed Posttest	
	Mean difference (SE)	p, d [CIs]	Mean difference (SE)	p, d [CIs]	Mean difference (SE)	p, d [CIs]
L2+L1 vs. L2+L1prac	-.24 (7.25)	1.00, -.01 [-.67, .64]	4.52 (2.82)	.391, .53 [-.15, 1.18]	3.55 (4.01)	.813, .34 [-.32, .99]
L2+L1 vs. L2-only	-4.86 (6.69)	.886, -.25 [-.92, .43]	1.37 (2.88)	.964, .16 [-.51, .83]	-.72 (4.24)	.998, -.06 [-.73, .61]
L2+L1 vs. Control	-2.32 (7.45)	.989, -.11 [-.79, .58]	46.40 (5.08)	.000, 3.25 [2.15, 4.20]	36.32 (5.99)	.000, 2.13 [1.23, 2.93]
L2-only vs. L2+L1prac	-4.62 (6.69)	.900, .23 [-.43, .88]	-3.14 (3.06)	.734, .34 [-.32, .99]	-4.27 (3.27)	.565, .49 [-.19, 1.14]
L2+L1prac vs. Control	-2.08 (7.44)	.964, -.09 [-.76, .57]	41.88 (5.19)	.000, 2.89 [1.89, 3.76]	32.77 (5.35)	.000, 2.17 [1.29, 2.95]
L2-only vs. Control	2.54 (6.90)	.983, .13 [-.56, .81]	45.03 (5.22)	.000, 3.06 [2.00, 3.98]	37.05 (5.52)	.000, 2.38 [1.44, 3.20]

Note. Grey shading indicates reliable and meaningful ES because CIs for d do not pass through zero.

TABLE 8

Within-Group Comparisons for Ongoing IMP in the Description Oral Production Test (Mean Difference, Mean Standard Error (SE), p , and Cohen's d ES with CIs for d)

	Pretest vs. Posttest		Pretest vs. Delayed Posttest		Posttest vs. Delayed Posttest	
	Mean difference (SE)	p, d [CIs]	Mean difference (SE)	p, d [CIs]	Mean difference (SE)	p, d [CIs]
L2+L1 ($n=17$)	-44.71 (5.67)	.000, 2.81 [1.81, 3.68]	-41.19 (7.05)	.000, 2.29 [1.38, 3.09]	3.51 (3.04)	.265, -.31 [-.98, .37]
L2+L1prac ($n=19$)	-39.95 (5.73)	.000, 2.34 [1.47, 3.11]	-37.41 (5.66)	.000, 2.17 [1.33, 2.92]	2.54 (1.97)	.214, -.33 [-.96, .32]
L2-only ($n=17$)	-38.48 (4.79)	.000, 2.73 [1.74, 3.59]	-37.07 (5.02)	.000, 2.55 [1.60, 3.39]	1.41 (2.32)	.551, -.14 [-.81, .53]
Control ($n=16$)	4.01 (6.59)	.552, -.20 [-.89, .50]	-2.56 (8.19)	.759, .12 [-.57, .81]	-6.57 (6.29)	.313, .34 [-.37, 1.03]

Note. Grey shading indicates reliable and meaningful ES because CIs for d do not pass through zero.

Within-Group Changes in Ongoing IMP Use Over Time. See Table 8 for all within-group comparisons. For the Control group, scores did not change over time (negligible ES for all comparisons). For all treatment groups, we found major improvement between Pretest and Posttest (large ES) and between Pretest and Delayed Posttest (large ES). There was no reliable change for any treatment group between Posttest-Delayed Posttest (negligible ES). The parallel coordinate plots in Figure 3 show these trajectories at the level of individual learners.

