Applied Psycholinguistics **34** (2013), 1163–1193
doi:10.1017/S0142716412000197

The impact of instruction on second-language implicit knowledge: Evidence against encapsulation

PAUL D. TOTH
Temple University

PEDRO GUIJARRO-FUENTES
University of Plymouth

Received: June 24, 2010 Accepted for publication: October 28, 2011

ADDRESS FOR CORRESPONDENCE
Paul D. Toth, Department of Spanish & Portuguese, Temple University, 441 Anderson Hall, 1114 West Berks Street, Philadelphia, PA 19122-6090. E-mail: ptoth@temple.edu

ABSTRACT
This paper compares explicit instruction in second-language Spanish with a control treatment on a written picture description task and a timed auditory grammaticality judgment task. Participants came from two intact, third-year US high school classes, with one experiencing a week of communicative lessons on the Spanish clitic *se* ($n = 15$) and the other exposed to *se* only incidentally ($n = 20$). Explicit instruction consisted of grammar rules with sentence-level examples, followed by communicative tasks. Three test versions were administered within a split-bloc design as a pretest, immediate posttest, and delayed posttest 6 weeks after instruction. The instructed group increased targetlike uses of *se* on both tasks and sustained gains through the delayed posttest, although first-language transfer errors persisted. Meanwhile, overgeneralization errors centered on semantic and syntactic contexts similar to the instructional object, aligning with the unergative–unaccusative distinction among intransitive verbs. It is argued that the data provide evidence for the permeability of second-language implicit knowledge to explicit instruction and against total encapsulation as a model of the mind.

Although research since the 1980s has documented positive effects for instruction on second-language (L2) acquisition (deGraff & Housen, 2009; Ellis, 2002), many scholars remain skeptical. In some studies, positive results have failed to match the predictions of linguistic theory, whereas in others they have simply been inconclusive or unsupportive. Early research into the gap between L2 instruction and acquisition found similar developmental sequences for accuracy with L2 English grammatical morphemes despite considerable differences in learners' instructional experiences (Bailey, Madden, & Krashen, 1974; Krashen, Sferlazza, Feldman, & Fathman, 1976). These results coincided nicely with the theoretical positions of Chomsky (1965) and Fodor (1983), wherein linguistic competence, or the implicitly known, abstract system underlying communication, is "informationally

© Cambridge University Press 2012 0142-7164/12 $15.00

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

encapsulated" from other cognitive "modules," including the conscious reasoning that derives explicit grammatical knowledge. Clearly, these two mental capacities, which shall be called "implicit" and "explicit L2 knowledge" here, are not the same; and the disconnect between explicitly knowing *about* L2 structure and employing an implicit communicative system is one that any L2 learner or instructor can appreciate. Based on this common experience and Krashen and colleagues' (1976) findings, Krashen (1982) packaged the encapsulation argument within an instructional program that eschewed explicit "learning" and held exposure to L2 "comprehensible input" as the only necessary variable for "acquiring" implicit L2 knowledge.

Since then, concern over the role of instruction in acquisition has inspired much research on links between explicit and implicit L2 knowledge. Although common developmental sequences were repeatedly found in studies of L2 development, explicitly instructed learners advanced more quickly through the stages and reached higher levels of ultimate attainment than naturalistic learners (Doughty, 1991; Pica, 1983; Pienemann, 1984). Explicitly negotiating errors with learners also appeared to make L2 input more comprehensible, thus creating conditions that favored implicit knowledge development (Long, 1983; Pica, 1987). However, although L2 research reported positive, if indirect, impacts for explicit instruction (Norris & Ortega, 2000), critics also noted that common assessments such as open-ended judgment and sentence completion tasks were problematic because they allowed learners to utilize explicit knowledge during test taking (Doughty, 2003; Norris & Ortega, 2003; Truscott, 1998). Consequently, recent studies of L2 instruction have attempted to provide more valid assessments of implicit knowledge via tasks without extended opportunities for planning or reflection.

The present study investigates learning outcomes for L2 Spanish learners at an American high school who received 270 min of explicit instruction on the pronominal clitic *se* over one week. They were compared to a similar control group who received only communicative Spanish input during the same period. Following recent suggestions for L2 assessment (Doughty, 2003; Ellis, 2009), testing included a written picture description task and a timed, picture-based, auditory grammaticality judgment (GJ) task. These assessments were believed to validly assess implicit L2 knowledge, given the aural mode and time pressure of the judgment task, and evidence on both tasks that linguistic properties too subtle for explicit monitoring shaped postinstruction performance. The results showed that instructed learners' improvement with *se* came with overgeneralization reflecting the unergative–unaccusative distinction among intransitive verbs. We conclude that although implicit and explicit L2 knowledge are in many ways autonomous, they cannot be utterly impervious to each other. Evidence that growth in implicit knowledge coincided with the instructional development of explicit knowledge requires the revision of complete encapsulation as a valid model of the mind.

THE IMPLICIT–EXPLICIT (DIS)CONNECTION AND THE ROLE OF INSTRUCTION

For Fodor (1983), the "informational encapsulation" of implicit linguistic knowledge meant that language processing occurred "autonomously," with no sharing

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

of "resources" from other modules such as memory or attention (p. 37). At times, an a priori commitment to this model has led generative L2 theorists to circular, unfalsifiable interpretations of learner data that disregard alternative explanations. For example, in seminal work by White (1991), Francophone middle school L2 English learners improved markedly on adverb placement 5 weeks after explicit instruction by accepting correct English word orders and rejecting French-like, incorrect ones. However, many also overgeneralized their rejection to superficially similar sequences that were in fact acceptable in English. A year later, with no further instruction on adverbs and the English program shifting from all-day immersion to just 90 min per week, many learners also reverted to accepting an ungrammatical, French-like word order, although performance with correct English sequences remained higher than on the pretest. Nonetheless, L2 applications of Chomsky's (1981) principles and parameters theory assumed that structural contrasts between the first language (L1) and the L2 were binary switches that, once activated, brought a full shift in implicit L2 knowledge that should preempt L1 transfer and rein in overgeneralization. Because White's results failed to show this, Schwartz and Gubala-Ryzak (1992; Schwartz, 1993) argued that instruction-derived, explicit L2 knowledge could only have affected learners' "learned linguistic behavior," leaving their implicit knowledge unchanged. Although White (1992) later noted that Chomskyan theory need not preclude more than one active parameter within individuals, she agreed that overgeneralization and diminishing instructional effects could only mean that implicit L2 knowledge was unaffected. Hence, the results posed no challenge to encapsulation, despite the fact that the same assumptions under investigation were also used to discount unexpected data.

Since then, encapsulation proponents have continued to interpret disconnects between explicit and implicit L2 knowledge as meaning that no connection exists at all. For example, evidence of instructed learners coming to correctly acquire subtle L2 properties despite having received inadequate explicit rules is often taken to mean not just that universal grammar (UG) constrains implicit L2 knowledge but also that explicit knowledge is irrelevant to its development (Belikova, 2008; Bruhn-Garavito, 1995; White, 2003). Meanwhile, after finding that instructed L2 Spanish learners made errors reflecting overgeneralized explicit rules whereas naturalistic learners were more targetlike, Rothman (2008), citing Schwartz & Gubala-Ryzak (1992), argued that the instructed learners must not have accessed true implicit L2 knowledge on the tests, but rather allowed the "competing system" of explicit knowledge to interfere with their performance (p. 99). Thus, it has been argued that explicit knowledge is both utterly inert *and* an impediment to implicit L2 development.

Although Rothman's (2008) untimed tasks may well have allowed learners to reflect on imprecise, explicit "rules of thumb," neither their nontargetlike performance nor the stronger results for uninstructed learners provide compelling evidence for encapsulation. Uninstructed learners could have used self-taught explicit rules during the tests, and instructed learners might have performed better if they had received more adequate explicit information. It appears that what most strongly links variable learner behavior to the conclusion that implicit and explicit L2 knowledge are utterly impervious is a commitment to Chomskyan–Fodoran orthodoxy (cf. Rothman, 2008; Schwartz, 1999; White, 2003). The possibilities

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

that implicit L2 knowledge could be truly variable, or that explicit knowledge might facilitate implicit development, are generally rejected even though they need not invalidate UG-based analyses.

It is interesting that within Chomsky's (1995, 2007) minimalist program, many generative L2 researchers have begun considering how implicit syntactic knowledge may "interface" with nonsyntactic knowledge, including the conceptual and pragmatic knowledge that underlies lexical selection and such discursive phenomena as topicalization and pronoun use (e.g., Rothman & Slabakova, 2011). However, whereas some believe that the real-time processing of such interfaces could explain L2 variability (Sorace, 2011), none have explored how explicit L2 knowledge and discursive monitoring might affect implicit knowledge through these same interfaces. As generativists, Gregg (2001) and White (2007) concede that UG works best as a "property" theory of L2 knowledge rather than as a "transition" theory of acquisition. Still, many continue to draw encapsulation-based implications for L2 pedagogy, where the teacher's essential role is simply providing L2 input. Rothman (2010), for example, explains that explicit L2 knowledge benefits learners mostly by "trickling down" from nonnative teachers who may use it to improve their monitored L2 "performance" and thus provide more accurate L2 input (p. 57).

Although nongenerative, cognitive L2 theories generally offer more cogent explanations for how explicit instruction might benefit implicit knowledge, and why overgeneralization and L1 transfer might create targetlike changes, most of these lack a theory of language that adequately embraces the range of phenomena covered in generative theory. Connectionist and emergentist approaches are more tolerant of variability in mapping form onto meaning, as they attribute long-term L2 development to frequency of use and an ever more fine-grained "tuning" of such mappings that gradually eliminates L1 transfer and overgeneralization (Ellis, 2005; Hopper, 1998; Tomasello, 2003). However, their explanation for the grammatical meanings that learners link to forms is imprecise (Chater & Manning, 2006; Dörnyei, 2009), as all connections derive from environmental stimuli interacting with cognitively generic, inductive processes with no linguistic specialization. Hence, how grammatical features such as pronominal and interrogative reference, definiteness, case, aspect, or causativity could be made available for association with auditorily perceived forms is unclear (Jackendoff, 2002).

