

Input and output grammar instruction in tutorial CALL with a complex grammatical structure

Joseph Collentine and Karina Collentine

Abstract

Conceptual replication within a computer-assisted language learning (CALL) environment provides an understanding of the generalizability of second language acquisition (SLA) research (Porte, 2013; Smith & Schulze, 2013). The present study replicates Collentine (1998a), a classroom-based experiment framed within a larger discussion on the relative benefits of input- and output-based instruction. Collentine (1998a) compared the benefits of Processing Instruction (VanPatten, 2004) and output-based instruction, both of which elevated the Spanish subjunctive's communicative value. The results showed that input- or output-oriented instruction informed by how learners process grammatical information can affect the acquisition of complex grammatical phenomena. This conceptual replication not only seeks to corroborate the original study's findings in a new learning context. It also tests the finding's generalizability to a tutorial CALL environment built on 3D simulations and emerging web-app technologies. The participants were foreign-language learners of Spanish in a classroom-based curriculum (N = 50). The results indicate that, in the classroom and in a CALL environment, both input- and output-oriented approaches can promote the acquisition of a complex grammatical structure if practice is meaningful (informed by psycholinguistic processing principles) and deliberate, and if feedback is provided.

KEYWORDS: 3D ENVIRONMENT; CONCEPTUAL REPLICATION; EMERGING TECHNOLOGIES; SECOND LANGUAGE ACQUISITION; SUBJUNCTIVE; TUTORIAL CALL

Affiliation

Northern Arizona University, Department of Modern Languages, Flagstaff, AZ, USA.
email: Joseph.Collentine@nau.edu (corresponding author)

Introduction

Chun (2012) as well as Smith and Schultze (2013) advocate conceptual replication to test key second language acquisition (SLA) tenets in computer-assisted language learning (CALL) environments. Conceptual replication seeks to confirm the generalizability of a study's findings with a different methodology (Porte, 2013). Conceptual replication also helps materials designers to apply SLA learning principles to a range of new learning contexts. Thus, conceptual replication can help designers of CALL to create software that is theoretically sound.

As Mackey (2012) notes, one of the key issues of SLA replication going forward continues to be the relative importance of input-oriented – and thus output-oriented – approaches to language instruction. The present study is a conceptual replication of Collentine (1998a). Within a broader debate on the relative impact of input- and output-oriented instruction, the study showed that both modes equally promote grammatical development where psycholinguistic processing principles inform instructional design and the target structure is complex, like the Spanish subjunctive.

Review of the literature

Input and output in SLA and the need for conceptual replication

Long after Krashen's (1982) original proposal promoting the primacy of input, the relative impact of input- and output-oriented instruction on grammatical development is still debated. VanPatten and Benati (2010) argue that, while all theories recognize input's importance, output's contribution is debatable since research does not convincingly show that output affects the development of long-distance syntactic relationships nor that it has appreciable effects on phenomena governed by Universal Grammar. Gass, Behney, and Plonsky (2013) reason that input and output work together in SLA, although input's primary role is to provide positive and negative evidence (i.e., during conversational interaction) and it may be inconsequential for the development of syntax.

Input is important for learners to obtain the data necessary to establish the building blocks of the second language (L2). Learners need multiple exemplars of the L2 to extract grammatical patterns (Ellis, 2002). Additionally, learners do not begin with target-like processors (e.g., parsers) to analyze grammatical features, although input-oriented instruction such as Processing Instruction (PI) can provide learners with exemplars and psycholinguistic strategies for establishing the L2 grammar (VanPatten, 2012). VanPatten (2004) has identified psycholinguistic principles explaining what features in input learners attend to and ignore. For example, the *primacy of content*

words principle and the *lexical preference principle* explain why learners tend to acquire structures that have a high communicative value, such as tense markers in a passage where understanding the chronology of events is important. Additionally, the *sentence location principle* elucidates why learners are likely not to notice grammatical features appearing in sentence medial or final position, such as subordinate-clause verbs. Noting that grammatical structures have various semantic functions, VanPatten (2012) submits that input's primary role is the construction of grammar, speculating that 'input processing cannot address the issue of what happens after grammatical features are acquired' (p. 455).

Output appears to be particularly important to move learners *beyond* the initial stages of acquisition. Swain (1985) hypothesized that output was essential for learners to move from semantic to syntactic processing, where messages are conveyed and interpreted by considering the interaction of lexical, morphological and syntactic features. Swain (1995) emphasizes that output promotes the acquisition of morphosyntactic abilities: 'Output, thus, would seem to have a potentially significant role in the development of syntax and morphology' (p. 128). If output is accompanied by feedback (e.g., positive or negative feedback from a tutor) or if it leads to hypothesis testing, learners acquire greater metalinguistic awareness about the L2 grammar (Gass, 2010). Skehan (2009) argues that production (when time-pressures are relatively low) allows the learner to automatize the conceptualization and formulation of utterances, which facilitates the morphosyntactic complexification of L2 messages.

Toth (2006) asserts that whenever instruction makes 'form-meaning connections amenable to noticing and analysis' (p. 367) both input- and output-oriented instruction can effectively promote grammatical development. Moreover, complex grammatical structures may need both input- and output-oriented instruction, since they tend to be multifarious phenomena, which are most likely represented cognitively as procedural knowledge. DeKeyser and Prieto Botana (2013) contend that frequent 'comprehension and production' (p. 453) practice provides learners with the cognitive conditions under which they can proceduralize grammatical knowledge. One such complex structure is the Spanish subjunctive in adjectival clauses.

(1)

El alumno que estudie más siempre saldrá mejor.

[NP/antecedent El alumno] [CP que [V-subjunctive estudie] más] siempre saldrá mejor.

'The student who ends up studying most will always do better'

Collentine (1998a, 2002, 2013) argues that one of the difficulties in the subjunctive's acquisition is that students must learn to interpret and produce the

structure while attending to numerous morphosyntactic relationships. To produce a sentence such as (1) learners must be at an advanced stage in their development since they must reliably produce embedded clauses (cf., Collentine, 2013). To produce or interpret such sentences, they must be able to determine whether the antecedent *el alumno* has a definite or a variable referent and process long-distance syntactic relationships (cf. Farley, 2001) since the subordinate-clause verb *estudie* indicates that the antecedent is variable via the subjunctive inflection.