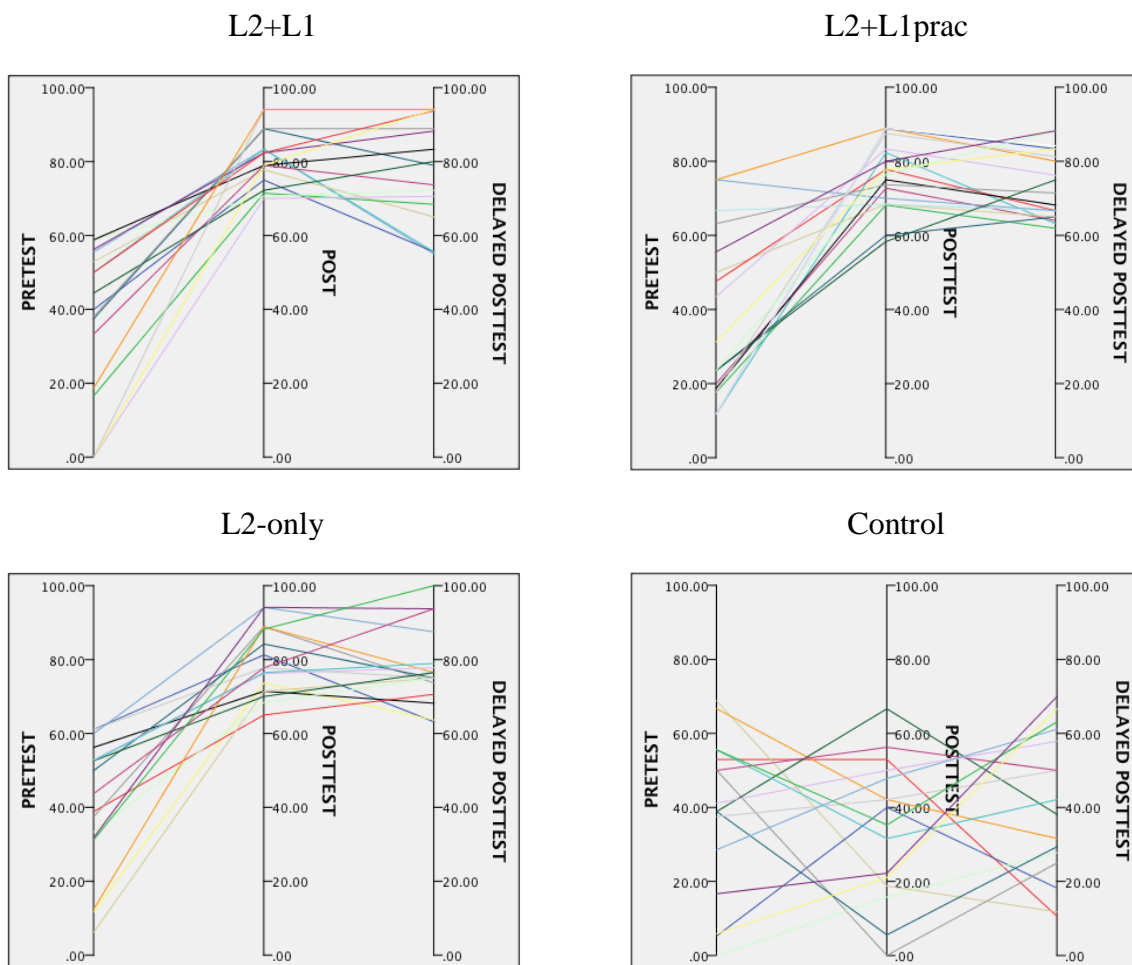
Taken together, our results indicate that the L2+L1, L2+L1prac, and L2-only treatments all led to more appropriate use of both habitual and ongoing IMP immediately after instruction (i.e., at Posttest). However, six weeks later, at Delayed Posttest, we found clearly different

patterns: For habitual IMP, only the L2+L1 group retained their gains at Delayed Posttest; for ongoing IMP, all treatment groups retained their gains.

<FIGURE 3 HERE>

FIGURE 3

Parallel Coordinate Plots of Ongoing IMP in the Activity Description Oral Production Test



DISCUSSION

The present study examined the extent to which different types of EI (L2 only vs. L2 + L1) about viewpoint aspect and comprehension practice (of L2 sentences only vs. of L2 + L1 sentences) improved L2 learners' oral production of the French IMP to describe ongoing and habitual events, immediately after instruction (at Posttest) and then six weeks later (at Delayed Posttest).