Still, cognitive L2 approaches have extensively considered how explicit and implicit knowledge may work in tandem while nonetheless representing two distinct knowledge bases (DeKeyser, 2003; Paradis, 2009; Ullman, 2005). Some propose that knowledge *about* the L2 facilitates knowledge *of* the L2 by helping learners "notice" form–meaning relationships that would otherwise go unprocessed (e.g., Schmidt, 2001; Talmy, 2008; VanPatten, 1996). Without an insistence on encapsulation, "monitoring" L2 production via explicit knowledge is also widely believed to facilitate form–meaning alignments, as explicit knowledge that may be more sophisticated than the extant implicit system bootstraps targetlike forms into place during utterance planning and assembly (e.g., de Bot, 1996; Izumi, 2003; Kormos, 2006). If this occurs at developmentally opportune times, when the implicit L2 system can absorb the linguistic entailments of the language being "pushed" by explicit knowledge, this may be precisely the activity that moves implicit knowledge

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

forward. Hence, if classroom data can demonstrate distinct contributions from explicit and implicit L2 knowledge jointly underlying postinstructional changes in linguistic behavior, it would beg the question of how mutual influence could *not* be implicated as the most logical explanation.

IMPLICIT KNOWLEDGE OF ARGUMENT STRUCTURE AND SPANISH *se*

To characterize implicit linguistic knowledge, generative theory distinguishes two classes of intransitive verbs with differing semantic and syntactic properties (Burzio, 1986; Perlmutter, 1978). *Unergative* verbs denote controlled actions or processes, such as *nadar* (swim), *llorar* (cry), and *jugar* (play), where a semantic agent (i.e., a noun phrase [NP]) initiates an event as in (1), and becomes the syntactic subject as an "external argument" of the verb phrase (VP). Meanwhile, *unaccusatives* denote changes of state or location, such as *nacer* (be born), *aparecer* (appear), and *llegar* (arrive), where a semantic *non*agent undergoes an event as in (2), becoming the syntactic subject as an "internal argument" of VP, which is also the point of origin for direct objects of transitive verbs. Thus, the semantic distinction between causing and undergoing an event is reflected syntactically in differing paths to subject position.

(1) Unergatives: $[_S\ [_{NP}$ *Julia*$]\ [_{VP}$ *nadó*$]\ ]$ — *Julia* = AGENT, external argument
Julia swam.

(2) Unaccusatives: $[_S\ [_{NP}$ *Julia*$]\ [_{VP}$ *nació* $_]\ ]$ — *Julia* = PATIENT, internal argument
Julia was born.

Cross-linguistic evidence supports this analysis, as unaccusative subjects exhibit some of the same collocational and combinatorial properties of direct objects in many languages, whereas unergative subjects do not (Alexiadou, Anagnostopoulou, & Everaert, 2004; Levin & Rappaport-Hovav, 2005). Still, in both English and Spanish, surface differences between the two verb classes are subtle, as both frequently appear in subject–verb sequences that would be difficult for a nonlinguist to differentiate.

Following Pustejovsky (1995) and Levin and Rappaport-Hovav (2005), two classes of transitive verbs can also be distinguished based on whether they allow a spontaneous, inchoative reading when agents are omitted. The first, called "alternators" here, depict agentive events when transitive as in (3a), but allow spontaneous, inchoative readings when an agent is omitted, as in (3b). Note that (3b) is similar to the unaccusative in (2), as the subject is nonagentive and originates as an internal argument. However, unlike (2), all dialects of Spanish require the clitic *se* in (3b) to derive intransitivity, thus demonstrating one of its essential functions as an argument "absorber" (Koontz-Garboden, 2009; Reinhart & Siloni, 2005). Because absorbing the agent is primarily a syntactic operation, the full semantic impact of *se* is determined by the properties of the elements that remain. Thus, although alternators allow an inchoative reading in (3b), a missing agent could also convey a passive or impersonal reading, particularly if the word order

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

is verb–subject, as in *Se secaron los platos*, "The dishes were dried" or "One dried the dishes."

(3) a. [S [NP *Pedro*] [VP *secó* [NP *los platos*]]] *Pedro* = AGENT;
Pedro dried the dishes *los platos* = PATIENT

b. [S [NP *Los platos*] [VP ***se*** *secaron* —]]. *los platos* = PATIENT
The dishes Ø dried.

Meanwhile, in the second verb class, "accusatives," external causation is always implied, whether the agent is given as in (4a) or omitted as in (4b). As such, the still tacit agent in (4b) excludes the possibility of an inchoative reading (Pustejovsky, 1995), and can only be interpreted with passive or impersonal meaning. Like alternators, however, *se* is required to derive single-argument syntax, and in most dialects, an inanimate patient will agree with the verb as its subject but usually stay to the right, as shown in (4b).

(4) a. [S [NP *Pedro*] [VP *lavó* [NP *los platos*]]] *Pedro* = AGENT;
Pedro washed the dishes *los platos* = PATIENT

b. [S — [VP ***Se*** *lavaron* [NP *los platos*]]]. *los platos* = PATIENT
Were washed the dishes
"The dishes were washed" or "One washed the dishes."

Thus, although both inchoative alternators (3b) and unaccusatives (2) lack specified agents to initiate the event, passive accusatives (4b) and inchoative alternators both require *se* to derive intransitivity. With inherently intransitive unergatives (1) and unaccusatives (2), however, *se* is not required because there are no missing arguments to be absorbed.[1] Thus, targetlike implicit knowledge of *se* requires a sensitivity to verbal argument structure that employs the clitic only for missing arguments, and not for all nonagentive subjects.

The learning challenge for English-speaking learners is further complicated by *se*'s various L1 equivalents. In (3b) and (4b), where *se* absorbs an agent, English uses a zero-derived inchoative and passive copula, respectively. As (5a) and (5b) demonstrate, however, *se* may also absorb patient arguments with the same verbs to yield reflexives or reciprocals in English, depending on whether the agent is singular as in (5a) or plural as in (5b).

(5) a. *Pedro* ***se*** *lavó*. "Pedro washed **himself**."
b. *Los niños* ***se*** *secaron*. "The children dried **themselves / each other**."

For English-speaking learners, then, the similar structures "NP *se* V" and "*se* V NP" appear to have inconsistent meanings corresponding to numerous L1 forms. The systematicity of *se*'s link to omitted arguments, compared with the miscellany of free-morphemes in English, has led Reinhart and Siloni (2005) to propose a cross-linguistic parameter for "reflexivization" where argument reduction occurs syntactically in Spanish but lexically in English. Because *se*'s use belies subtle and complex distinctions beyond the realm of traditional pedagogical rules, the

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

question at hand is whether instruction aimed at raising awareness of these issues can facilitate an implicit knowledge of *se* that both counteracts L1 transfer and generalizes beyond the tokens in the input. If so, then implicit L2 development may be shaped not only by subtle principles of UG but also by the explicit information and guided interactions of instruction.

THE PRESENT STUDY

Research Questions

The following research questions thus guided our inquiry in this study:

1. Will explicit instruction on *se* improve its use with alternators and accusatives on an auditory, timed GJ task and a written picture description task?
2. Will instruction lead to overgeneralization errors with both unergatives and unaccusatives, or just the nonagentive unaccusatives?
3. Will instruction preempt L1 transfer errors that omit *se* with intransitive alternators?

Method

Participants. Our data came from 2 intact classes of third-year L2 Spanish in a US suburban public high school. In one class, 19 learners experienced communicative, explicit instruction on nonagentive *se* over three 90-min lessons. In the second, 24 control-group learners received three 90-min lessons during the same week, where they practiced writing to pen pals with no explicit grammatical focus. Participants' ages ranged from 15 to 17 years, with 11 females and 8 males in the instructed group and 15 females and 9 males among the controls.

All participants completed a questionnaire to indicate their native language, place of birth, extent of current and prior exposure to Spanish, and any previous knowledge of rules for *se*. Individuals were later excluded from the analyses if they (a) had ever lived in a Spanish-speaking country, (b) spoke Spanish at home, (c) did not speak English as their L1, (d) indicated prior knowledge of nonagentive *se*, (e) scored higher than 50% correct use of nonagentive *se* on the picture description pretest, or (f) were absent from two or more lessons or any test administration. As a result, 4 learners' data in each class were excluded, such that there were 15 in the instructed group and 20 in the control group. Given these relatively small numbers, an a priori statistical power analysis confirmed that, with a significance level of $\alpha = 0.05$ and a minimum power $(1 - \beta)$ of 0.80, a moderate effect size ($\eta_p^2 = 0.04$) would be detectable in the factor interactions of the repeated-measures analyses of variance (ANOVAs) to be used.[2]

Participants also included 31 adult native Spanish speakers from all parts of Latin America, each taking the tests once individually to compare their performance with the learners. These were all immigrants to the United States who had spent at least their first 18 years in a Spanish-speaking country, such that they could all be said to have fully acquired Spanish as an L1. Fourteen were women and 17 were men, with ages ranging from 19 to 51 years.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Finally, the instructor in both groups was a 23-year-old female nonnative Spanish speaker completing her third student-teaching internship semester. Data collection took place 8 weeks after she started working full-time in these classes. Beforehand, her Spanish had been assessed as "advanced–low'" on the proficiency scale of the American Council for the Teaching of Foreign Languages (1999), indicating that she could produce paragraph-level spoken discourse without systemic errors. All treatment classes were videotaped, and a review of the recordings showed her use of Spanish *se* to be nativelike. The learners' regular teacher was also present in the classroom at all times, but she generally remained silent and did not participate in any aspect of instruction.

Instructional treatments. The two Spanish classes met every other day for 90 min, the control group from 12:10 to 1:40 p.m. and the instructed group from 1:50 to 3:20 p.m. A pretest and immediate posttest were given before and directly following the three-lesson treatment in each group. A delayed posttest was then administered 6 weeks later to assess longer-term treatment effects. The instructor confirmed that after the three lessons, *se* occurred only incidentally in both groups' classroom discourse and was never the explicit focus of instruction.