Mackey (2012) calls for conceptual replications of research on input processing principles, referencing DeKeyser, Salaberry, Robinson, and Harrington's (2002) *Language Learning* critique of VanPatten's psycholinguistic principles. Making particular note of Collentine (1998a), DeKeyser *et al.* (2002) surmised that input- and output- oriented instruction may be equally effective when learners make form-meaning connections with complex phenomena. The central question that Collentine examined was whether PI was superior to output-oriented instruction if VanPatten's processing principles were applied to both input- and output-oriented instruction. Up to that point, PI was largely shown to be superior to output-oriented instruction. Yet, that research did not entail complex structures on the whole and studies' output-oriented treatments were not informed by processing principles. Collentine focused on the Spanish subjunctive in adjectival clauses and both instructional modes (i.e., input, output) elevated the subjunctive's communicative value, revealing both instructional modes to be beneficial. In other words, Collentine showed that the acquisition of complex grammatical structures responds both to input or output when processing principles inform instructional design. To date, this relationship has not been studied beyond the classroom to other learning contexts. From 1998 up to 2014, the study has been cited more than 40 times in the SLA literature as one that raised questions about the relative impact of input- and output-oriented instruction and as one that provided insights into classroom learning of the subjunctive. Even so, the core finding that the acquisition of complex phenomena like the subjunctive is likely to occur when either mode is informed by processing principles has not received enough attention in the literature, as most of the discussion to date has focused on the relative impact of PI at the earliest stages of acquisition (cf., VanPatten, 2012). The present conceptual replication explores the generalizability of Collentine to CALL to examine whether the tenets hold up where instruction entails emerging technologies.

It should be noted that Farley (2001) presents results in line with Collentine (1998a), showing that, when instruction in either mode is meaningful and informed by processing principles, subjunctive development will occur. Notwithstanding Farley's important contribution to SLA's understanding

of the relationship between linguistic complexity and instructional mode, we conceptually replicate Collentine (1998a) because the focal processing principle informing the instructional design required more attention to form-meaning connections than the focal principle informing Farley, which centered on a structural principle relating to acoustic salience (cf., Collentine, 2002).¹ Collentine (1998a) did contain at least two notable limitations. First, it did not employ a delayed posttest, making unclear the treatments' long-term impact. Second, Collentine's (1998a) PI treatment employed only referential tasks (where learners use the subjunctive to make form-meaning connections), lacking any of the prescribed affective tasks (where learners use the subjunctive to express opinions or react emotionally). As will be shown below, while the present replication does not provide more insight into the long-term effects issue, it does bring the PI treatment in line with VanPatten's (2004) methodology by employing both referential and affective activities.

Apart from assessing the generalizability of Collentine's (1998a) insights into input- and output-oriented instruction, this replication study also provides insights into the potential of emerging technologies to support deliberate grammar practice with high contextualization in the form of a 3D world and by providing stimulus-rich, meaningful practice with web-app technologies. The use of these technologies affords us an additional advantage that meets conceptual replication's goal of informing SLA theory construction. Emerging technologies can be equipped with script-based tracking, which provides a depiction of the learning process that can be compared with the experiment's outcomes by recording learners' actions and responses to a database (Chapelle, 2001; Fischer, 2007). A shortcoming of much classroom-based research on the efficacy of instructional treatments is that we generally only advance SLA theory with information about the *product* of any given intervention, extrapolated from pretest-posttest comparisons. Script-based tracking data provides insights into the *process* of learning that can be compared with the product (Fischer, 2007). Research conclusions stemming from classic pretest-posttest comparisons are, for all intents and purposes, premised on the notion that learners performed uniformly during the treatment phase. However, even if all learners in a treatment phase completed the same 10 items, they likely completed the activity with varying success. Indeed, the only between-subjects variation considered in such classic analyses is what is recorded on the pretest and the posttest(s). Tracking data can be analyzed so as to allow a complementary analysis of the variable performance of learners during the treatment (e.g., with regression analyses; cf., K. Collentine, 2013). Essentially, with script-based tracking, the product can be understood in the context of the process.

SLA informed input and output instruction in tutorial CALL

In exploring input- and output-oriented instruction in CALL, it is important to consider the SLA premises informing the design of software. We focus on input and output in tutorial CALL, where instruction involves a machine as tutor and a human student, rather than in tool-oriented CALL, where a machine mediates human interaction as might be the case in computer-mediated communication (Hubbard & Siskin, 2004; Schulze, 2008). Hubbard and Siskin (2004) argue that tutorial CALL has erroneously been viewed as a behaviorist approach to learning, as the term 'drill and practice' is often rephrased derisively as 'drill and kill'. Hubbard and Siskin (2004) contend that, if properly designed, tutorial CALL can be informed by current information-processing theories.

To understand how we can help learners make form-meaning connections with complex structures like the subjunctive by elevating its communicative value in tutorial CALL (Collentine, 1998a, 2002; VanPatten, 2004), it is informative to update our understanding of 'practice' as a cognitive activity. According to skill acquisition theory (DeKeyser, 2007), 'practice' is the process of obtaining declarative knowledge (e.g., the Spanish subjunctive inflections) and building skills, sometimes referred to as procedural knowledge (e.g., correctly utilizing the subjunctive given certain pragmatic and syntactic conditions). To build new and reinforce existing knowledge as well as to proceduralize knowledge, practice should not be incidental but rather deliberate (Cerezo Ceballos, 2010; DeKeyser, 2007). In deliberate practice, learners complete activities that afford ample opportunities to process the structure. Additionally, learners complete activities where they do something meaningful with the structure (Leow, 2007). Receptive practice provides multiple exemplars of the target structure, and productive practice provides myriad opportunities to generate the structure. Meaningful practice is best attained by contextualizing the structure within some concrete situation (VanPatten & Benati, 2010). Contextualization requiring receptive practice can involve comprehension or interpretation, or even recognizing whether the structure relates to a given context (Cerezo Ceballos, 2010). Contextualization requiring productive practice can involve using the structure in speech or writing toward any number of communicative ends. For example, contextualizing the subjunctive in adjectival clauses (e.g., *Busco un apartamento que esté cerca del centro* 'I am looking for any apartment that is near downtown') may involve learners identifying correct uses of the structure or generating it appropriately in a situation where someone is considering housing options. Skill acquisition theory also acknowledges that deliberate practice greatly benefits from feedback mechanisms (Lyster & Saito, 2010). In deliberate practice, feedback models

correct usage or provides metalinguistic clues as to a structure's use (Ranta & Lyster, 2007), which are forms of positive evidence. Feedback can also draw attention to an error or prompt repair, which are forms of negative evidence.

