Optimizing Language Instruction: Matters of Explicitness, Practice, and Cue Learning

Catherine A. Stafford
University of Wisconsin–Madison

Harriet Wood Bowden
University of Tennessee–Knoxville

Cristina Sanz
Georgetown University

Input exposure is essential for nonprimary language learning, but the importance of explicit instruction and corrective feedback continues to be debated. If instruction is required, how might it be optimized in terms of its nature and timing? In this study, 65 Spanish-English bilinguals were introduced to Latin through an interactive computer program designed to promote initial learning of thematic role assignment by drawing learners’ attention to efficient input processing strategies when new, informative morphosyntactic cues to thematic role assignment were set up in competition with more familiar but potentially misleading cues. Participants in 4 experimental groups completed input-based, task-essential practice with interpreting agent/patient roles in Latin sentences and received feedback throughout the practice session. Two of the groups additionally received prepractice explanation of how thematic roles are assigned in Latin via morphosyntactic cues. Groups also varied systematically in whether they received less or more explicit feedback during practice. Results suggest that practice and less explicit feedback were sufficient to trigger improvement in the ability to interpret Latin case

This study was funded by grants to Stafford from the University of Wisconsin–Madison Graduate School, sponsored by the Wisconsin Alumni Research Foundation. The study is part of The Latin Project, with materials developed by the authors with support from Georgetown’s GSAS and Spencer Foundation grants to Sanz. We thank Bill Garr and Ru San Chen of Georgetown’s UIS/CNDLS for their invaluable assistance. We are also grateful to three anonymous Language Learning reviewers for their extremely helpful feedback and to Lourdes Ortega for her extraordinarily meticulous work as editor. Any remaining errors are our own.

Correspondence concerning this article should be addressed to Catherine Stafford, University of Wisconsin–Madison, Department of Spanish and Portuguese, 1018 Van Hise Hall, 1220 Linden Drive, Madison, WI 53706. Internet: cstafford@wisc.edu
morphology; however, more explicit, metalinguistic feedback was necessary to promote improvement in production. Prepractice explanation without metalinguistic feedback during practice did not significantly influence development of either interpretive or productive abilities.

**Keywords** explicit instruction; feedback; Competition Model; computer-assisted language learning; Latin; input-based instruction

Input exposure is required for nonprimary language learning, but the nature and timing of instruction, including whether it should incorporate overt grammar explanation and whether explanation should be provided preemptively or reactively, remain topics of debate. Several researchers have found a facilitative role for grammar explanation either prior to or during second-language (L2) exposure (e.g., Farley, 2004; Fernández, 2008; Henry, Culman, & VanPatten, 2009; Rosa & Leow, 2004; see Norris & Ortega, 2000, for a meta-analysis), whereas others have found little or no benefit for grammar explanation when L2 exposure consisted of task-essential practice that obliged learners to connect form and meaning (e.g., Benati, 2004; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Sanz, 2004; Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996; Wong, 2004, 2009; see Sanz & Morgan-Short, 2005, for a review).

In this study we investigated the comparative effectiveness of four instructional treatments that differed in their degree of instructional explicitness as well as the timing of instruction on initial learning of Latin. The computer-based treatments were devised to help participants alter their input processing strategies so that they would rely less on familiar morphosyntactic cues and more on those that are more effective for processing Latin. All participants completed input-based, task-essential practice with assigning thematic roles (deciding who did what to whom) at the sentence level in Latin. All participants also received right/wrong feedback throughout the practice session. Two groups additionally received prepractice explanation of how thematic roles are assigned in Latin (the +GE groups), and two groups received feedback during practice that included metalinguistic information in addition to the right/wrong indication (the +EF groups).

**Explicit Information, Structured Input, and L2 Learning**

The explicit instructional technique known as processing instruction (PI; VanPatten, 2005) seeks to promote the creation of form-meaning connections through input-based activities that are task-essential (Loschky & Bley-Vroman, 1993) in the sense that successful task completion entails that learners connect
targeted forms and meanings in appropriate ways. Several studies have isolated and assessed PI’s two main components—explicit information (EI) and structured input (SI)—in order to determine whether it was one, the other, or a combination of the two components that was responsible for promoting L2 learning. A key feature of EI is pointing out to learners the suboptimal first-language (L1) strategies they tend to employ during L2 processing; SI activities are designed to draw learners’ attention to what are often less salient, more complex features of the L2 so that they are noticed and processed in ways that promote form-meaning connections. Feedback provided during SI activities indicates whether learners’ responses are correct but does not include any additional explanation.