The instructed group's lessons were designed by the first author in collaboration with the instructor and a research assistant. The instructor reviewed and rehearsed final drafts of the lessons during two 2-hr meetings in the weeks prior to data gathering. The week immediately before, the first author and research assistant each visited the school to observe the instructor's teaching style and offer suggestions for the upcoming treatments. Each lesson in the instructed group was built around culturally authentic L2 texts that exemplified particular features of nonagentive *se*. In the first lesson, "impersonal *se*" was presented within a critic's review of two Costa Rican restaurants. In the second, "passive *se*" was exemplified in a recipe for Mexican stuffed chilies. In the third, "spontaneous *se*" (i.e., inchoative *se*) was demonstrated through a newspaper report of a Chilean earthquake. All three lessons followed a similar sequence of communicative tasks that moved from comprehension to production, with an explicit grammar explanation in the middle, as in Table 1.

In the first lesson, explicit rules contrasted impersonal *se* in generic events with transitive sentences omitting *se* and referencing specific "doers." The second lesson further highlighted passive *se*'s association with missing agents and outlined verbal agreement with undergoer arguments that would normally be direct objects in sentences without *se*. Finally, in the third lesson, inchoative or "spontaneous" *se* was presented as a way to exclude causation from events completely. Again, sentences with causative agents were contrasted with those where *se* represented an inchoative change of state. In all three lessons, sentence-level examples and English-translation equivalents accompanied the explanations.[3]

As indicated in Table 1, each lesson began with comprehension tasks where learners listened to a recording of the primary text while reading along on a handout, and then identified its essential meaning and supporting details. A grammar explanation followed, with the instructor having learners find examples of *se* in the text. Then, based on visual prompts and key words, the instructor guided the class

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Table 1. *Task sequence used in all three 90-min lessons of the instructed group's treatment*

Time	Task	Goal
5 min	1. Warm-up/prereading	Orient learners to topical content of text
15 min	2. Reading/listening comprehension	Processing text content; exposure to *se*
5–10 min	3. Explicit grammar presentation	Provide a rule for *se* with examples
5–10 min	4. Grammatical analysis	Identify examples of *se* in the text
20 min	5. Guided speaking (whole class with feedback)	Teacher-guided communicative practice using *se* as in the rule and the text
25 min	6. Information exchange (small groups, whole-class follow-up)	Independent communicative practice using *se* as in the rule and the text

in performing a language function with *se* similar to that of the text, including a restaurant review on Day 1, a recipe on Day 2, and an account of an accident on Day 3. Finally, learners wrote their own versions of each language function before working in groups of two to three on information exchanges, where they listened to each other's compositions and made decisions, such as identifying the restaurant being reviewed, naming the dish in the recipe, or stating whether they believed the accident to be true. Learners were encouraged to help each other with any difficulties, and whole-class follow-ups were used to resolve remaining questions.

Meanwhile, the control group was designed to assess ordinary "maturation" for *se* in the classroom without organized intervention, as Mackey and Gass (2005, p. 115) suggest. As such, control lessons were less structured than the instructed group and involved no explicit focus on *se*. Learners brought in one or two gifts from home to send in a "culture box" to L2 English learners in Bolivia with whom they had already established contact as pen pals. They spent the three lessons working in groups of three to four to prepare their gifts, decorate a box, and write short descriptions in Spanish of each gift and its significance. The instructor provided grammar or vocabulary help upon request. Before enclosing their gifts in the box, each group presented their work to the class in Spanish. Throughout this time, although teacher input may have included *se* incidentally, and some individuals may have needed help with expressions requiring *se,* at no time was this or any other grammatical structure the overt focus of instruction.

Data gathering and assessment. The assessment tasks included (a) a written, sentence-level picture description task, where learners used cued verbs and nouns to describe corresponding pictures within an overall time limit for the task; and (b) an auditory picture-based GJ task, where learners rated how well prerecorded sentences represented corresponding pictures within fixed time intervals for each

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press



Figure 1. Two items from the picture description task ("to cook," "the fish" and "to be born," "the baby").

item. Based on current literature (Ellis, 2009; Norris & Ortega, 2003), it was assumed that the timed GJ task would operationalize implicit knowledge of *se*, whereas the globally timed picture description task would require both implicit and explicit knowledge to some degree.[4]

For the picture description task, learners wrote single sentences for 15 pictures that each appeared with a given verb and NP in their test booklet, as in Figure 1. Approximately 8 min was given for the entire task. Each of the four verb classes in Examples (1) through (4) was represented by three test items to assess grammatical uses of *se*, overgeneralization, and L1 transfer. Items were randomized; and, except for unergatives, the pictures and NPs depicted nonagents.

For the picture-based GJ task, the test books contained 88 numbered pictures, and for each one, participants heard a recorded sentence and rated its acceptability as a depiction of the picture. A 4-s gap was used between sentences, given that pilot testing of 2- through 8-s intervals with another third-year L2 Spanish class ($n = 26$) had learners rating the task "challenging" but not "overwhelming" or "easy" at 4 s. Following Juffs's (2001) GJ design suggestions, participants responded on an answer sheet using a 6-point Likert scale, as shown in Figure 2, with 6 indicating *total acceptability*, 1 indicating *total unacceptability*, and intermediate values in between.[5] Participants were directed to respond intuitively about the extent to which the sentence could be used to describe the picture and to select "Don't Know" if they were unsure. Prerecorded instructions took participants through three examples before starting. Then, randomly ordered items offered grammatical and ungrammatical uses of *se* with the verb classes in (1)–(4) to assess targetlike use, overgeneralization, and L1 transfer. The pictures for

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Recorded sentences:

"La ventana se cerró."

"*Las flores se crecen."*

Visuals appearing in the test book:

1. 

2. 

Corresponding items on student answer sheet:

CIRCLE YOUR ANSWER:

How well does the sentence describe the picture?

(worst) (best)

1. 1 2 3 4 5 **6** Don't Know

2. 1 **2** 3 4 5 6 Don't Know

Figure 2. Items from the auditory picture-based grammaticality judgment task ("The window closed," "The flowers (*se) are growing").

alternators and accusatives all depicted entities undergoing an agentless event, with three grammatical sentences per verb class using inchoative or passive *se* and three ungrammatical items per class omitting *se* in the same constructions. For unergatives and unaccusatives, the pictures showed entities initiating (i.e., unergatives) or undergoing (i.e., unaccusatives) events, with four grammatical sentences per class omitting *se* and four ungrammatical items per class using an overgeneralized *se*. All other items were distracters; the task took about 25 min total.

The researchers worked with the instructor and an assistant to design three versions of the two tasks based on third-year Spanish vocabulary. During test taking, learners could also ask for any vocabulary help as needed. Near-final drafts of the three versions were given to three native Spanish speakers to check for validity, and adjustments were made accordingly. Appendixes A and B list the items in each verb class on each test version's two tasks. During data gathering, final drafts of the three versions were administered in a split-block design to control for test-learning effects, so that the three versions were evenly and randomly distributed to learners as the pretest, and then each person cycled through the remaining versions for the two posttests (i.e., ABC, BCA, or CAB).

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

The equivalency of test versions was evaluated by conducting univariate ANOVAs on the immediate posttests within each verb class for the two tasks. No significant differences for test version were found in any verb class on either task, except for using *se* with accusatives on the picture description task, $F(2, 32) = 3.85$, $p = .03$, $\eta_p^2 = 0.20$. Here, a Tukey honestly significant difference post hoc analysis showed Version C to be significantly higher than A ($p = .03$), but not B. A review of accusative items across test versions revealed that on Version C, the verb *destruir*, "to destroy," beside a picture of a house collapsing elicited more tokens of *se* (e.g., *Se destruyó la casa,* "The house was destroyed") than the other accusative items on both that version and the others. This may have been because *destruir* often appeared in the last lesson's reading about an earthquake in Chile, whereas the other accusative items were either used in the two earlier lessons or drawn from common third-year vocabulary. In the end, *destruir* was retained in the means for Version C, given that a univariate ANOVA on accusative scores for the delayed posttest found no effect for test version, such that the disproportionate impact of that item had been neutralized.

Scoring and coding procedures. Scoring for both tasks began by derandomizing the items by verb class and assigning scores for targetlike uses of *se*, overgeneralization, and L1 transfer. For the picture description task, scoring isolated the use or omission of intransitive *se* so that inferential statistics could assess group results. Targetlike and overgeneralized *se* were quantified by assigning 1 point to all answers using *se* intransitively, whereas all other answers received a zero, regardless of grammaticality. Thus, targetlike alternators and accusatives (e.g., *Se secó la ropa*, "The clothes dried," or *Se lavó la ropa*, "The clothes were washed") as well as *se* overgeneralized to unaccusatives and unergatives (e.g., **Se nació el bebé*, "The baby was born," or **Juan se nadó*, "John swam") each received 1 point. Then, to assess L1 transfer errors, alternator and accusative answers were rescored so that intransitives without *se* each received 1 point (e.g., **La ropa secó*, "The clothes dried," or **La ropa lavó*, "The clothes were washed"), whereas all others received a zero. Throughout the coding, other formal features, such as subject–verb agreement or tense–aspect suffixes, were ignored unless they interfered with assessing the sentence's transitivity, in which case they, too, received a zero. After the coding was complete, an average for each participant was determined by dividing the points earned for targetlike use, overgeneralization, and L1 transfer by the number of relevant items. Individual means thus ranged from 0.00 to 1.00, with intermediate values representing a participant's percent use (or omission) of *se*. Three-way repeated-measures ANOVAs were then performed on the scores for targetlike use, overgeneralization, and transfer, as described in the Results Section, with treatment as a between-subjects factor and time of test and verb class as within-subjects factors.

On the GJ task, like the production task, items were first derandomized by verb class and grammaticality. The numbers on the Likert scale indicating learner judgments were entered into calculations of individual means within each verb class and grammaticality category. Scores thus ranged from 1 to 6, with those below the midpoint at 3.5 indicating poor acceptability for the sentence as a depiction of the picture, and higher scores indicating greater acceptability. Answers of "Don't

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Know" were treated as null data points, so that individual scores reflected only the numerical data given (Juffs, 2001). Individual scores were then entered in three-way, repeated-measures ANOVAs to assess targetlike use, overgeneralization, and L1 transfer, with treatment as a between-subjects factor and verb class and time of test as within-subjects factors.