How might deliberate practice be affected in tutorial CALL with readily available technologies? Web-app technologies built with HTML5, JavaScript, JQuery, and CSS3 can provide practice and feedback with an array of stimulus-rich presentational features and interactivity (Godwin-Jones, 2014). Collettine (1998b) as well as Russell (2012, 2014) posit that the vast array of multimedia features that CALL offers makes it particularly suitable for enhancing input (e.g., textual enhancement, the ability of students to control aural media). Receptive practice can take the form of the learner demonstrating comprehension or selecting an interpretation by a mouse click or screen touch. Current and emerging technologies (e.g., audio files, YouTube) offer a wide variety of input options that can be combined with engaging activities (e.g., audio samples embedded in an HTML5 interface). The computer as tutor can offer multiple exemplars of the targeted structure. CSS3 can provide various textual enhancements when providing exemplars or feedback (e.g., colorization, shadow effects). Production practice can be built on shallow morphological parsing (e.g., *canten* > *cant/* Stem + *e/*Present Subjunctive + *n/*3rd Person Subject) with JavaScript that analyzes output and provides feedback (e.g., Is the response relevant?, Is it close?; cf., Schulze, 2008). Presently, natural language processing (NLP) or other types of artificial intelligence (e.g., parsers, speech synthesis) for the purposes of providing focused corrective feedback are neither mainstream nor mature technologies (Beatty, 2012; Schulze, 2008). Accordingly, in tutorial CALL learners are largely limited to producing written rather than oral output, and such production must be brief in order to provide useful feedback. Still, software design motivated by NLP that restricts its evaluation of a learner's output to single words and short phrases (e.g., shallow morphological parsing with regular expressions) are posited to promote form-focused learning (Schulze, 2008). Emerging technologies can also provide rich contextualization for learning grammar. This is not only achieved with graphics, video and audio. Discussing how technology can promote SLA, Doughty and Long (2003) assert: 'New knowledge is better integrated into long-term memory and more easily retrieved if tied to real-world events and activities' (p. 58). 3D worlds can provide a visual and temporal situation for practice relating to a target structure. Kim, Lee, and Thomas (2012) argue that 3D worlds offer learners 'true simulations of real world conditions' (p. 3).

Research questions

This conceptual replication study addresses the following questions.

1. Does input- and output-oriented instruction in CALL where psycholinguistic processing principles inform instructional design affect the acquisition of complex grammatical phenomena like the subjunctive by foreign-language learners of Spanish?
2. What insights can tracking technologies provide into the relationship between learning processes and products?

Method

Participants

In the original Collentine (1998a) experiment (OE), participants were in a curriculum that promoted proficiency and that heavily emphasized grammatical competence. In the present conceptual replication experiment (RE), the participants' curriculum emphasized proficiency and was much more in line with the ACTFL 5-C competences (cf., http://www.actfl.org/sites/default/files/pdfs/public/StandardsforFLLEXsumm_rev.pdf).

Consequently, the RE participants' curriculum downplayed the role of grammar in favor of cultural knowledge and interdisciplinary connections. Thus, while the OE took place at the end of the second semester, the RE took place at the middle of the fourth semester, a course that elevated subjunctive instruction (as well as other constructs such as relative pronouns) to a level on par with that experienced by the participants in the original experiment. Both the OE and the RE took place following an introduction to the subjunctive in nominal clauses. It is our estimation that, although the OE students had fewer contact hours, the RE group's readiness as a function of curricular grammatical sequencing was adequately comparable. The RE did not include a control group because the OE indicated that participants receiving no treatment did not apply their knowledge of the pretest to their performance on the posttest. The OE had 54 participants, 18 in each of the groups (input, output, and control). The RE had 50 participants, 25 in the input treatment (INTR) and 25 in the output treatment (OUTR). Each RE group was an intact class.

Treatments

The major factors of both the OE and the RE are the target structure, language mode, and instructional approach.² Both experiments fostered the learners' abilities to use the subjunctive in adjectival clauses, as detailed in example (1). Both the OE and the RE contained an INTR and an OUTR.

In both the OE and the RE, the INTR entailed PI (VanPatten, 1993, 2004) and the OTR entailed production where the subjunctive's 'communicative value would be evident' (Collentine, 1998a: 580). In other words, in both the OE and the RE 'the principal factor distinguishing the two experimental groups was processing mode: one group cultivated its subjunctive abilities through input (albeit a specific type of input) and the other through output' (Collentine, 1998a: 580). In the OE and RE, the INTR and OTR learners were given strategies for appropriately processing the subjunctive in adjectival clauses. The processing principles informing the INTR and OTR were the *primacy of content words principle* and the *lexical preference principle*. A corollary to these principles is that learners process more meaningful morphology before less or non-meaningful morphology, and the assumption is that the subjunctive morphology is not very meaningful in adjectival clauses (cf., Collentine, 2010, 2013).³ Each INTR group was told that learners tend not to notice the meaning conveyed by the indicative and the subjunctive, and so they should examine the subordinate-clause mood to assess the referential nature of the antecedent. They were told to notice that the subjunctive connotes variable antecedents (e.g., *Compro un carro que vaya rápido* 'I'll buy a car [i.e., any car] that goes fast') while the indicative connotes definite antecedents (e.g., *Compro un carro que va rápido* 'I'll buy a car [e.g., which I have already seen] that goes fast'). Each OTR group was told that learners tend to forget about selecting subordinate-clause mood. They were told to assess the antecedent's referential status based on contextual clues and then produce the subjunctive if the antecedent was variable and the indicative if the antecedent had a referent. As discussed below, practice in the OE and the RE in both the INTR and the OTR sought to affect the acquisition of the subjunctive by raising its communicative value such that the mood of the subordinate clause was the sole mechanism by which the antecedent's referential or variable status was connoted. Providing practice where the subordinate-clause mood was the sole means of disambiguating the referential status of an antecedent converted the subjunctive into a more meaningful morpheme. Finally, so that processing was largely non-paradigmatic, examples and practice materials in the OE and the RE were limited to variable versus referential antecedents, present (indicative or subjunctive), and third person referents.

The OE and the RE had the same three phases and timeframes. First, students completed the pretest and a short test of their knowledge of the subjunctive morphology (see *Testing materials*). Second, each INTR and OTR occurred over two class periods, beginning with explicit information about the target structure and processing considerations, followed by practice (see *Instructional materials*). The OE took place entirely in a classroom setting,

with J. Collentine delivering the explicit information and strategies orally and with a handout; he also guided the INTR and OTR groups through the practice materials. The RE took place in a computer lab. J. Collentine delivered orally the explicit information and strategies, and tutorial CALL software provided the practice. The RE employed 3D technology and browser-based technologies built on HTML5, CSS3, JavaScript and JQuery for content delivery and interactivity, designed and authored by the researchers. In the RE, both researchers were available to manage the treatment procedures and to provide technical assistance. Finally, the posttest occurred the day after the intervention (see *Testing materials*).