Summary of Results

For past habitual events, all treatment groups showed major improvement in the use of IMP immediately after instruction (large ES between Pretest-Posttest). This appeared to be temporary improvement for the L2+L1prac and L2-only groups (with large negative ES Posttest-Delayed Posttest), but more durable for the L2+L1 group (negligible ES between Posttest—Delayed Posttest). Between-group comparisons with the Control group confirmed these trends (see also *d* change scores in supplementary materials): ES at Posttest were large because of more appropriate use of IMP in all treatment groups. At Delayed Posttest, however, ES between Control and L2+L1 were large because the L2+L1 group's Pretest-Posttest improvement was maintained, but ES between Control and L2+L1prac and Control and L2-only were negligible because the L2+L1prac and L2-only groups' Pretest-Posttest improvement was lost. In summary, all treatments appeared to improve learners' habitual IMP use in oral production in a discourse-level test immediately after the instruction, but six weeks later only the effects of L2+L1 treatment - the only treatment that included EI about the L1 - were detectable.

For past ongoing events, we found major improvement for all treatments between Pretest-Posttest (large ES), and these gains were retained at Delayed Posttest (negligible ES between

Posttest and Delayed Posttest). Comparisons with Control showed large ES for all treatment groups at both Posttest and Delayed Posttest because of more appropriate IMP use in the treatment groups. In contrast to our findings for habitual IMP, then, all treatments appeared to improve learners' use of ongoing IMP in oral production immediately after the instruction with effects additionally detectable six weeks later.

These oral production results are consistent with Author's previously discussed findings for comprehension, which showed that the L2+L1 treatment (i.e., providing L1 EI with L1 practice alongside a core of L2 EI with L2 practice) improved the speed (self-paced reading test) and accuracy (sentence judgement test in reading and listening) of L2 comprehension of habitual and ongoing IMP immediately after instruction with gains retained six weeks later.

However, we found some differences between our findings for oral production (current study) and comprehension (Author) for the effects of the L2+L1prac and L2-only treatments, which could be related to the nature of the tests used. First, for habitual IMP, the current study found improved oral production for both these groups at Posttest, but Author found no improvements in reading comprehension (at either Posttest or Delayed Posttest). Second, for ongoing IMP, the current study found improved oral production for these groups at Post and Delayed Posttest, but Author found no improvements in reading comprehension for L2+L1prac, and only limited improvement for L2-only. Two factors could potentially explain these results. First, the comprehension tests were arguably more controlled than the production tests because the comprehension tests required learners to respond to IMP uses in specific sentences. The production tests, however, were less controlled (especially the Picture-Based Narratives) thus providing, to some extent, more flexibility in how particular viewpoint aspect meanings were expressed. Second, the comprehension tests required learners to read and judge L1 and L2

sentences (read an L1 context, read an L2 sentence, and judge how well they were matched for meaning), whereas the current study required L2 oral production without L1 production. It is possible that requiring learners to work with and switch between L1 and L2 (as in Author) was more challenging than producing L2 sentences only (current study), especially for learners whose training did not involve L1-based training (as discussed in Author). These could be possible explanations for why the L2+L1prac and L2-only appeared to perform better at Posttest in the oral production tests than in the comprehension tests.

Taken together, then, two trends emerge from the current study's oral production findings and those for comprehension as reported in Author. First, at immediate Posttest, all treatments improved their oral production of ongoing and habitual IMP, but only the L2+L1 and (to a lesser extent) L2-only treatments improved comprehension. Second, at Delayed Posttest, only the L2+L1 treatment led to improved production and comprehension of *both* ongoing *and* habitual IMP. Thus, our findings indicate that oral production and comprehension improvement was only found to be detectable six weeks after instruction for learners whose treatment included L1 EI, combined with L1 practice and the core, L2 EI and practice.