Finally, reliability analyses were conducted on the immediate posttest results for each test version to assess internal consistency among the major response categories. For the picture description task, the six items representing *se* with alternators and accusatives (three items each) together yielded acceptable Cronbach α values: Version A = 0.91, Version B = 0.79, and Version C = 0.87. Likewise, the six items for overgeneralization of *se* with unergatives and unaccusatives yielded high α values: Version A = 0.95, Version B = 0.99, and Version C = 0.97. However, when picture descriptions were scored to assess omission of *se* with alternators and accusatives, only the α for Version C was acceptable, whereas Versions A and B showed poor consistency, with two of six items from A and three items from B automatically excluded due to a lack of tokens for the structure: Version A = –0.36, Version B = 0.28, and Version C = 0.88. The high individual variability on Versions A and B thus indicated that the various lexical items did not uniformly evoke zero-derived intransitives, which for alternators would mean L1 transfer. Instead, participants' other responses, such as transitives and, to a lesser extent, targetlike *se*, varied considerably by lexical item. Rather than exclude this data, however, it will be presented with the understanding that it unreliably assesses zero-derived intransitivity as a construct. The Discussion Section will then consider why L1 transfer may have varied among individual alternator verbs.

Last, for the GJ task, the six items representing grammatical uses of *se* with alternators and accusatives yielded the following acceptable values for Cronbach α: Version A = 0.78, Version B = 0.87, and Version C = 0.77. For the 14 ungrammatical items overgeneralizing *se* to unergatives and unaccusatives and omitting *se* with intransitive alternators and accusatives, acceptable values for α were found: Version A = 0.75, Version B = 0.85, and Version C = 0.86.

Results

Picture description task. Table 2 displays means and standard deviations for the picture description data, beginning in rows (a) and (b) with the instructionally targeted *se* with alternators and accusatives, followed in (c) and (d) by overgeneralization of *se* to unergatives and unaccusatives, and in (e) and (f) by ungrammatical, zero-derived intransitivity with alternators and accusatives. The treatment groups' scores over the three test administrations are given in the middle 6 columns of the table, whereas native speakers are on the far right.

Taking each pair of results in turn, rows (a) and (b) show that instructed learners notably increased their use of *se* with alternators, and to a lesser extent accusatives, from the pretest to the immediate and delayed posttests, whereas the control group remained near zero throughout. On the immediate posttest, the instructed group rose to 38% for alternators and 20% for accusatives, and then decreased

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Table 2. *Picture descriptions: Mean (standard deviations) percentage of use or omission of se in single-argument sentences within verb classes*

Verb Class & Sentence Type	Instructed			Control			
	Pre	Post	Delayed	Pre	Post	Delayed	Native
(a) Alternator & *se*, e.g.,	0.00	0.38	0.27	0.02	0.08	0.03	0.30
La ropa se secó.	(0.00)	(0.45)	(0.36)	(0.07)	(0.24)	(0.15)	(0.23)
(b) Accusative & *se*, e.g.,	0.00	0.20	0.13	0.00	0.02	0.02	0.05
Se lavó la ropa.	(0.00)	(0.33)	(0.25)	(0.00)	(0.07)	(0.07)	(0.12)
(c) Unergative & *se*, e.g.,	0.00	0.16	0.11	0.00	0.00	0.00	0.00
**María se nadó.*	(0.00)	(0.35)	(0.30)	(0.00)	(0.00)	(0.00)	(0.00)
(d) Unaccusative & *se*, e.g.,	0.02	0.22	0.18	0.02	0.00	0.02	0.00
**El bebé se, nació.*	(0.09)	(0.37)	(0.28)	(0.07)	(0.00)	(0.07)	(0.00)
(e) Alternator w/o *se*, e.g.,	0.09	0.09	0.28	0.02	0.00	0.30	0.00
**La ropa secó.*	(0.15)	(0.15)	(0.31)	(0.07)	(0.00)	(0.30)	(0.00)
(f) Accusative w/o *se*, e.g.,	0.04	0.00	0.07	0.00	0.00	0.00	0.00
**La ropa lavó.*	(0.11)	(0.00)	(0.19)	(0.00)	(0.00)	(0.00)	(0.00)

on the delayed posttest to 27% for alternators and 13% for accusatives. These delayed posttest means were similar to the native speakers' 30% for alternators, but considerably higher than their 5% for accusatives, due to the native speakers' frequent use of transitives and copular passives on this task, neither of which require *se*. This is interesting because it reflects the prominence of external agents in accusative verb semantics (Pustejovsky, 1995), especially given that no agents were depicted in the drawings. Meanwhile, learners' greater production of *se* with the same verbs suggests a competing instructional effect.

The three-way repeated-measures ANOVA found no significant differences for the interaction between treatment, time of test, and verb class, nor for the two-way interactions between time of test and verb class, or treatment and verb class.[6] However, there were significant effects for the interaction between time of test and treatment, $F\ (2, 66) = 4.68$, $p = .01$, $\eta_p^2 = 0.12$, and for the three main effects of verb class, $F\ (1, 33) = 12.99$, $p < .01$, $\eta_p^2 = 0.28$, time of test, $F\ (2, 66) = 8.01$, $p < .01$, $\eta_p^2 = 0.20$, and treatment, $F\ (1, 33) = 10.88$, $p < .01$, $\eta_p^2 = 0.25$, all of which yielded large effect sizes. The data in Table 2 suggest that the interaction between time of test and treatment stemmed from the instructed group's increased scores on the two posttests, compared with the unchanged control group scores. Two-tailed, independent samples *t* tests on the immediate and delayed posttests within each verb class confirmed this conclusion.[7] For alternators, the instructed group was significantly higher than the controls on both the immediate posttest, $t\ (19.84) = 2.30$, $p = .03$, $d = 0.83$, and the delayed posttest, $t\ (17.60) = 2.36$, $p = .03$, $d = 0.87$. For accusatives, however, group differences were only near significant, with the immediate posttest yielding stronger results, $t\ (15.08) = 2.12$, $p = .05$, $d = 0.75$, than the delayed posttest, $t\ (15.94) = 1.78$, $p = .09$, $d = 0.60$. Nonetheless, the ANOVA's strong main effect for treatment can be attributed to higher scores in the instructed group, whereas the main effect for time of test was due to higher scores on both posttests, and the effect for verb class stems from an overall higher use of *se* with alternators than accusatives.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Turning to overgeneralization of *se* with unergatives and unaccusatives, rows (c) and (d) in Table 2 again indicate near flat means for the control group. Meanwhile, the instructed group's near zero pretest means for both verb classes was followed by immediate posttest increases to 16% for unergatives and 22% for unaccusatives, with minor delayed posttest decreases to 11% for unergatives and 18% for unaccusatives. The three-way repeated measures ANOVA showed no significance for the interaction among treatment, time of test, and verb class, nor for the interactions between time of test and verb class, or treatment and verb class. As in the previous analysis, however, the interaction between time of test and treatment was significant, $F\ (2, 66) = 3.42$, $p = .04$, $\eta_p^2 = 0.09$, as were the main effects for verb class, $F\ (1, 33) = 5.18$, $p = .03$, $\eta_p^2 = 0.14$, and treatment, $F\ (1, 33) = 8.77$, $p = .01$, $\eta_p^2 = 0.21$, with time of test approaching significance, $F\ (1.6, 51.2) = 2.96$, $p = .07$, $\eta_p^2 = 0.08$.[8] To examine the interaction between time of test and treatment, two-tailed independent samples *t* tests were again conducted within each verb class on the two posttests. For unergatives, the differences between the control and instructed groups were not significant, but for the unaccusatives they were: on the immediate posttest, $t\ (14.00) = 2.32$, $p = .04$, $d = 0.84$, and on the delayed posttest, $t\ (15.52) = 2.19$, $p = .04$, $d = 0.76$. Meanwhile, the main effect for verb class was easily attributable to higher overall scores for unaccusatives, and the effect for treatment to higher overall scores for instructed learners. Finally, the weak main effect for time of test is not surprising, given that scores across treatments and verb classes were collapsed within each test administration. Thus, the near static control group scores and weaker postinstruction increases for unergatives diluted unaccusative increases in the instructed group. In short, the increased use of *se* with unaccusatives, although not as great as with alternators or accusatives, follows a similar developmental trajectory to those instructionally targeted verb classes, even though such overgeneralization would not have been supported by either the explicit grammar rules or ambient L2 input.

Finally, returning to Table 2, rows (e) and (f) provide means for percent omission of *se* with alternator or accusative verbs in intransitive constructions. For alternators, this meant L1 transfer of the English zero-derived inchoative (e.g., **La ropa secó*, "The clothes dried"), whereas for accusatives they were simplified passives without L1 or L2 morphosyntactic analogues (e.g., **La ropa lavó*, "The clothes washed"). From the means, it is apparent that both groups similarly increased their use of alternators without *se* on the delayed posttest, despite explicit information to the contrary in the instructed group. In addition, the two groups' low pretest and immediate posttest means for alternators, and their consistently low means for accusatives, belie differences in the structural alternatives that they employed. The instructed group made large postinstruction gains using *se* with alternators and accusatives whereas the control group did not. Hence, some instructed learners used both ungrammatical bare intransitives and *se* intransitives together after instruction, whereas the control group's only alternatives were either copular passives or transitives.