Instructional materials

In comparing the two studies' instructional materials, we use the term *process* to refer to either producing or interpreting Spanish utterances. Additionally, a *target sentence* is one where learners processed a sentence containing an antecedent and an adjectival clause (e.g., *Quiere un coche que {cuesta [indicative], cueste [subjunctive]} poco*. 'She wants a car that {costs, might cost} little money). We likewise use *target verb* to refer to an adjectival clause verb inflected in either the indicative or the subjunctive. We compare each treatment according to linguistic channel (speaking, writing, listening reading), contextualization, linguistic features processed, and feedback.

PI treatments

Regarding the linguistic channel, both the OE and the RE participants processed target sentences in written and aural exemplars. The OE input was provided by an in-person instructor or on paper, whereas the RE input came in the form of digital audio or words on a web page. Both the OE and the RE groups worked with five reading and five listening activities. Concerning contextualization, the OE and RE practice activities were grouped into three sets, each relating to a particular situational context (e.g., shopping for a new car, furniture, gifts for a party). The classroom-based OE contextualized practice-activity sets with sentences (written or aural), short paragraphs (two to three sentences, written or read out loud), and caption cartoons. The CALL based RE contextualized practice-activity sets with a 3D virtual world built with the Unity game engine (see unity.com). Before each set, the OTR – as well as the INTR – explored a 3D marketplace for five minutes to search for gifts for a family member, to shop for new furniture, and to buy decorations for a party. Learners were instructed to interact with non-player characters (NPCs) and objects to find their gifts.



Figure 1: 3D world sample

With respect to linguistic features, PI employs two types of structured-input activities, both encouraging form-meaning connections. Referential activities require that the learner demonstrate the correct interpretation of a target sentence, e.g., *Quiere un coche que cueste [subjunctive] poco*. In both the OE and the RE, the subjunctive's communicative value was raised since the antecedent's referential status could never be ascertained by lexical redundancy such as an accompanying article, since, for example, *un coche* can have a definite or variable referent depending on the target sentence's context. Instead, target sentences could only be interpreted correctly by noticing the target verb's mood and ascertaining its connotation *vis-à-vis* the antecedent. This design feature ensured that the subjunctive morphology was more meaningful than it otherwise would be. Affective activities required learners to respond to target sentences with an opinion or personal preference. Accordingly, the OE contained seven sentence-level referential activities, where participants interpreted or reacted to target sentences in isolation to promote sentence-level processing. The OE also contained three discourse-level referential activities, where learners interpreted or reacted to target sentences embedded in a paragraph. The RE contained four sentence-level and two discourse-level referential activities, as well as two sentence-level and one discourse-level affective activities. Feedback was similar in the OE and RE. In both cases, the 'tutor' indicated whether a response was correct or not. In both cases, feedback for incorrect answers was also accompanied by reminders to consider whether the target verb was in the indicative or subjunctive, and what that implied about the referential status of the antecedent.

Output treatments

With respect to the linguistic channel, whereas the OE participants processed target sentences in writing and in speech, the RE participants only wrote target sentences using a keyboard. The OE participants worked with five writing and five oral activities, and the RE with 10 writing activities. In both the OE and the RE, the OTR contextualization was the same as that employed in the INTR. As for the features learners produced, in both the OE and the RE, the generation of the target verb's mood (indicative or subjunctive) was never determined or resolved by lexical redundancy such as by the antecedent's article but rather by the antecedent's referential status with respect to the target sentence's context: with the goal of elevating the subjunctive's communicative value and so the morphology's meaningfulness, the target verb needed the indicative if the antecedent had a referent (e.g., a particular inexpensive *coche* 'car') and the subjunctive if the antecedent was variable (e.g., any inexpensive *coche* 'car'). Nevertheless, whereas the OE participants processed whole and partial target sentences, the RE participants processed only partial target sentences. For both the OE and the RE, the production of partial target sentences entailed focusing learners on one or more features. OE and RE learners processed target verbs in eight simple-sentence activities. In both studies, learners might provide only the target verb in a sentence completion such as *Quiere un coche que _____ poco*. Alternatively, they might provide both the antecedent and the target verb, as in *Quiere _____ que _____ poco*. In a sentence completion activity learners might need to produce the antecedent, the relative pronoun (e.g., *que* 'that') and the target verb. It should be noted that the OE learners produced whole sentences in role-play activities, which was not technologically possible for the RE learners. OE and RE participants also processed target verbs in two discourse-level paragraph-completion activities.

Feedback was qualitatively different in the OE and RE. Whereas the instructor provided feedback in the OE, software routines gave feedback in the RE. In both the OE and the RE, learners were told whether their responses were correct or not, as well as whether their answers were *close* to being correct. In either case, after an incorrect answer, all participants were prompted to modify their responses, and the correct response was provided after a third try. Determining that a response was *close* in the OE required a judgment on the part of an instructor, and could relate to various linguistic features. A close response in the RE meant the following. First, consider a subjunctive form like *tengan* 'they might have' from the verb *tener* 'to have'. The stem is *teng*, the mood inflection *a*, and the person-number inflection *n*. If the RE participant generated the correct mood inflection along with either the correct verb stem or the correct person-number inflection, the software told the learner that the response was close whereupon he or she was prompted to modify

the response. If the RE participant needed to produce both a target verb and the antecedent, the software routines would independently assess both parts (e.g., checking the antecedent for an appropriate determiner, or lack thereof, or for number) and provide feedback as to whether each part of the partial sentence was correct, close or incorrect. This was accomplished by colorizing input boxes containing incorrect responses and informing the learner on the screen whether a response was incorrect or close; no further clue was given.

Sample CALL (RE) practice materials

Regarding sentence-level practice, both the INTR and OUTR RE groups would complete a task where they followed a protagonist as she considered potential gifts and described her decisions and/or considerations, as in Figure 2.

Your aunt, Tía María, wants do a bunch of upgrades to her house, adding some flowers for the party as well as new items like furniture, rugs, etc. Determine whether she is stating that she is open to possibilities or not.



“Esa planta con muchas flores es mi favorita.”

Figure 2: Sample 1: Practice context.

The INTR group would then select the best description of the situation by clicking one of the subordinate clauses in Figure 3 and feedback ensued.

Va a comprar una planta que ...

produce muchas flores.

produzca muchas flores.