L1 Explicit Instruction to Address Crosslinguistic Influence

Consistent with the Unified Competition Model (MacWhinney, 2005, 2012, 2016), our finding that additional L1 EI plus practice benefitted L2 learners' use of habitual IMP but appeared to provide no additional benefits (compared to L2 EI and practice) for ongoing IMP could be explained by the nature of the crosslinguistic learning problem: L1-L2 form-meaning mapping similarities and differences for ongoingness and habituality.² SLA research on the acquisition of polyfunctional aspectual forms (like IMP) has shown that a form's different

functions tend to be acquired in stages, rather than all at once (Andersen, 1984; Bardovi-Harlig, 2000; Salaberry, 2008). For example, IMP's ongoing function has been found to be acquired before its habitual function among English-speaking learners of L2 French (Howard, 2005). This is consistent with evidence suggesting that how meanings/functions are expressed in the L1, especially the validity of linguistic cues crosslinguistically, combined with associative learning mechanisms, could explain this L2 developmental phenomenon (Bylund & Athanasopoulos, 2014; N. Ellis, 2006; MacWhinney, 2005). A meaning that is reliably cued by a single form is thought to increase speakers' sensitivity/attention to that meaning, which, as a result, is understood to be an important factor facilitating L2 learning due to the relative ease of mapping a new L2 form to one, consistent form and its associated concept in the L1. Conversely, a meaning cued by a variety of forms and/or covertly can result in reduced sensitivity/attention to that meaning, thus contributing to L2 learning difficulty, due to a more challenging mapping of an L2 form to a concept expressed by multiple L1 forms (Andersen, 1984; DeKeyser, 2005). Indeed, English form-meaning mappings for past ongoingness and habituality can be categorized in these ways: past ongoingness is reliably cued by *be V+ing* (one-to-one form-meaning mapping), but past habituality is cued by a variety of forms, including *-ed* (which also expresses past time), *would*, *used to* (many-to-one form-meaning mapping). As previously discussed, the Unified Competition Model predicts greater difficulty for expressing habituality than ongoingness for English-speaking learners of French because of these L1-L2 differences for cue validity. As a result, English speakers could be less sensitive to the concept of habituality than ongoingness because of how these meanings are cued in their L1 (Bylund & Athanasopoulos, 2014; MacWhinney, 2005).

It is possible that L1 EI might have helped improve learners' habitual IMP use by increasing their sensitivity to (a) the concept of habituality and (b) the ways in which habituality is expressed in their L1 and its subsequent mapping in the L2. In contrast, the relative conceptual saliency of ongoingness to these speakers, because of its one-to-one form-function mapping in their L1 English, may have meant that EI about L1 ongoingness provided no additional benefits, compared to the French comprehension practice + EI. (See Tolentino & Tokowicz, 2014 for evidence that EI, albeit about the L2, provided no additional benefits for forms that were similar in the L1-L2 compared to input practice via pairs of L2 sentences with a difference between them that isolated the target feature).

Relevant to this discussion is that L1 practice without L1 EI (i.e., the L2+L1prac treatment) appeared less helpful for habitual IMP use than L1 practice with L1 EI. This could suggest that L1 practice alone was not sufficient to clarify L1 form-meaning mappings for habituality, and leads us to consider that the combination of L1 EI plus L1 practice may have been the determining factor. We note that a L1 EI without L1 practice treatment was not investigated in the current study, and that the effectiveness of L1 EI without L1 practice remains to be empirically tested.

Our finding that the patterning of results deviated between treatment groups only at Delayed (but not at Posttest) could indicate that providing L1 EI about habituality led to more durable learning effects. As previously noted, our L1 EI about habituality was followed by comprehension practice in interpreting the meanings of L2 and L1 forms. It is possible that the frequency of this practice provided opportunities for consolidation and rehearsal of new conceptual representations that were less likely to deteriorate over time (DeKeyser, 2017). It is

also possible that the form-meaning mapping representations did deteriorate though to a lesser extent and/or more slowly, and so were not detected after 6 weeks.

Limitations and Future Research

Due to the small number of participants in each group, we note that our findings are tentative. We also note that we did not elicit the IMP's habitual and ongoing functions in a single test, but instead used different tests for each function. For these reasons, our conclusions require replication. It is possible that our findings for IMP's different functions (i.e., habitual vs. ongoing) may be partly related to test differences. The habitual test was a semi-spontaneous, discourse-level oral production test which required learners to construct a narrative, whereas the ongoing test was more controlled and mechanical in order to set up contexts to elicit ongoingness. It is possible performance was less demanding in the ongoing test and allowed (more) access to a more explicit knowledge type. However, we note that no change was found for the Control group, thus weakening the likelihood that artefacts of the test design are entirely responsible for our findings. If test type alone explained our findings, then the Control group could have drawn on existing EI about L2 past ongoingness, which is certainly part of their school curriculum prior to the current study. Given the lack of gains in the Control group, we consider it is more likely that the ongoing test did not simply allow gains to be observed very easily. We also note, as previously discussed, that previous empirical and theoretical SLA research corroborates the notion that IMP's ongoing function is more easily acquired than the habitual function by English speakers, providing secondary support for our claims.