The repeated-measures ANOVA found no significant three-way interaction, nor two-way interactions between time of test and treatment, or treatment and verb class. Thus, how the treatments changed over time and how they handled each

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Table 3. *Grammaticality judgments: Mean (standard deviation) acceptability on a scale of 1–6 of single-argument sentences with or without se in each verb class*

Verb Class & Sentence Type	Instructed			Control			
	Pre	Post	Delayed	Pre	Post	Delayed	Native
(a) Alternator & *se*, e.g.,	3.56	4.83	4.60	3.42	3.45	3.57	5.20
La ropa se secó.	(0.47)	(0.76)	(0.89)	(1.12)	(1.19)	(1.05)	(0.67)
(b) Accusative & *se*, e.g.,	3.70	4.68	4.16	3.38	3.50	3.43	4.51
Se lavó la ropa.	(0.66)	(0.59)	(0.75)	(1.13)	(1.11)	(0.99)	(0.71)
(c) Unergative & *se*, e.g.,	4.07	4.12	4.14	3.93	3.55	3.95	2.95
**María se nadó.*	(0.82)	(0.73)	(0.66)	(1.19)	(1.09)	(1.22)	(1.08)
(d) Unaccusative & *se*, e.g.,	4.10	4.27	4.33	4.14	3.95	4.03	3.37
**El bebé se nació.*	(0.61)	(0.42)	(0.50)	(0.89)	(1.11)	(1.11)	(0.93)
(e) Alternator without *se*, e.g.,	4.16	3.97	3.94	3.65	4.02	3.93	2.85
**La ropa secó.*	(0.74)	(1.30)	(0.74)	(1.16)	(1.03)	(1.06)	(1.10)
(f) Accusative without *se*, e.g.,	3.39	3.36	2.98	3.14	3.02	2.73	2.80
**La ropa lavó.*	(0.98)	(1.07)	(0.88)	(1.21)	(1.14)	(1.33)	(1.07)

verb class did not statistically differ. There was, however, a significant effect for the interaction between verb class and time of test, F (2, 66) = 13.03, $p < .01$, $\eta_p^2 = 0.28$, showing that the verb class means changed differently over time. There were also significant main effects for verb class, F (1, 33) = 36.08, $p < .01$, $\eta_p^2 = 0.52$, and time of test, F (2, 66) = 21.49, $p < .01$, $\eta_p^2 = 0.39$, but no main effect for treatment. Given the lack of significant effects for treatment or its interactions, no univariate analyses on individual test administrations were conducted. Clearly, these results show that omitting *se* in intransitive sentences was concentrated among alternators, most prominently on the delayed posttest. Given the low reliability of these items on two of three test versions, along with the control group's parallel increases on the delayed posttest, the Discussion Section will argue that L1 transfer was highly determined by individual lexical items.

GJ task. The means for the picture-based GJ task are arranged in Table 3 similarly to the means in Table 2: Acceptance of *se* with intransitive alternators and accusatives appears in rows (a) and (b); overgeneralization to unergatives and unaccusatives appears in (c) and (d); and acceptance of intransitive alternators and accusatives without *se* are in (e) and (f). The six middle columns display the instructed and control group means for each test administration, with native speaker means on the far right.[9]

Considering *se* with alternators and accusatives first, a pattern like that of the picture description task is evident. Similar pretest scores in both groups were followed by a marked increase on the immediate posttest in the instructed group and modest decreases on the delayed posttest 6 weeks later. From pretest to delayed posttest, the instructed group increased by 1.04 points for *se* with alternators, coming within 0.60 of the native speakers. For accusatives, the delayed posttest showed poorer retention of immediate posttest gains, but still a net increase from the pretest of 0.46 points, which was 0.35 short of the native speakers. On the repeated-measures ANOVA, there were no significant effects for the three-way

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

interaction, or for the two-way interactions between time of test and verb class, and treatment and verb class. This suggests that verb class means did not change differently from each other over time, nor did either group distinguish the two classes. However, the two-way interaction between time of test and treatment was significant, F (2, 66) = 8.56, $p < .01$, $\eta_p^2 = 0.21$, as were the main effects of time of test, F (2, 66) = 11.57, $p < .01$, $\eta_p^2 = 0.26$, and group, F (1, 33) = 11.73, $p < .01$, $\eta_p^2 = 0.26$. This indicates that the group means changed differently over time, as the instructed group's posttreatment performance suggests. Where the main effects are concerned, the means for the two posttests are clearly higher than the pretest, whereas the instructed group's means were higher overall than the control group. The main effect for verb class, meanwhile, was not significant.

Given the interaction between time of test and treatment, two-tailed independent samples t tests were conducted within each verb class and test administration to locate differences in how the two groups changed over time. Although the treatments were not different on the pretest for either verb class, results were significant on both posttests for both verb classes: For alternators on the immediate posttest, t (33) = 3.94, $p < .01$, $d = 1.38$, and on the delayed posttest, t (33) = 3.05, $p < .01$, $d = 1.05$; for accusatives on the immediate posttest, t (30.35) = 4.07, $p < .01$, $d = 1.33$, and on the delayed posttest, t (33) = 2.38, $p < .05$, $d = 0.83$. Thus, the instructed group's gains in accepting *se* clearly distinguished them from the controls, even 6 weeks after instruction.

For overgeneralization of *se* with unergatives and unaccusatives, the means in Table 3, rows (c) and (d), reveal that unaccusatives were higher than unergatives in both groups on all three test administrations. The gap between verb classes widened in the instructed group from 0.03 on the pretest to 0.15 on the immediate posttest and 0.19 on the delayed posttest, whereas gaps in the control group started at 0.21 on the pretest, widened to 0.50 on the immediate posttest, and then narrowed to 0.07 on the delayed posttest. The smaller standard deviations in the instructed group than in the controls were also noteworthy, indicating less variation around the mean, particularly on the posttests. The three-way, repeated measures ANOVA showed no significant interactions between any factors, and no significant main effects other than for verb class, which yielded a robust result clearly attributable to higher means for unaccusatives than unergatives, F (1, 33) = 4.36, $p < .01$, $\eta_p^2 = 0.12$. Although the GJ results differ from the picture description results in that changes over time were not statistically detectable, the GJ data clearly show a preference for overgeneralizing *se* to unaccusatives over unergatives.

Finally, for accepting intransitive alternators and accusatives without *se*, the means in rows (e) and (f) of Table 3 indicate that change over time was not as great as in other comparisons of this study. The three-way repeated measures ANOVA showed no significant effects for any of the interactions or main effects except verb class, which was quite robust, F (1, 33) = 34.74, $p < .01$, $\eta_p^2 = 0.51$. Although the means for alternators in the instructed group fell only 0.22 from pretest to delayed posttest, staying within the "acceptable" portion of the scale (i.e., $M > 3.5$), the means for intransitive accusatives in both groups were the only ones to receive "unacceptable" means, coming within 0.20 points of the native speakers. Thus, although instruction did not offset L1 transfer enough to

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

reduce the acceptability of bare intransitive alternators, the much lower means for accusatives in the same syntactic configuration suggested a sensitivity to the external causation that these verbs imply.

Summary. Regarding the first research question about whether instruction would increase *se*'s targetlike use and acceptance among alternators and accusatives, the answer is "yes," particularly for alternators, where gains clearly distinguished instructed learners from controls on both tasks, even 6 weeks after instruction. For accusatives, gains were similarly sustained on the GJ task but weaker on picture descriptions, as learners, like native speakers, often used alternatives to *se* when responding. Meanwhile, the answer to the second question about whether instruction would trigger overgeneralization is again "yes," although the picture description data are more compelling, with significant increases from pretest to both posttests in the instructed group but little change in the control group. The results on both tasks also confirm that, regardless of group, unaccusative verbs were statistically preferred for overgeneralization over unergatives. Finally, with regard to whether instruction would preempt L1 transfer of zero-derived inchoative alternators, the answer is "no," given that scores for both producing and accepting this ungrammatical form remained relatively unchanged across test administrations. Furthermore, the clear difference between alternators and accusatives without *se* indicates that the transfer of English zero-derived inchoativity shaped these results, rather than a more generic tendency toward simplified detransitivization that would have encompassed both verb classes.

DISCUSSION

These results suggest that the 270-min explicit instructional treatment had a significant impact on implicit L2 knowledge. However, the effect was clearly both indirect and incomplete because noninstructional factors contributed to overgeneralization, and L1 transfer could not be preempted. Although the failure to contain these errors does suggest autonomous, learner-internal contributions to L2 development, their co-occurrence with targetlike improvements 6 weeks after instruction makes full encapsulation an implausible explanation.

Improvements with se

The impact of instruction on implicit L2 knowledge is most evident from the sustained improvement in targetlike uses of *se* on both tasks. Although DeKeyser (2003) notes that it is nearly impossible for assessments to completely isolate implicit knowledge, the fact that the GJ task involved auditory prompts under time pressure at the very least aligns the data with Ellis' (2009) factor analysis of L2 assessments, where timed and untimed GJ tasks loaded on two different factors that he attributed to implicit and explicit knowledge, respectively. In addition, despite the differing scoring schemes on the two tasks, both produced significant and large-sized effects for the interaction of time of test with treatment. Thus, if picture descriptions drew on explicit more than implicit knowledge, then both knowledge bases must have been fairly well aligned. If, however, implicit knowledge had

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

been impervious to explicit knowledge, one might have expected weaker GJ improvements given less time to monitor responses, or perhaps reduced group differences on the delayed posttest as explicit knowledge diminished. Instead, this study corroborates most classroom and laboratory instructional research, where explicit interventions brought about significant increases in targetlike outcomes that often lasted through posttesting several weeks later (deGraff & Housen, 2009; Ellis, 2002; Norris & Ortega, 2000).

These results are also consistent with White's (1991) study, where targetlike improvements in English adverb placement were still detectable through the final posttest 1 year later, despite the lack of further instruction on the topic and an L2 English program that had shifted from all-day immersion to just 90 min per week. As discussed earlier, her results were taken to mean that implicit L2 knowledge had been unaffected, mainly because increases in L1 transfer errors were interpreted in light of encapsulation and a belief that true L2 development entails once and for all parameter shifts, with no variability or backsliding (Schwartz & Gubala-Ryzak, 1992; Truscott, 1998; White, 2003; Zobl, 1995). However, given meta-analyses of more than 20 years of L2 research showing stronger results for explicit instruction over more subtle, implicit strategies (Norris & Ortega, 2000; Spada & Tomita, 2010), a valid alternative view would be that explicit and implicit knowledge work in tandem, and that explicit knowledge may be helpful, if not necessary, to contextualize, organize, and prepare learners for L2 processing in a way that builds and sustains targetlike implicit knowledge (e.g., Ellis, 2005; Paradis, 2009; Schmidt, 2001; VanPatten, 1996). Given the lack of subsequent instruction on adverbs and the precipitous decrease in class time the following year, one can only wonder how White's learners might have performed if adverbs had been revisited and reinforced in the intervening time, or at the very least their English immersion had continued.