Figure 3: Sample 1: Input sentence completion

The OUTR group would provide the appropriate target verb in the blank and feedback followed.

Va a comprar una planta que (producir) muchas flores.

Spanish characters: á é í ó ú ñ

Figure 4: Sample 1: Simple output sentence completion

As another example, both groups would complete a task where they followed the protagonist as she chose new décor for her house in anticipation of a party, such as in Figure 5.

Tía María wants do more upgrades to her house, adding art and photography. Determine whether she is stating that she is open to possibilities or not.



“No importa cuál. Me encantan los cuadros con pájaros volando.”

Figure 5: Sample 2: Practice context

The INTR group would then indicate which of two audio links – aural representations of the possible subordinate clauses – best described the situation, as shown in Figure 6.

Prefiere ver pájaros que ...

1

2

Figure 6: Sample 2: Input sentence completion

In this same context, the OTR group might produce the subordinate-clause verb.

As a final example, the OTR group also produced extended output in sentence completion activities to raise learners' metalinguistic awareness of target-sentence components, as is predicted to be the case in output production. These required that the learners produce not only the subordinate-clause verb, but also the antecedent noun phrase along with the relative pronoun, as in Figure 7.

Va a servir _____ (bebida) caliente _____ (Q) _____ (ser) totalmente natural.

Spanish characters:

Figure 7: Extended output sentence completion

Testing materials

An in-person, paper-based pretest/posttest procedure addressed the first research question. As in the OE, the RE employed a listening-comprehension instrument (10 items), a reading-comprehension instrument (10 items), and a written-production instrument (20 items) to measure the effects of the experiment. The participants also completed a ten-item test of their abilities to produce five indicative and five subjunctive forms, the results of which constitute a covariate in the analysis. Time constraints did not permit the inclusion of a distractor activity between the instruments in the RE, as was the case in the OE. The pretest and the posttest utilized different versions, whose combined correlation was corroborated with learners who had studied the structure, $r(22) = 0.44$, $p = 0.03$. Two raters assessed the assumed answers for the three instruments, both of which agreed completely on all items.

In the listening-comprehension test, participants indicated whether a sentence appropriately depicted a situation. For instance, the participants might hear from the instructor two versions of a sentence that differed only in subordinate-clause mood, such as *Paco quiere un coche que {cuesta [indicative], cueste [subjunctive]} poco* 'Paco wants a car that doesn't cost a lot'. Subsequently, based on a drawing and its caption, participants would indicate which version of the sentence best fit the depicted situation. The reading-comprehension test paralleled the listening-comprehension test, differing only in that participants read two versions of a sentence. The production test differed from the listening- and reading-comprehension tests in that the participants were given the subordinate clause verb's infinitive in parentheses and they had to write the correct conjugation (e.g., *Paco quiere un coche que _____ (costar) poco*).

On all tests, a correct answer received 1 point and an incorrect answer 0. For the sake of consistency, a Python script employing pattern-matching routines assessed the participants' production-test answers.

Tracking technologies

Script-based tracking technologies recorded to a database the NPCs and objects with which the learners interacted in the 3D world in addition to the resulting linguistic information that they read (e.g., questions that they chose to the pose, answers that they received, textual information about potential gifts). Script-based tracking technologies built into the web apps also recorded to a database learners' responses and accuracy during practice.

Analysis

To address the first research question regarding treatment gains, we employed two 2×2 repeated measure analyses of variance (ANOVA), one for the interpretation data (combining the listening and reading data) and another for the production data, as in Collentine (1998a). Each analysis had one within-subjects variable, test time (pretest and posttest), and two between-subjects variables, treatment (input and output) as well as the subjunctive morphology test results representing a covariate.

To address the second research question concerning insights that tracking technologies could provide into learning processes and products, we employed two regression analyses. We used a stepwise regression analysis where a predictor variable (i.e., independent variable) leaves the model if its contribution falls below a threshold of significance as measured by Akaike's (1974) information criterion, which produces a model with enough complexity (i.e., predictor variables) to optimize the amount of response-variable variation while not overfitting the model (Larson-Hall, 2009). Since the treatments differed substantially as concerns practice, we present one tracking-data analysis for the INTR and another for the OUTR. Regarding predictor variables, for both treatments, we operationalized learning processes as the language that learners read in the 3D world (contextualization phase) and their failure rate during the practice phase on their posttest gains. Recall that each student explored the 3D world as he or she wished, and so what learners read during the contextualization process would have varied between students. Failure rate indicates the extent to which learners varied in how well they performed during practice phases. After screening a variety of predictor variables, three factors demonstrated a high degree of variability among the participants. The variable *marked contextualizer verbs* represents the number of non-present tense verbs that a participant read in the 3D world (e.g., future, conditional, preterit, and imperfect). While none of these tenses constitute instances of the subjunctive, they indicate the extent to which a learner was exposed to exemplars of Spanish's inflectional complexity during

the treatment. The variable *subordinate conjunctions* represents the number of *que* ‘that’ and *si* ‘if’ subordinators read in the 3D world. Since the subjunctive is largely limited to dependent clauses, this measures the extent to which a participant was exposed to complex syntax in the contextualizer phase. The variable *failure rate* stems from the practice tracking data. For the input group, this metric represents the proportion of erroneous choices to all choices that a participant recorded during the practice phase (e.g., selecting a sentence with the indicative where the subjunctive was necessary). For the output group, this metric represents the proportion of incorrect items that a learner produced and recorded to all recorded production items (e.g., producing the subjunctive where an indicative form was necessary) including ‘close’ answers, as determined by the shallow morphological parsing routines built into the JavaScript code. The dependent variable was a combined listening, reading, and production difference score (posttest–pretest) on items targeting the subjunctive.

Results

The results indicate that both instructional approaches were equally effective at promoting learners’ subjunctive abilities in adjectival clauses, and neither had a clear advantage.