The study by VanPatten and Oikkenon (1996) was one of the first studies to investigate the contribution of EI to overall PI benefits for L2 learning of Spanish direct object clitics among L1 English high school students. One treatment group received full PI (i.e., both EI and SI) and the other two groups received one or the other of the PI components. After reading a grammar explanation, the EI group completed distractor activities and the PI group completed the same sequence of SI activities as the SI group. The SI and full PI groups significantly outperformed the EI group on an interpretation posttest. On a production test, however, only the full PI group performed significantly better than the EI group; the SI group was not statistically distinguishable from either the full PI or the EI group.

Several partial replications of this study subsequently appeared, including Wong (2004, 2009), Benati (2004), and Farley (2004). By and large, the authors of these and other PI studies have claimed to have provided evidence that SI is the element of PI that facilitates second language acquisition (SLA) and that any additional benefit for preemptive instruction is minimal at best. However, counterevidence for this claim is found in the production test results of all five studies cited except for Benati’s; namely, neither the SI group in the original study nor the SI groups in the two partial replications by Wong differed significantly from the corresponding EI groups in posttest production performance, and Farley’s PI group scored significantly better than his SI group. However, the objective of these PI studies was to examine whether EI provided any benefit over and above what learners gained through completion of SI activities, not to compare EI and SI as independent methods of instruction. Indeed, the EI groups in these studies had considerably less time on task than the PI and SI groups, but as Wong (2004) argued, this difference was irrelevant for their purposes.

More recent information processing studies have compared more and less explicit training conditions directly. Morgan-Short et al. (2010) controlled time
on task in their study of artificial language learning under explicit and implicit training conditions. The linguistic targets for the experiment were noun-article and noun-adjective agreement. Participants in both of the training conditions completed interpretation and production practice in a computer game format. Those assigned to the explicit condition listened to computer-delivered grammar explanation before practice. To control time on task, those in the implicit condition listened to additional sentences in the artificial language, but without any overt grammar explanation. Learning was assessed with a grammaticality judgment task once when participants reached a criterion of low proficiency (operationalized as above-chance performance on the interpretation activity) and again after they completed a second training session. Results of the second assessment indicated significant improvements in judging noun-article agreement in both training conditions, but only the implicit training group improved significantly in judging noun-adjective agreement. Therefore, not only did L2 outcomes suggest no benefit for provision of explicit metalinguistic information, but in fact more learning was evidenced among those who learned under implicit conditions.

Fernández (2008) focused on online processing rather than learning outcomes in her study of the effects of EI on L1 English adults’ processing of L2 Spanish object-verb-subject (OVS) sentences in one experiment and of subjunctive constructions in a second one. Her objective was to investigate whether provision of EI before SI helped learners process more efficiently or faster than learners who did not receive EI. Three assessment measures were used: response times for SI items, the number of trials it took participants to provide four consecutive correct responses, and accuracy after this criterion was met. For the processing of OVS sentences, groups were not significantly different from each other on any measure, suggesting no role for EI. However, for processing of subjunctive constructions, the PI group significantly outperformed the SI group on all three measures. Fernández posited that EI was beneficial when participants had to process just one form, as she claimed was the case with subjunctive sentences, but not with OVS constructions, in which participants had to process more than one form, and the relationship between forms, in order to assign agent and patient roles respectively to sentence-final noun phrases and sentence-initial object clitics.

Henry et al. (2009) did a partial replication of Fernández’s (2008) study in which the target form was nominative and accusative case marking on articles in L2 German sentences presented in either SVO or OVS word order. They employed Fernández’s measure of number of trials to criterion and found, contrary to the original study, that the participant group that received EI in
addition to SI reached criterion significantly faster than the group that completed SI only, suggesting that EI helped participants correctly process OVS sentences sooner than participants whose treatment did not include EI. Henry et al. explained the discrepancy between the studies’ results in terms of morphological complexity and its effect on how “portable” (p. 572) the EI was. The target structures in Fernández’s study involved four accusative clitic pronouns (third-person singular, plural, masculine, and feminine) that participants would have to maintain in memory to process sentences correctly. In contrast, the target in Henry et al.’s study involved distinguishing between nominative and accusative articles in their masculine singular forms (der and den) only. The authors posited that this dichotomous distinction would have been far easier to maintain in memory—thus making it more portable—for processing L2 input.

Taken together, the findings of the preceding studies suggest that the degree of effectiveness of preemptive instruction—when effectiveness is appreciable—interacts with the L2 target form(s). Questions that emerge from the results of these studies concern the utility of feedback as an alternative or supplement to EI and the relative efficacy of more and less explicit feedback for promoting L2 learning. We turn now to some recent studies that have investigated these questions.