Overgeneralization of se

Whereas generativists have at times interpreted instruction-induced overgeneralization as evidence for encapsulation (Rothman, 2008; Schwartz & Gubala-Ryzak, 1992; White, 1992, 2003), these learners' preference for overgeneralizing *se* to unaccusatives suggests that implicit knowledge must be permeable to explicit instruction, and that it consists of a sensitivity to the subtle linguistic principles associated with UG. Although linguists may debate the extent to which the unergative–unaccusative distinction is syntactic or lexicosemantic in nature (Alexiadou et al., 2004), generativists attribute knowledge of the distinction to UG given its basis in the syntax–semantics interface and the fact that it corresponds to subtle collocational properties that distinguish these verbs and their derived forms. Research in both L1 and L2 contexts has shown the unergative–unaccusative distinction to correlate with significant differences in accuracy on judgment and production tasks, as well as reaction times on NP recognition and recall tasks (Friedmann, Taranto, Shapiro, & Swinney, 2008; Hirakawa, 2001; Montrul, 2004; Toth, 1999, 2000). The present study corroborates this work, in that scores for *se* with unaccusatives were never lower than unergatives on any task or test administration, or in any group, including the native speakers. It furthermore

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

corroborates the results of Toth (2000) in that the beginning university-level L2 Spanish learners there also overgeneralized *se* to unaccusatives after instruction on nonagentive alternators. Both studies thus suggest UG-constrained implicit knowledge within a "poverty of the stimulus" logic, given that such errors defied attested forms in the L2 input, explicit information given through instruction, and analogous L1 structures that might be transferred. As such, it appears that overgeneralization could only be based on the semantic and syntactic similarities between the nonagentive subjects of unaccusatives and intransitive alternators, despite superficial similarities with unergatives as verbs appearing in subject–verb sequences in both Spanish and English.

However, although supporting the possibility that UG constrains implicit knowledge development, the data here also provide evidence against encapsulation in that the gap between unaccusatives and unergatives widened on both tasks *after* explicit instruction, even though effect sizes were only sufficient to yield a main effect for verb class, rather than an interaction between verb class and time. Still, despite small participant numbers, on the picture description task, there was a significant interaction between treatment and time of test, meaning that the instructed group departed from the controls after instruction. Accompanied by the main effect for verb class, unaccusatives reached their greatest acceptability on the delayed posttest. Meanwhile, on the GJ task, the gap was also widest at the delayed posttest, and Figure 3 shows a clear contrast between the two treatments. Here, individual GJ scores for *se* with unaccusatives are plotted on the horizontal axis against those for *se* with unergatives on the vertical axis. Although most scores from both groups fall within the upper right-hand quadrant where *se* is "acceptable" in both verb classes, a greater proportion of instructed learners clearly fall below the diagonal line representing equal scores for either class, showing a greater preference for *se* with unaccusatives than the controls. Hence, although *se* with unaccusatives earned higher means in all groups, on both tasks, and on every test administration, the pattern was most pronounced on the posttests in the instructed group. Even though this was not a statistically significant interaction of the ANOVA, it still begs the question of why larger numbers of instructed individuals should behave this way, if there is no interface between explicit and implicit L2 knowledge.

Detransitivization and L1 transfer

Meanwhile, the data on detransitivized alternators and accusatives without *se* provides evidence that, even as sensitivity to verbal argument structure may underlie implicit L2 knowledge, L1-influenced variability in how it is morphosyntactically realized must also be explained. Despite instruction on *se* and images of entities undergoing events on the picture description task, instructed learners demonstrated sensitivity to the implied external causation of accusatives when they produced more transitive answers in this class than for alternators. This explains the significant main effect for verb class, which did not appear on the more receptive GJ task. However, the results on both tasks for bare, detransitivized accusatives and alternators demonstrate how argument structure sensitivity was fortified by L1-analogous forms. Bare accusatives, implying external causation and having no L1 analogue, consistently received near zero means on picture descriptions in both

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

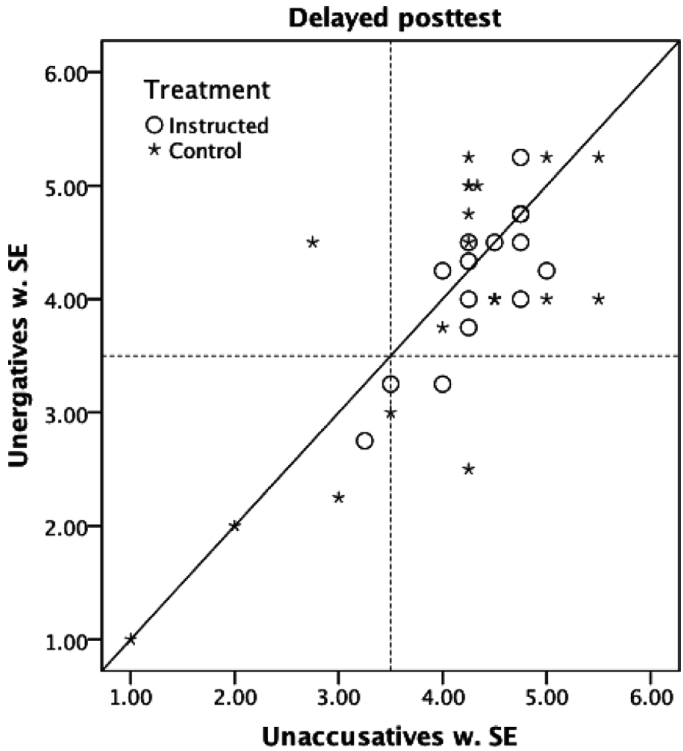
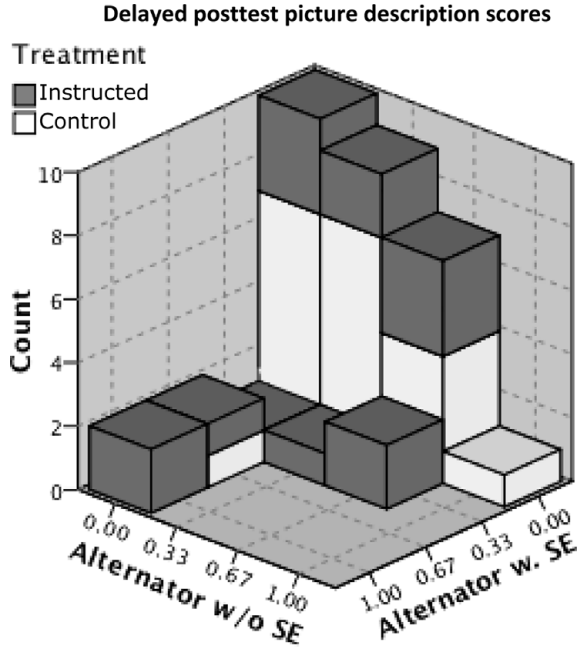


Figure 3. Scatterplot of GJ scores for overgeneralized *se*.

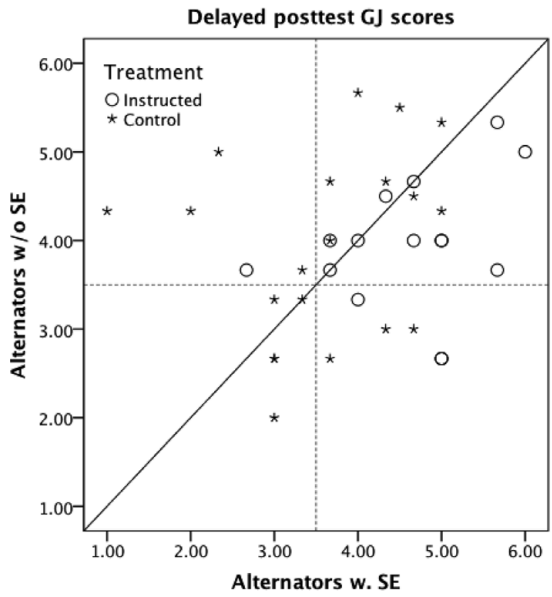
groups (Table 2, row (f)) while also being the only GJ structure to fall into the "unacceptable" range, below 3.5, throughout all testing (Table 3, row (f)). Despite instruction, however, bare intransitive alternators, which both lacked presupposed external causation and corresponded to the L1 zero derivation, increased markedly in both groups on the picture description task (Table 2, row (e)) while decreasing only slightly for instructed learners on the GJ task (Table 3, row (e)). This raises the question, then, of the status of targetlike alternators with *se* in the instructed learners' implicit knowledge and the extent to which the group means represented variability within individuals.

Figure 4 plots individual delayed posttest scores for alternators with and without *se* against each other for the picture description task in (a) and the GJ task in (b). Although variable reliance on either form within individuals is evident on both tasks, the trend among instructed learners is toward targetlike use. In (a), the vast majority of control participants appear in the farthest row to the back and right side of the three-dimensional graph, using varying percentages of alternators without *se* but none with *se*. Only one control learner used *se* with alternators two out of three times, while avoiding the L1-transferred form. Meanwhile, although nearly

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press



(a)



(b)

Figure 4. Scatterplots of delayed posttest scores for L1 transfer versus targetlike use of *se*.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

half the instructed learners also fell in the back right-hand row of the chart, where only alternators without *se* were used, the remainder did produce *se* in varying quantities, such that the majority of this subgroup trended toward the left side of the graph, where *se* is more often employed than omitted. However, because this task allowed greater time to respond, and answers could avoid intransitivity altogether, the GJ results in (b) may better assess implicit knowledge. Here, again, although the majority of learners placed in the upper right-hand quadrant of the graph, where both structures received acceptable ratings, the controls clearly show greater dispersion than the instructed group. In addition, whereas most control learners fall to the left of the diagonal line, indicating higher scores for alternators without *se* than with *se*, most instructed learners show the opposite pattern, falling to the right of the diagonal, where alternators with *se* receive higher scores. On both tasks, then, targetlike use of *se* is more a trend than a categorical shift among instructed learners, many of whom nonetheless continued relying on L1-analogous alternatives on both tasks.