Table 1: Mean subjunctive interpretation scores by instructional group and test time

Instructional group	Test	M	n	SD	95% Confidence interval
Input	pretest	6.3	25	3.1	[5.1, 7.5]
	posttest	8.7	25	2.0	[7.9, 9.5]
Output	pretest	5.6	25	3.2	[4.3, 6.9]
	posttest	7.4	25	3.9	[5.8, 8.9]

Table 2: Repeated measures ANOVAs for subjunctive interpretation

Effect	Source	SS	df	MS	F	p	η^2
Within-subjects	Test time	92.9	1	92.9	13.1	0.001	0.218
	Test time × Subjunctive morphology	28.8	1	28.8	4.1	0.050	0.080
	Test time × Instructional group	1.8	1	1.8	0.3	0.613	0.005
	Error	333.3	47	7.1			
Between-subjects	Intercept	385.9	1	385.9	53.9	0.000	0.534
	Subjunctive morphology	245.9	1	245.9	34.3	0.000	0.422
	Instructional group	30.4	1	30.4	4.2	0.045	0.083
	Error	336.5	47	7.2			

Regarding the interpretation scores, which combined the learners' listening and reading scores, the ANOVA found a significant effect for instructional group, $F(1, 47) = 4.2, p = 0.045, \eta^2 = 0.083$. This reflects the fact that the INTR group scored better on both the pretest and posttest. Additionally, the analysis showed a significant effect for test time, $F(1, 47) = 13.1, p = 0.001, \eta^2 = 0.218$. However, the analysis found no interaction between instructional group and test time, $F(1, 47) = 0.3, p = 0.613, \eta^2 = 0.005$. Thus, as in the OE, both the INTR and the OTR RE groups realized significant and comparable gains from their respective instructional approaches in terms of their abilities to interpret the subjunctive in adjectival clauses.

The pretest interpretation scores were lower in the OE (i.e., INTR: $M = 1.4, SD = 1.3$; OTR: $M = 1.6, SD = 2.0$) than they were in this experiment. This conceptual replication study was conducted later in the learners' Spanish curriculum, such that prior knowledge most likely played a role in this experiment. Indeed, it is worthwhile to note that, in the original experiment the covariate prior knowledge of the subjunctive morphology did not affect the interpretation results, yet here it was a significant factor, accounting for a high degree of the results, $F(1, 47) = 34.3, p < 0.001, \eta^2 = 0.422$. Still, it must be kept in mind that the effect for test time actually takes into account the variation in prior knowledge of subjunctive morphology, yielding greater confidence in the assertion that both groups benefited from their respective instructional approaches in terms of their interpretation abilities with the subjunctive.

Table 3: Mean subjunctive production scores by instructional group and test time

Instructional group	Test time	<i>M</i>	<i>n</i>	<i>SD</i>	95% Confidence interval
input	pretest	5.4	25	3.6	[4.0, 6.7]
	posttest	6.5	25	3.0	[5.3, 7.6]
output	pretest	4.5	25	3.8	[3.0, 6.0]
	posttest	6.6	25	3.7	[5.0, 8.1]

Table 4: Repeated measures ANOVAs for subjunctive production

Effect	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2
Within-subjects	Test time	110.3	1	110.3	16.9	0.001	0.264
	Test time × Subjunctive morphology	57.0	1	57.0	8.7	0.005	0.156
	Test time × Instructional group	8.4	1	8.4	1.3	0.263	0.027
	Error	307.6	47	6.5			
Between-subjects	Intercept	96.7	1	96.7	11	0.002	0.190
	Subjunctive morphology	404.3	1	404.3	46	0.001	0.494
	Instructional group	5.9	1	5.9	0.7	0.417	0.014
	Error	413.4	47	8.8			

Concerning the production scores, the ANOVA found no effect for instructional group, $F(1, 47) = 0.7, p = 0.417, \eta^2 = 0.014$. Yet, there was a significant effect for test time, $F(1, 47) = 16.9, p = 0.001, \eta^2 = 0.264$, and no interaction was found between instructional group and test time, $F(1, 47) = 1.3, p = 0.263, \eta^2 = 0.027$. Again, as in the OE, both the INTR and the OTR RE groups realized significant and comparable gains from their respective instructional approaches in their production of the subjunctive in adjectival clauses.

The pretest production scores were lower in the original experiment (i.e., input group: $M = 0.6, SD = 0.9$; output group: $M = 0.8, SD = 1.2$), doubtless a function of prior subjunctive knowledge. Indeed, while prior subjunctive morphological knowledge did not affect the production results in the original experiment, here it was a significant factor, $F(1, 47) = 46, p = 0.001, \eta^2 = 0.494$. Similar to the interpretation analysis, since the effect for test time factors in the variation in prior knowledge of subjunctive morphology, we can be confident in our assertion that both groups benefited from their respective instructional approaches in terms of their production abilities with the subjunctive.

The analysis of the tracking data revealed that only the OTR group's gains were significantly affected by the language that they read in the 3D contextualization phase and their failure rate during the practice phase. Before conducting the stepwise regression analysis, the predictor variables were screened for basic assumptions (e.g., normality, collinearity). For both groups, subordinate conjunctions did not meet the normality requirements and the data were transformed with a square-root calculation since the distribution was moderately skewed.

The stepwise regression analysis revealed no relationship between the INTR group's overall gains from the experiment and the tracking variables. However, the analysis revealed that the OTR group's gains were significantly affected by the subordinate conjunctions that they read in the 3D world ($B = -0.34, ns.$) and their failure rate in the practice phase ($B = 0.30, ns.$), $F(2,22) = 3.42, p = 0.051, R^2 = 0.228$. Indeed, individual correlation coefficients between each predictor and the dependent variable indicated that exposure to subordinate conjunctions correlated negatively with output gains, $r(23) = -0.37, p = 0.034$, and failure rate correlated positively with output gains, $r(23) = 0.34, p = 0.049$. In other words, students benefiting from the output treatment were exposed to less complex syntax in the contextualization phase; they also had a higher failure rate during the practice phase. Interpreting the subordinate conjunctions predictor is not straightforward. This predictor indicates that more successful learners in the output treatment processed less syntactic complexity in the contextualization phase. It is not unreasonable to conclude, thus, that successful learners were exposed to less complicated language in the 3D contextualization phase. This variable may indicate that readily accessible

contextualization (e.g., not syntactically complex) better prepares learners to benefit from the practice phase. To interpret the failure rate predictor, recall that there was a qualitative difference in how the input and the output group's failure rates were calculated. The input group's failure rate reflected only incorrect answers. The output group's failure rate reflected both entirely incorrect and *close* answers. A *close* answer in the output treatment resulted when some aspect of the verb other than the mood inflection (i.e., the stem or the person-number inflection) was deemed incorrect by shallow morphological parsing. Being informed that the response was 'close', the learner knew that some part of the verb was correct and something required repair. Thus, the learner likely employed metalinguistic knowledge about the morphology of the verb to correct the response, and such deliberate processing of the verbs during practice probably resulted in greater gains. To support this conjecture, we correlated only the number of 'close' counts and the output group's overall gains (using the same dependent variable as in the regression analysis). The analysis showed that 'close' failure instances significantly predicted overall gains, $r(23) = 0.43, p = 0.034$.