Notwithstanding these limitations, our findings provide a number of directions for future research on differences between instructional components and their impact on L2 learning. For

example, it is unclear whether systematic production practice (L2 vs L2+L1), instead of comprehension practice, would lead to the same learning gains, or the extent to which altering the amount or spacing of practice would affect the findings. As previously noted, learners completed extensive L2 practice, but very little L1 practice in comparison. Although additional L1 practice without L1 EI (the L2+L1prac group) appeared to provide few additional learning benefits, providing larger amounts of L1 practice may lead to different results. Also, future research might even explore the effects of providing *only* L1 EI and L1 practice (i.e., without L2 treatments) to advanced-level learners for features with L1-L2 form-meaning differences to isolate the effects of clarifying L1 form-meaning mappings on L2 learning.

CONCLUSION

The current study examined the extent to which differences in the type of EI and comprehension practice improved the appropriacy of IMP use in L2 oral production. We provided three comprehension-based treatments: one group received EI about the L2 plus extensive L2 comprehension practice (L2-only group); a second group received the same L2 EI, L2 practice, plus additional L1 comprehension practice (L2+L1prac group); and a third group received the same L2 EI, L2 practice, L1 practice, plus additional EI about the L1 (L2+L1 group). A Control group received no instruction and completed only the Pretest, Posttest, and Delayed Posttest. This design allowed us to examine how differences in the type of EI (L2 vs. L2+L1) and type of comprehension practice (L2 vs. L2+L1) impacted L2 learning of viewpoint aspect in L2 French. Compared to L2-only and L2+L1prac, results showed that providing additional L1 EI benefitted the oral production of *both* habitual and ongoing IMP at six weeks after treatment. The other two treatment groups made gains at Posttest for both IMP meanings,

but these were only maintained at Delayed for ongoing IMP. Taken together, we argue that these results suggest that additional L1 EI benefitted learning of habitual IMP because it helped learners concretize a concept of past habituality that was more useful, to them as L1 English speakers, for learning French IMP. This helped learners to work out complex relations between L1-L2 form-meaning mappings, hypothesized to be a cause of L2 learning difficulty. Since, compared to habituality, ongoingness is relatively less complex in the L1 and is expressed morphologically, by one reliable cue, in both the L1 and L2, additional EI about the L1 appeared to provide no extra learning benefits. These results suggest that tailoring instruction, specifically the nature of the EI, to the nature of the learning problem facilitated L2 learning.

REFERENCES

- Andersen, R. W. (1984). The one to one principle of interlanguage construction. *Language Learning*, 34(4), 77–95.
- Andringa, S., de Glopper, K. & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning*, 61(3), 868-903
- Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, 37(2), 286–309.
- Author.
- Ayoun, D. (2004). The effectiveness of written recasts in the second language acquisition of aspectual distinctions in French: A follow-up study. *The Modern Language Journal*, 88(1), 31-55.

- Ayoun, D. (2013). *The second language acquisition of French tense, aspect, mood and modality*. Amsterdam: John Benjamins
- Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.
- Bartning, I. (1997). L'apprenant dit avancé et son acquisition d'une langue étrangère. Tour d'horizon et esquisse d'une caractérisation de la variété avancée [The advanced learner and their acquisition of a foreign language. Overview and outline of the advanced learner variety]. *Acquisition et interaction en langue étrangère*, 9, 9–50.
- Bartning, I. (2009). The advanced learner variety: Ten years later. In E. Labeau & F. Myles (Eds.), *The Advanced Learner Varieties: The Case of French* (pp. 11–40). Bern, Switzerland: Peter Lang.
- Calabria, M., Costa, A., Green, D. W., & Abutalebi, J. (2018). Neural basis of bilingual language control. *Annals of the New York Academy of Sciences*, 1426(1), 221–235.
- Cintrón-Valentín, M. C., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in psychology*, 7, 1284.
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge, UK: Cambridge University Press.
- DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen and M. Sato (Eds.), *Routledge Handbook of Instructed Second Language Acquisition* (pp. 15–32). London: Routledge.
- DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, 55(S1), 1–25.