It is also worth remembering that the immediate posttest results for Versions A and B of the picture description task showed poor consistency in eliciting detransitivized alternators and accusatives without *se*. Two items were excluded for lack of variation on Version A, and three were excluded from B. All of these were accusatives, and individual responses indicated a strong avoidance of bare detransitivization there, as the near zero group means in Table 2, row (f) show. Still, for items included in the reliability analyses, which were mostly alternators in Versions A and B, internal consistency was poor, meaning that individual responses for each item varied more than the overall means for this part of the test. Thus, as different lexical items elicited varying amounts of *se*'s omission within the same individual, a picture emerges of a lexically driven reliance on L1 transfer reminiscent of connectionist learning models, rather than systematic behavior. This interpretation is further supported by the higher overall use of *se* with accusative *destruir*, "destroy," on Version C's picture descriptions for the immediate posttest, which contributed to higher means on that version than the other two. The fact that *destruir* had appeared recently and saliently in the lesson just before the test, and that the difference between test versions later disappeared on the delayed posttest, suggests the initial influence of lexical exemplars for *se* spreading to a more general, rulelike pattern. Thus, whereas the use or omission of *se* showed variability within learners and suggested the presence of both structures in their implicit L2 knowledge, for many the vacillation may be best understood as lexically determined transfer competing with the conversion of *se* into a broader systemic rule.

Theoretical implications

Taken together, these results suggest that explicit instruction had a significant impact on implicit L2 knowledge, although continued overgeneralization and transfer errors show that it was clearly neither direct nor complete. On these grounds, an encapsulationist might still deny an interface between implicit and explicit knowledge, and following Chomsky (1965, pp. 3–4), attempt to tease apart those aspects of the data representing true implicit "competence" from those

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

representing explicit monitoring and other manifestations of "performance." Making inferences about a construct that is not directly observable is difficult, and apart from Ellis's (2009) proposal that timed GJ tasks could effectively assess implicit L2 knowledge, generativists have often done so with untimed tasks, based on whether learners demonstrate sensitivity to subtle distinctions that lie beyond the explicit knowledge of most nonlinguists. However, by both this criterion and Ellis's (2009), it appears that the targetlike improvements and accompanying overgeneralization represent changes in instructed learners' implicit L2 knowledge. Meanwhile, the persistence of transfer errors suggests an L1-based holdover in implicit knowledge, which may have been more lexically based than systemic by the delayed posttest. Still, whatever role lexical chunking may have played in adding *se* to implicit knowledge, overgeneralization did not target verbs with superficial similarities to the target (i.e., all intransitives), but rather only those with similar underlying semantic and syntactic characteristics (i.e., unaccusatives). Thus, for many, *se* appeared to become a systemic marker for nonagentive, derived subjects in Spanish.

Still, one could speculate, as Carroll (2001, p. 341) does, that explicit instruction may have only benefited implicit L2 knowledge through incidental, inductive input processing, and that metalinguistic noticing, awareness, and monitoring could not in principle have affected the subtle linguistic properties underlying surface forms. This would be the most conservative interpretation of the data; to refute it, one might argue that an implicit instructional treatment would have been needed for comparison with the explicit one. However, research that has made such comparisons has consistently found not only that explicit treatments generally produce stronger results but also that it is very difficult to ensure that only implicit knowledge is activated during implicit treatments (deGraff & Housen, 2009; deGraff & Housen, 2009; de Jong, 2009; DeKeyser, 2003). According to DeKeyser (2003) and Ellis (2005), many cognitive psychologists have concluded that implicit and explicit knowledge cannot but work in tandem, with explicit knowledge continually activated during speech processing to help learners interpret what they hear, assign it pragmatic intention, and direct their attention accordingly. Although as already noted, many L2 generativists are currently considering how implicit syntactic knowledge may "interface" with discourse pragmatics and other nonsyntactic cognitive resources, the links to "executive control" and explicit knowledge remain unexplored (Sorace, 2011, p. 23). Hence, it is not clear why, necessarily, explicit knowledge that a particular form with an understood meaning belongs in a particular place could never affect the processing of L2 linguistic data to restructure implicit knowledge. It not only seems counterintuitive that such activity would be no more beneficial than unfocused exposure to comprehensible L2 input, but a number of studies on monitoring in language production suggest a supportive relationship (de Jong 2009; Kormos, 2006; Toth, 2006, 2008). Although the details of how jointly operating implicit and explicit knowledge facilitate L2 development are far from resolved, at present there appear to be insufficient empirical grounds for maintaining that outside linguistic input is necessarily more privileged in shaping implicit knowledge than the "voice in one's head" where explicit knowledge acts upon speech perception and production. Contra encapsulationist reasoning, the mere fact of an indirect or delayed effect for explicit instruction cannot be

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

accepted as prima facie counterevidence, given that exposure to external input itself has only delayed and indirect effects on implicit knowledge as well.

Limitations of the study

Classroom research is always vulnerable to greater contextual constraints than the laboratory, and this study is no exception. However, these limitations create opportunities for others to follow up, and the payoff is greater ecological validity when studying L2 acquisition in one of the most common places where it occurs. Where the present study is concerned, time limitations allowed only two assessment tasks to be administered beyond the three treatment lessons, and participants included just two intact classes where permission was granted to conduct research. Future work might therefore expand the sample size and offer additional or alternative treatments and assessments. For example, an implicit–explicit instructional comparison would not only more conclusively address issues treated here simply with reference to other research but also better clarify the magnitude of the interaction between verb class and time of test for unaccusative overgeneralization. Greater information gathering about learners' proficiency in Spanish and their treatment of unaccusative and unergative verbs in English would furthermore provide a stronger basis for interpreting instructional effects. Likewise, additional assessments of implicit knowledge combined with qualitative data on learners' explicit knowledge of the target structure would provide greater evidence for how the two knowledge bases interact. Finally, because little is known about how the descriptive adequacy of explicit knowledge affects developments in implicit knowledge, exploring this relationship further, through both short-term and longitudinal work, with both quantitative and qualitative methods, will surely allow a more nuanced and psychologically plausible account of the factors influencing implicit L2 knowledge than we are currently able to articulate.

CONCLUSION

It is hoped that this study will serve as an invitation for generative and nongenerative researchers alike to continue exploring the relationship between the linguistic reality of what classroom L2 learners come to know and the instructional reality that shapes how they come to know it. Given that concurrent patterns of improvement, overgeneralization, and L1 transfer implicate both instruction-derived explicit knowledge and UG- and L1-derived implicit knowledge, it appears that the most straightforward explanation for these results is a permeable modularity between the two knowledge bases, rather than the complete encapsulation inspired by Chomskyan and Fodoran theory.

Although distinguishing the implicit linguistic system from other aspects of L2 knowledge has greatly facilitated theoretical inquiry, Jackendoff (2002) warns that treating a heuristic idealization as a true material boundary may thwart opportunities to place the work of linguistics in a broader psychological context and thereby enhance its explanatory power (pp. 33–34). The built-in weaknesses of a commitment to rigid encapsulation are apparent in Gregg (2001) and White's (2007) suggestions that UG ought only to be understood as a "property" theory

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

of L2 knowledge, and not a "transition" theory of development. Although Gregg refers to work in L2 processing as a fruitful avenue of inquiry for transition theories, and in fact some L2 processing work has been conducted within a generative framework (e.g., Juffs, 2004; Juffs & Harrington, 1995), pursuing this agenda ultimately threatens the solid walls of encapsulation, as the explicit knowledge that affects conscious noticing, attention, utterance planning, and interpretation becomes implicated in the process. Thus, within strict encapsulation, White (2003) can only conclude from her classroom studies that explicit instruction failed to affect implicit L2 knowledge and that the relevant input triggers for doing so remain unclear (p. 175). Although much work remains to determine how L2 input affects implicit knowledge, the burden of explaining variability in learner performance would be considerably lighter, and the result more psychologically plausible, if there were no need to hypothesize about how such a rigidly isolated module, whose activity can never be directly observed, would develop as a result of external stimuli. Lightening this burden would surely not only allow for significant theoretical advances, as linguistics becomes more pertinent to the broader study of L2 acquisition, but also provide a richer basis for determining the relationship between learning outcomes and instructional practices, as the importance of what teachers do and how they do it becomes much more nuanced than simply providing L2 input.

APPENDIX A

Lexical items for each verb class on the three test versions of the picture description task

Verb Class	Version A	Version B	Version C
Alternator	*encender* "turn on"	*cocinar* "cook"	*cerrar* "close"
	abrir "open"	*apagar* "turn off"	*romper* "break"
	romper "break"	*derretir* "melt"	*mover* "move"
Accusative	*oír* "hear"	*traer* "bring"	*destruir* "destroy"
	limpiar "clean"	*servir* "serve"	*pedir* "order"
	pagar "pay"	*mirar* "watch"	*reparar* "fix"
Unergative	*llorar* "cry"	*caminar* "walk"	*correr* "run"
	nadar "swim"	*viajar* "travel"	*jugar* "play"
	caminar "walk"	*esquiar* "ski"	*patinar* "skate"
Unaccusative	*llegar* "arrive"	*descansar* "rest"	*salir* "leave"
	nacer "be born"	*bucear* "dive"	*volver* "return"
	desaparecer "disappear"	*asistir* "attend"	*descansar* "rest"

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

APPENDIX B

Lexical items for each verb class on the three versions of the picture-based grammaticality judgment task

Verb Class	Version A	Version B	Version C
Alternator	*apagar* "turn off"	*abrir* "open"	*descomponer* "break"
	cerrar "close"	*romper* "break"	*cocinar* "cook"
	descomponer "break"	*mojar* "get wet"	*ensuciar* "get dirty"
Accusative	*tocar* "play (music)"	*construir* "build"	*servir* "serve"
	cortar "cut"	*lavar* "wash"	*lavar* "wash"
	necesitar "need"	*destruir* "destroy"	*vender* "sell"
Unergative	*viajar* "travel"	*nadar* "swim"	*esquiar* "ski"
	bailar "dance"	*llorar* "cry"	*andar en bici* "bike"
	patinar "skate"	*jugar* "play"	*trabajar* "work"
	jugar "play"	*charlar* "chat"	*caminar* "walk"
Unaccusative	*descender* "descend"	*aparecer* "appear"	*descender* "descend"
	salir "leave"	*crecer* "grow"	*llegar* "arrive"
	descansar "rest"	*llegar* "arrive"	*bucear* "dive"
	volver "return"	*nacer* "be born"	*asistir* "attend"

ACKNOWLEDGMENTS

This research was supported with funding from the University of Wisconsin Graduate School and Center for Latin American, Caribbean, and Iberian Studies; Temple University Provost's and Faculty Development Offices; and the University of Plymouth Business School. We express our deepest gratitude to the several individuals who assisted in developing the teaching and testing materials for this project, as well as those who helped to implement the instructional treatments. They include Kelly Ferguson, Daniel Friedrich, Angie Garritano, Ellen Grunder, Sara Hagen, Denise Hanson, Johanna Hernández, Lindsay March, Ashley Moncayo, Aneidys Rodríguez, and Joanna Steele. Invaluable advice on the design and statistical procedures for this project was also provided by Adam Gamoran and Rick Nordheim at the University of Wisconsin and Richard Heiberger and Josh Klugman at Temple University. Thank you all very much for your generosity.