Discussion and conclusions

Regarding the first research question, the present conceptual replication corroborates Collentine's (1998a) original finding that input- and output-oriented instruction where psycholinguistic processing principles inform instructional design affect the acquisition of complex grammatical phenomena like the Spanish subjunctive. Indeed, the generalizability of this insight beyond the classroom to CALL would not have been possible or defensible without the sort of replication reported here. The only differences between the original and this CALL replication study were attributable to changes in the non-major factors (e.g., the present learners brought more language study to the experiment) rather than to the transfer of the learning situation from the classroom to a CALL setting. As in most SLA research, the results indicate that prior knowledge of a construct plays an important role in the efficacy of instructional interventions. Complex structures such as the subjunctive in adjectival clauses develop out of various declarative knowledge stores. Consideration of the effect of the morphological covariate in the ANOVA analyses demonstrates that prior knowledge of the inflectional morphology of a morphosyntactic structure influences learning, a factor that has been largely ignored to date in subjunctive intervention studies (Collentine, 2010, 2013).

This conceptual replication is also important to the extent that it provides guidance to CALL materials designers on how to promote SLA based grammar development with emerging technologies. Concerning the second research question, the study shows that CALL software's ability to embed script-based

tracking technologies to record aspects of the learning process provide important insights into instructional design and SLA processes that impact learner outcomes (i.e., product). One implication stemming from the tracking analysis of the output-oriented treatment is that contextualization (i.e., providing a situational setting for a practice activity) may need to include language that is simple rather than complex. If practice places demands on learners' limited attentional resources and if the contextualization of that practice is also cognitively demanding, the task becomes complex overall and learners may not have enough remaining resources to benefit from such practice (Skehan, 2009). Interestingly, the tracking analysis showed no relationship between what the learners receiving PI did during their treatment and their gains. Perhaps a fine-grained analysis would reveal a more informative picture of the relationship between PI processes and products. Of course, the results may be accounted for by the fact that researchers such as VanPatten contend that it is the whole PI methodology that affects acquisition (VanPatten, 2004). Nonetheless, the tracking analysis also indicates that feedback promoting repair and metalinguistic reflection are important affective agents in output-oriented instruction. This is an especially important finding for tutorial CALL, since it suggests that designers should explore further the potential benefits of feedback routines built on shallow parsing technologies (e.g., regular expressions pattern matching and JavaScript routines).

In terms of the larger question of the relative importance of input and output in L2 instruction, this conceptual replication and the original study indicate that what most accounts for grammatical gains is not whether instruction is input- or output-oriented. Rather, instruction should require that learners focus on the meaning of the target structure (e.g., output instruction does not involve mechanical exercises) and practice should be deliberate rather than incidental (DeKeyser *et al.*, 2002; Toth, 2006; VanPatten, 2002). There is not a single way to accomplish this goal. In the input-oriented PI treatment it is important to emphasize that, not only did the learners receive information on how to process the target structure when reading or listening, they processed a particular type of input, namely, structured input (along with affective activities). Structured input is not authentic input. One of its characteristics is that it makes otherwise unnoticeable grammatical features more meaningful. In the output-oriented treatment, the learners also received processing strategies and they were prompted to produce verbs whose mood inflection was critical to communicating a message. It turns out that the present CALL conceptual replication revealed insights into the learning process in output-oriented instruction that is difficult to see in classroom research. The feedback on the learners' errors that were 'close' to correct may well have fed salient data back to the developing system by requiring learners to conduct a 'sophisticated analysis' of the structure's component parts (Toth, 2006: 328).

There are limitations to the generalizability of this study. Both input and output were narrowly operationalized. PI has many components and requirements, and it would be difficult for day-to-day practitioners or content designers to faithfully provide all the necessary requirements without a broad understanding of SLA theory (cf., Farley, 2001; VanPatten, 2002). Nevertheless, its utility seems indisputable given the range of structures with which its efficacy has been shown (VanPatten, 2002, 2012). Additionally, future research would do well to explore the benefits of more robust parsing technologies and feedback on learners' production, as well as the effects of synchronous computer mediated communication (SCMC) on grammar learning since humans can provide feedback that is flexibly responsive (Schulze, 2008). While the study employed a covariate to factor in the effects of previous subjunctive morphological knowledge, there was no screening of learners based on their entire Spanish academic career, such that it is unclear what were the entire effects of previous knowledge on the results. A final consideration is that the present study examined learners from only a single level of instruction and it included no delayed posttest. A more comprehensive picture of the relative benefits of input and output would emerge from a study of multiple instructional levels and throughout a longer timeframe.

While Chun (2012) expresses concern about the brief shelf life of CALL technologies, which would diminish the utility of CALL replication studies, we contend that the technologies employed in this study (i.e., 3D, web apps, shallow parsing) will not be replaced in the short or long term. HTML5 and CSS3 are relatively new technologies and developers are increasingly looking to web apps for content delivery and to provide interactivity. We nevertheless acknowledge that it is reasonable to question whether 3D worlds will maintain or gain in popularity. It is well known that sites such as Second Life have diminished in popularity, and mobile devices better support 2D game environments. In any event, it seems that there is reason to be optimistic of either input- or output-oriented tutorial CALL going forward.

Notes

1. While Farley (2001) contends that Collentine (1998a) did not consider processing principles in the PI treatment, Collentine (2002) clarifies how Collentine (1998a) designed both the PI and the output treatments to elevate the subjunctive's communicative value, in line with the *primacy of content words principle* and the *lexical preference principle*.

2. See Collentine (1998a) for a complete description of the treatment and instructional materials. Furthermore, Collentine's (2002) response to Farley (2001) contains additional descriptions of the Collentine (1998a) materials.

3. While many Spanish grammar books encourage learners to cue in on an antecedent's article to determine whether it is variable or definite (e.g., *Busco el/un* [definite/indefinite] *restaurante que es/sea* [indicative/subjunctive] *nuevo* 'I am looking for the/a restaurant that is new'),

it is an unreliable indicator. Collettine (2002) shows that in native speaker speech definite and indefinite articles are unreliable predictors of the mood of an adjectival clause (*Busco un restaurante que es/sea nuevo*). In any event, practice in both the original and the present study only used antecedents with indefinite articles regardless of whether the dependent clause of an exemplar contained indicative or subjunctive (i.e., in input practice) or whether a learner was to produce the indicative for the subjunctive (i.e., in output practice).