- DeKeyser, R. M., & Sokalski, K. J. (1996). The differential role of comprehension and production practice. *Language Learning*, 46(4), 613–642.
- Domínguez, L., Tracy-Ventura, N., Arche, M. J., Mitchell, R., & Myles, F. (2013). The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism: Language and Cognition*, 16(3), 558–577.
- 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.
- Ellis, N. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164-194.
- Ellis, N. (2017). Salience in Usage-Based SLA. In S. M. Gass, P. Spinner & J. Behney (Eds.), *Salience in Second Language Acquisition* (pp. 21–40). London: Routledge.
- Ellis, N. C. (2008). The associative learning of constructions, learned attention, and the limited L2 endstate. In Robinson, P. & Ellis, N.C. (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372-405). New York, NY: Routledge
- Ellis, N. C. (2018). Salience in usage-based SLA. In Gass, S. M., Spinner, P., & Behney, J. (Eds.), *Salience in Second Language Acquisition* (pp. 21-40). New York, NY: Routledge
- Ellis, N. C., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, 33(4), 589-624.

- Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J., & Sagarra, N. (2014). An eye-tracking study of learned attention in second language acquisition. *Applied Psycholinguistics*, 35(3), 547-579.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. London: Routledge.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics*. Los Angeles: Sage.
- Gass, S. M. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics*, 8(4), 329-350.
- Henry, N., Jackson, C. N., & Dimidio, J. (2017). The role of prosody and explicit instruction in processing instruction. *The Modern Language Journal*, 101(2), 294–319.
- Hoffmann, L.-F. (1995). *L'essentiel de la grammaire française*. Englewood Cliffs, NJ: Prentice Hall
- Horst, M., White, J., & Bell, P. (2010). First and second language knowledge in the language classroom. *International Journal of Bilingualism*, 14(3), 331–349.
- Howard, M. (2005). Les contextes prototypiques et marqués de l'emploi de l'imparfait par l'apprenant du français langue étrangère [The prototypical and marked contexts of Imperfect use by the advanced learner of French]. In E. Labeau & P. Larrivée (Eds.), *Nouveaux développements de l'imparfait* [New developments of the Imperfect] (pp. 175–197). New York, NY: Rodopi.
- Isabelli, C. A. (2008). First noun principle or L1 transfer principle in SLA?. *Hispania*, 465-478.

- Kihlstedt, M. (2002). Reference to past events in dialogue: The acquisition of tense and aspect by advanced learners of French. In R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense-aspect morphology* (pp. 323–361). Amsterdam: John Benjamins.
- Kihlstedt, M. (2015). Acquisition of the imparfait in L2 French in adults and children: The same or different? A comparative case study of Swedish adults and children in an immersion setting. *Language, Interaction and Acquisition*, 6(1), 74–106.
- Kupferberg, I. (1999). The cognitive turn of contrastive analysis: Empirical evidence. *Language Awareness*, 8(3-4), 210-222.
- Kupferborg, I., & Olshtain, E. (1996). Explicit contrastive instruction facilitates the acquisition of difficult L2 forms. *Language Awareness*, 5(3-4), 149-165.
- Larsen-Freeman, D. (2015). Research into practice: Grammar learning and teaching. *Language Teaching*, 48(2), 263-280.
- Larson-Hall, J. (2014). *A guide to doing statistics in second language research using SPSS and R*. London: Routledge.
- Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, 65(S1), 127-159.
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied linguistics*, 29(4), 694-716.
- Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Han, Z. & Park, E.S. (Eds.), *Understanding second language process* (pp. 27-44). Bristol, UK: Multilingual Matters