NOTES

1. However, *se* does optionally appear with some inherent intransitives as a telicity marker (McCready & Nishida, 2008). Although many believe that this "aspectual *se*" is related to its argument-absorbing function elsewhere, instruction here only addressed *se*'s link to missing arguments, and tokens of aspectual *se* only appeared incidentally in the ambient input.
2. Larson-Hall (2010), citing Cohen (1988) and others, labels the η_p^2 values of 0.01, 0.06, and 0.14 as small, medium, and large, respectively.
3. All teaching materials and grammar explanations are available from the first author upon request.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

4. Although corroborating measures of implicit knowledge would have been desirable, the roughly 35 min required to administer these two tasks was the maximum allowed in this school setting, given our use of three instructional days for the treatments.
5. A graded rather than binary GJ scale was used because beginning to intermediate learners are often reluctant to decisively reject ungrammatical L2 sentences, thus making a graded scale better suited to capturing their intuitions.
6. Note this lack of significance was found despite effect sizes that were greater than the minimum detectable $\eta_p^2 = 0.04$ determined in the statistical power analysis.
7. At times, Welch's procedure was applied to the *t* tests to adjust for nonhomogeneity of variance. This occurred when the instructed group began using *se* and hence exhibited considerably greater variation than the controls, who scarcely used *se* at all.
8. These results reflect a Greenhouse–Geisser adjustment for nonsphericity
9. On GJ tasks, it is common for native speaker standard deviations to be higher for ungrammatical items (rows (c)–(f)) than grammatical items, because some individuals only reluctantly reject items on a test, even if they are ungrammatical (e.g., Bley-Vroman & Yoshinaga, 2000).

REFERENCES

Alexiadou, A., Anagnostopoulou, E., & Everaert, M. (Eds.). (2004). *The unaccusativity puzzle: Explorations of the syntax–lexicon interface*. Oxford: Oxford University Press.

American Council on the Teaching of Foreign Languages. (1999). *ACTFL proficiency guidelines—Speaking*. Alexandria, VA: Author.

Bailey, N., Madden, C., & Krashen, S. D. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning, 24*, 235–243.

Belikova, A. (2008). Explicit instruction vs. linguistic competence in adult L2-acquisition. In H. Chan, H. Jacob, & E. Kapia (Eds.), *BUCLD 32: Proceedings of the 32nd Annual Boston University Conference on Language Development* (Vol. 1, pp. 48–59). Somerville, MA: Cascadilla Press.

Bley-Vroman, R., & Yoshinaga, N. (2000). The acquisition of multiple *wh*-questions by nonnative speakers of English. *Second Language Research, 16*, 3–26.

Bruhn-Garavito, J. (1995). L2 acquisition of verb complementation and binding principle B. In F. R. Eckman, D. Highland, P. W. Lee, J. Mileham, & R. Rutkowski-Weber (Eds.), *Second language acquisition theory and pedagogy* (pp. 79–100). Mahwah, NJ: Erlbaum.

Burzio, L. (1986). *Italian syntax: A government and binding approach*. Dordrecht: Reidel.

Carroll, S. E. (2001). *Input and evidence: The raw material of second language acquisition*. Philadelphia, PA: John Benjamins.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*, 335–344.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2007). Of minds and language. *Biolinguistics, 1*, 9–27.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*, 529–555.

deGraff, R., & Housen, A. (2009). Investigating the effects and efectiveness of L2 instruction. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 756–778). Malden, MA: Wiley–Blackwell.

de Jong, N. (2009). Second language learning of grammar: Output matters too. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 95–115). Bristol: Multilingual Matters.

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 311–348). Oxford: Blackwell.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Doughty, C. J. (1991). Second language instruction does make a difference. *Studies in Second Language Acquisition, 13*, 431–469.
Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Oxford: Blackwell.
Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305–352.
Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24*, 223–236.
Ellis, R. (2009). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31–64). Bristol: Multilingual Matters.
Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
Friedmann, N., Taranto, G., Shapiro, L. P., & Swinney, D. (2008). The leaf fell (the leaf): The online processing of unaccusatives. *Linguistic Inquiry, 39*, 355–377.
Gregg, K. R. (2001). Learnability and second language acquisition theory. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 152–180). Cambridge: Cambridge University Press.
Hirakawa, M. (2001). L2 acquisition of Japanese unaccusative verbs. *Studies in Second Language Acquisition, 23*, 221–246.
Hopper, P. J. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (Vol. 1, pp. 155–176). Mahwah, NJ: Erlbaum.
Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*, 168–196.
Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
Juffs, A. (2001). Discussion: Verb classes, event structure, and second language learners' knowledge of semantics–syntax correspondences. *Studies in Second Language Acquisition, 23*, 305–314.
Juffs, A. (2004). Representation, processing, and working memory in a second language. *Transactions of the Philological Society, 102*, 199–225.
Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in Wh-extraction. *Studies in Second Language Acquisition, 17*, 483–516.
Koontz-Garboden, A. (2009). Anticausativization. *Natural Language and Linguistic Theory, 27*, 77–138.
Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.
Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Englewood Cliffs, NJ: Prentice–Hall.
Krashen, S. D., Sferlazza, V., Feldman, L., & Fathman, A. (1976). Adult performance on the SLOPE test: More evidence for a natural sequence in adult second language acquisition. *Language Learning, 26*, 145–151.
Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.
Levin, B., & Rappaport-Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.
Long, M. H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly, 17*, 359–382.
Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Erlbaum.
McCready, E., & Nishida, C. (2008). Reflexive intransitives in Spanish and event semantics. In J. Dölling, T. Heyde-Zybatow, & M. Schäfer (Eds.), *Event structures in linguistic form and interpretation* (pp. 223–244). Berlin: Walter de Gruyter.
Montrul, S. (2004). Psycholinguistic evidence for split intransitivity in Spanish second language acquisition. *Applied Psycholinguistics, 25*, 239–267.
Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition*. Oxford: Blackwell.
Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Philadelphia, PA: John Benjamins.
Perlmutter, D. (1978). Impersonal passives and the Unaccusative Hypothesis. In J. Jaeger (Ed.), *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society* (pp. 157–189). Berkeley, CA: Berkeley Linguistics Society.
Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning, 33*, 465–497.
Pica, T. (1987). Second-language acquisition, social interaction, and the classroom. *Applied Linguistics, 8*, 3–21.
Pienemann, M. (1984). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6*, 186–212.
Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
Reinhart, T., & Siloni, T. (2005). The lexicon–syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry, 36*, 389–436.
Rothman, J. (2008). Aspect selection in adult L2 Spanish and the competing systems hypothesis. *Languages in Contrast, 8*, 74–106.
Rothman, J. (2010). Theoretical linguistics meets pedagogical practice: Pronominal subject use in Spanish as a second language. *Hispania, 93*, 52–65.
Rothman, J., & Slabakova, R. (Eds.). (2011). The mind–context divide: On acquisition at the linguistic interfaces [Special issue]. *Lingua, 121*, 567–688.
Schmidt, R. W. (2001). *Attention*. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.
Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*, 147–164.
Schwartz, B. (1999). Let's make up your mind: "Special nativist" perspectives on language, modularity of mind, and nonnative language acquisition. *Studies in Second Language Acquisition, 21*, 635–656.
Schwartz, B., & Gubala-Ryzak, M. (1992). Learnability and grammar reorganization in L2A: Against negative evidence causing the unlearning of verb movement. *Second Language Research, 8*, 1–38.
Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism, 1*, 1–34.
Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263–308.
Talmy, L. (2008). Aspects of attention in language. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 27–38). New York: Routledge.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Toth, P. D. (1999). Preemption in instructed learners of Spanish as a foreign language: Acquiring a rule for *se*. *Spanish Applied Linguistics, 3*, 195–246.
Toth, P. D. (2000). The interaction of instruction and learner–internal factors in the acquisition of L2 morphosyntax. *Studies in Second Language Acquisition, 22*, 169–208.
Toth, P. D. (2006). Processing instruction and a role for output in second language acquisition. *Language Learning, 56*, 319–385.
Toth, P. D. (2008). Teacher- and learner-led discourse in task-based grammar instruction: Providing procedural assistance for L2 morphosyntactic development. *Language Learning, 58*, 237–283.
Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*, 103–135.
Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141–178). Washington, DC: Georgetown University Press.
VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition: Theory and research*. Norwood, NJ: Ablex.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press

White, L. (1991). Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom. *Second Language Research, 7*, 133–181.
White, L. (1992). On triggering data in L2 acquisition: A reply to Schwartz and Gubala-Ryzak. *Second Language Research, 8*, 120–137.
White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.
White, L. (2007). Linguistic theory, universal grammar, and second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 37–56). Mahwah, NJ: Erlbaum.
Zobl, H. (1995). Converging evidence for the acquisition/learning distinction. *Applied Linguistics, 16*, 35–57.

https://doi.org/10.1017/S0142716412000197 Published online by Cambridge University Press