About the authors

Dr Joseph Collettine is Professor of Spanish at Northern Arizona University. He researches computer assisted language learning, corpus linguistics, and the acquisition of the subjunctive.

Dr Karina Collettine is Associate Professor of Spanish and Spanish Secondary Education at Northern Arizona University. Her current research interests include foreign language learning in 3D worlds and task-based computer-mediated communication.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716–723. <http://dx.doi.org/10.1109/TAC.1974.1100705>
- Beatty, K. (2012). *Teaching and researching: Computer-assisted language learning*. New York: Pearson Education.
- Cerezo Ceballos, L. (2010). *Talking to avatars: The computer as a tutor and the incidence of learner's agency, feedback, and grammatical form in SLA* (Unpublished doctoral dissertation). Georgetown University. Washington, DC.
- Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524681>
- Chun, D. (2012). Replication studies in CALL research. *CALICO Journal*, 29 (4), 591–600. <http://dx.doi.org/10.11139/cj.29.4.591-600>
- Collettine, J. (1998a). Processing instruction and the subjunctive. *Hispania*, 81 (3), 576–587. <http://dx.doi.org/10.2307/345673>
- Collettine, J. (1998b). Cognitive principles and CALL grammar instruction: A mind-centered, input approach. *CALICO Journal*, 15 (1), 1–18.
- Collettine, J. (2002). On the acquisition of the subjunctive and authentic processing instruction: A response to Farley. *Hispania*, 85 (4), 879–888. <http://dx.doi.org/10.2307/4141255>
- Collettine, J. (2010). The acquisition and teaching of the Spanish subjunctive: An update on current findings. *Hispania*, 93 (1), 39–51.
- Collettine, J. (2013). Subjunctive in second language Spanish. In Kimberly Geeslin (Ed.), *The handbook of Spanish second language acquisition*, 270–286. New York: Wiley and Sons. <http://dx.doi.org/10.1002/9781118584347.ch16>

- Collentine, K. (2013). Using tracking technologies to study the effects of linguistic complexity in CALL input and SCMC output. In P. Hubbard, M. Schulze, & B. Smith (Eds.), *Learner-computer interaction in language education: A Festschrift in honor of Robert Fischer*, 46–65. San Marcos, TX: CALICO.
- DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction*, 97–113. New York: Routledge.
- DeKeyser, R. & Prieto Botana, G. (2013) The acquisition of grammar by instructed learners. In K. Geeslin (Ed.), *The handbook of Spanish second language acquisition*, 449–465. New York: Wiley and Sons.
- DeKeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's 'Processing Instruction: An update'. *Language Learning*, 52 (4), 805–823. <http://dx.doi.org/10.1111/1467-9922.00204>
- Doughty, C. & Long, H. M. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Language Learning & Technology*, 7 (3), 50–80.
- Ellis, N. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24 (2), 143–188.
- Farley, A. P. (2001). Authentic processing instruction and the Spanish subjunctive. *Hispania*, 84 (2), 289–299. <http://dx.doi.org/10.2307/3657760>
- Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *Computer Assisted Language Learning*, 20 (5), 409–442. <http://dx.doi.org/10.1080/09588220701746013>
- Gass, S. (2010). The relationship between L2 input and L2 output. In E. Macaro (Ed.) *Continuum companion to second language acquisition*, 194–219. London: Continuum International Publishing.
- Gass, S., Behney, J., & Plonsky, L. (2013). *Second language acquisition: An introductory course* (4th ed.). New York: Routledge.
- Godwin-Jones, R. (2014). Towards transparent computing: Content authoring using open standards. *Language Learning & Technology*, 18 (1), 1–10.
- Hubbard, P. & Siskin, C. B. (2004). Another look at tutorial CALL. *ReCALL*, 16 (2), 448–461. <http://dx.doi.org/10.1017/S0958344004001326>
- Kim, S., Lee, J., & Thomas, M. (2012). Between purpose and method: A review of educational research on 3D virtual worlds. *Journal of Virtual Worlds Research*, 5 (1), 1–18.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Larson-Hall, J. (2009). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
- Leow, R. (2007). Input in the L2 classroom: An attentional perspective on receptive practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*, 21–50. New York: Cambridge University Press.

- Lyster, R. & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32 (2), 265–302. <http://dx.doi.org/10.1017/S0272263109990520>
- Mackey, A. (2012). Why (or why not), when and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics*, 21–46. Cambridge: Cambridge University Press.
- Porte, G. (2013). Who needs replication? *CALICO Journal*, 30 (1), 10–15. <http://dx.doi.org/10.11139/cj.30.1.10-15>
- Ranta, L. & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The awareness-practice-feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*, 141–160. Cambridge: Cambridge University Press.
- Russell, V. (2012). Learning complex grammar in the virtual classroom: A comparison of processing instruction, structured input, computerized visual input enhancement, and traditional instruction. *Foreign Language Annals*, 45 (1), 42–71.
- Russell, V. (2014). A closer look at the output hypothesis: The effect of pushed output on noticing and inductive learning of the Spanish future tense. *Foreign Language Annals*, 47 (1), 25–47. <http://dx.doi.org/10.1111/flan.12077>
- Schulze, M. (2008). AI in CALL: Artificially inflated or almost imminent? *CALICO Journal*, 25 (3), 510–527.
- Skehan P. (2009). Modeling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, 30 (4), 510–532. <http://dx.doi.org/10.1093/applin/amp047>
- Smith, B. & Schulze, M. (2013). Thirty years of the CALICO Journal – replicate, replicate, replicate. *CALICO Journal*, 30 (1), i–iv. <http://dx.doi.org/10.11139/cj.30.1.i-iv>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds), *Input in second language acquisition*, 235–253. Rowley, MA: Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds), *Principles and practice in the study of language*, 125–144. Oxford: Oxford University Press.
- Toth, P.D. (2006). Processing instruction and a role for output in second language acquisition. *Language Learning*, 56 (2), 319–385. <http://dx.doi.org/10.1111/j.0023-8333.2006.00349.x>
- VanPatten, B. (1993). Grammar teaching for the acquisition rich classroom. *Foreign Language Annals*, 26 (4), 435–450. <http://dx.doi.org/10.1111/j.1944-9720.1993.tb01179.x>
- VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, 52 (4), 755–803. <http://dx.doi.org/10.1111/1467-9922.00203>
- VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary*, 5–31. Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B. (2012). Input processing. In S. Gass & A. Mackey (Eds) *The Routledge handbook of second language acquisition*, 440–464. New York: Routledge.

VanPatten, B. & Benati, A. (2010). *Key terms in second language acquisition*. London: Continuum.