- Lonsdale, D., & Le Bras, Y. (2009). *A frequency dictionary of French: Core vocabulary for learners*. London: Routledge.
- Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In Crookes, G. & Gass, S.M. (eds). *Tasks and Language Learning. Integrating Theory and Practice* (pp. 123-167). Bristol, UK: Multilingual Matters
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: The database*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B. (2012). The logic of the unified model. In Gass, S.M. & Mackey, A. (eds). *The Routledge handbook of second language acquisition* (pp. 211-227). New York, NY: Routledge
- MacWhinney, B. (2018). A unified model of first and second language learning. In Hickmann, M., Veneziano, E. & Jisa, H. (Eds.), *Sources of Variation in First Language Acquisition: Languages, contexts, and learners* (pp. 287-312). Amsterdam: John Benjamins
- MacWhinney, B. (Ed.). (1987). *Mechanisms of language acquisition*. London: Psychology Press.
- Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, 66, 834-871.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, 50(3), 417-528.
- Parisse, C., & Le Normand, M. T. (2000). How children build their morphosyntax: The case of French. *Journal of Child Language*, 27(2), 267–292.
- Pica, T. (1983). Methods of morpheme quantification: Their effect on the interpretation of second language data. *Studies in Second Language Acquisition*, 6(1), 69–78.

- Plonsky, L. & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning* 64, 4, 878–912
- Ranta, L., & Lyster, R. (2017). Form-focused instruction. In Garrett, P. & Cotts, J. (Eds.), *The Routledge Handbook of Language Awareness* (pp. 58-74). New York, NY: Routledge
- Salaberry, R. M. (2008). *Marking past tense in second language acquisition: A theoretical model*. New York. NY: Continuum.
- Sanz, C. (2004). Computer delivered implicit vs. explicit feedback in processing instruction. In VanPatten, B. (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 245-260). Mahwah, NJ: Lawrence Erlbaum
- Sanz, C., & Morgan–Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer–assisted study. *Language Learning*, 54(1), 35–78.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime 2.0: User’s guide*. Pittsburgh: Psychology Software.
- Shintani, N., Li, S., & Ellis, R. (2013). Comprehension–Based versus Production–Based grammar instruction: A Meta-Analysis of comparative studies. *Language Learning*, 63(2), 296–329.
- Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In *Ferguson, C.A. & Slobin, D.I. Studies of Child Language Development* (pp. 175-208). New York, NY: Holt, Rinehart & Winston
- Smith, C. (1997). *The parameter of aspect*. New York, NY: Springer:

- Spada, N., Lightbown, P., & White, J. (2005). The importance of form/meaning mappings in explicit form-focused instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 199-234). Berlin: DeGruyter
- Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological science*, 10, 281-284.
- Starren, M. (2001). *The second time: The acquisition of temporality in Dutch and French as a second language*. Utrecht, NL: DeGruyter
- Tagliamonte, S., & Lawrence, H. (2000). "I Used to Dance, but I Don't Dance Now" The Habitual Past in English. *Journal of English Linguistics*, 28(4), 324-353.
- Tolentino, L. & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning*, 64(2), 279-309
- VanPatten, B. (2017). Processing instruction. In S. Loewen and M. Sato (Eds.), *Routledge Handbook of Instructed Second Language Acquisition* (pp. 166-180). London: Routledge.
- VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, 18(4), 495-510.
- VanPatten, B., Collopy, E., Price, J., 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.
- White, J. L., & Horst, M. (2012). Cognate awareness-raising in late childhood: Teachable and useful. *Language Awareness*, 21(1-2), 181-196.

- Wong, M. H. I., Zhao, H., & MacWhinney, B. (2018). A cognitive linguistics application for second language pedagogy: The English preposition tutor. *Language Learning*, 68, 438-468.
- Wu, Y. J., & Thierry, G. (2010). Investigating bilingual processing: the neglected role of language processing contexts. *Frontiers in Psychology*, 1, 178.
- Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In Miller, D., Bayram, F., Rothman, J. & Serratrice, L. (Eds.), *Bilingual Cognition and Language: The state of the science across its subfields* (pp. 37-56). Amsterdam: John Benjamins
- Zhao, H., & MacWhinney, B. (2018). The Instructed Learning of Form–Function Mappings in the English Article System. *The Modern Language Journal*, 102, 99-119.

NOTES

1. Based on a meta-analysis of reliability coefficients in L2 research, Plonsky and Derrick (2016) propose that .83 (median = .92) should be considered a general (not absolute) threshold for an acceptable estimate of interrater reliability.
2. Given that both tests involved clear contrasts between PC (for perfective events) and IMP (for habitual, ongoing events), it seems unlikely that this aspect of the test design can explain performance differences for the IMP's habitual and ongoing functions.

APPENDIX

Description of the core L2-Only Treatment (Received by all Treatment Groups) and the Additional L1 EI and Practice Used in Session 1: Ongoingness (Present vs Past). For all Materials, see Author and IRIS

	Core L2-only treatment	Additional L1 components				
Pre-practice EI	<p>[Watch a six-second video clip of man eating an apple. The apple was never fully eaten.]</p> <p>To describe this you could say: <i>Il mange une pomme</i> Or <i>Il mangeait une pomme</i></p> <p>The difference between these two is: Il mange = ongoing action RIGHT NOW Il mangeait = ongoing action IN THE PAST</p> <p>The ends of the verbs distinguish between an ongoing action in the present <i>versus</i> past e.g. [Four verbs presented in pairs, aurally and in writing]:</p> <table><tr><td><i>Présent</i> RIGHT NOW</td><td><i>Imparfait</i> IN PAST</td></tr><tr><td>regarde [ʁəɡard]</td><td>regardait [ʁəɡardɛ]</td></tr></table>	<i>Présent</i> RIGHT NOW	<i>Imparfait</i> IN PAST	regarde [ʁəɡard]	regardait [ʁəɡardɛ]	<p>[Same video as L2-only treatment]</p> <p>To describe this you could say: <i>He is eating an apple</i> Or <i>He was eating an apple</i></p> <p>The difference between these two is: ‘he is eating’ = ongoing action RIGHT NOW ‘he was eating’ = ongoing action IN PAST”</p> <p>To identify ongoing meaning in the present <i>versus</i> the past, you need to focus on the auxiliary.</p> <p>Look/listen out for ‘is’ or ‘was’ to indicate whether it is an ongoing action taking place RIGHT NOW (present) or it is one IN THE PAST (past).”</p>
<i>Présent</i> RIGHT NOW	<i>Imparfait</i> IN PAST					
regarde [ʁəɡard]	regardait [ʁəɡardɛ]					
Practice	<p>96 French items (48 listening, 48 reading).</p> <p>Aim: Identify whether an ongoing event is taking place:</p> <p>“MAINTENANT” (right now) or “DANS LE PASSÉ” (in the past)</p>	<p>Additional 32 English items (16 listening, 16 reading).</p> <p>Aim: identify whether an ongoing event is taking place:</p> <p>“RIGHT NOW” or “IN THE PAST”</p>				

	<p>Example (English glosses not provided):</p> <p>Il...</p> <p>(1) fait du shopping ('is shopping')</p> <p>(2) faisait du shopping ('was shopping')</p>	<p>Example:</p> <p>He...</p> <p>(1) is eating a sandwich</p> <p>(2) was eating a sandwich</p>
<p>EI given immediately after incorrect responses during practice</p>	<p>After incorrectly responding 'MAINTENANT':</p> <p><i>"NOTE: The IMPARFAIT expresses an ongoing event DANS LE PASSÉ, not an ongoing event taking place MAINTENANT"</i></p> <p>After incorrectly responding 'DANS LE PASSÉ':</p> <p><i>"REMEMBER: The present tense in French expresses an ongoing event taking place MAINTENANT; not an ongoing action DANS LE PASSÉ"</i></p>	<p>After incorrectly responding 'RIGHT NOW':</p> <p><i>"The present tense in English ('is +ing') and in French expresses the same meaning: ongoing action taking place RIGHT NOW"</i></p> <p>After incorrectly responding 'IN THE PAST':</p> <p><i>"The past tense in English ('was +ing') is the same as the IMP in French (-ait). They both express an ongoing action IN THE PAST"</i></p>