**Preemptive Instruction, Feedback, and L2 Learning**

A different strand of PI work has focused on manipulating the degree of explicitness of feedback provided during SI activities while keeping the provision of EI constant. Sanz (2004) conducted the first study of this kind with L1 English adults learning L2 Spanish direct object clitics. Both groups in this study completed SI-type treatments, and although all participants were alerted when they responded incorrectly during practice, those in a more explicit feedback group received additional metalinguistic information. Results showed no reliable between-groups difference in L2 performance on interpretation and production posttests, suggesting that there was no advantage for providing metalinguistic explanation as part of feedback.

In a computer-based study that likewise targeted acquisition of L2 Spanish clitics, Sanz and Morgan-Short (2004) isolated the effects of preemptive grammar explanation and explicit feedback provided during task-essential practice. Four participant groups were formed on the basis of provision or not of preemptive explanation and concurrent feedback during input-based practice. All groups improved significantly in their ability to interpret and produce object clitics and there was no significant between-groups difference in
posttest performance, leading the researchers to conclude that task-essential practice was necessary and sufficient to facilitate L2 learning.

Rosa and Leow (2004) also used a computer-based design to investigate learning of past conditional constructions in L2 Spanish by L1 English adults. Three task features were manipulated for this study: provision of preemptive explanation, provision of feedback, and explicitness of feedback. Participants in five experimental groups solved multiple-choice format jigsaw puzzles that required them to join a main clause with its corresponding subordinate clause to construct conditional sentences. A control group was exposed to the structure through a reading passage. The experimental groups significantly outperformed the control group on immediate and delayed recognition and production tests. Specific results for exemplars not seen during treatment revealed that the group that received explicit feedback alone and the two groups that received two sources of information—preemptive grammar explanation plus either more or less explicit feedback—outperformed the other two groups that received information through just one source. The results of this study should be interpreted with caution, however, in view of the fact that practice consisted of only 18 exemplars, which may not have provided enough evidence for participants in the less explicit groups to induce a rule.

**Pending Issues**

In summary, results of the foregoing studies by and large reveal minimal benefits of preemptive instruction for L2 outcomes when L2 exposure involves task-essential practice. Nevertheless, EI may help learners make more complex form-meaning connections like those associated with the subjunctive (Farley, 2004; Wong, 2009). In studies that have focused on L2 processing rather than outcomes, Fernández (2008) and Henry et al. (2009) found that EI may play a role in altering participants’ online processing at early stages of L2 learning. Finally, with regard to the role of metalinguistic feedback provided during L2 practice, whereas Rosa and Leow’s (2004) results indicated a facilitative role for some L2 outcomes, Sanz’s (2004) and Sanz and Morgan-Short’s (2004) did not. This complicated landscape of findings is a reminder of the interactive relationships between types of instructional interventions, particular grammatical structures, and methods of assessment—relationships that require further clarification.

A factor that may have influenced the results of the preceding studies is previous L2 exposure. With the exception of Morgan-Short et al.’s (2010) participants, who were introduced to an artificial language, participants in these
studies ranged from novice to advanced L2 learners. Although the reported pretest scores suggested minimal knowledge of targeted structures, the potential influence of previous L2 instruction cannot be discarded. To control for effects of previous instruction, the current study introduced na"ıve learners to Latin, specifically to morphosyntax associated with assigning thematic agent and patient roles. As is further detailed in the Method section, participants’ knowledge of Spanish and English was expected variously to facilitate and inhibit successful initial Latin learning and it was this competition (MacWhinney, 2005) that was of empirical interest. The research questions that guided the study were the following: (1) Does prepractice grammar explanation influence initial learning of Latin morphosyntax associated with thematic role assignment? (2) Does explicit corrective feedback during practice influence initial learning?

Method

Participants
Participants in the study were 65 Spanish-English bilingual adults living in the United States. The sample included heritage Spanish speakers as well as early and late bilinguals who were proficient to highly proficient English speakers. We chose this type of sample for several reasons. First, they were na"ıve learners of Latin grammar, which was confirmed by poor performance on a Latin grammar pretest (see Latin Test Battery subsection). Second, knowledge of Latin cognates in both Spanish and English was likely to make participants fast Latin vocabulary learners, which was beneficial for the study, given its focus on learning of morphosyntax. Finally, investigating the effects of practice, explicitness, and cue learning with these bilingual participants helps broaden the SLA field’s spectrum of knowledge about different types of learners, a call that has been made recently by several SLA researchers (e.g., Cook, 2009).

Participants were recruited on two large university campuses and in the surrounding communities and they were sought through advertisements in Spanish language newspapers, postings on listservs, flyers posted in area libraries and businesses, and by word of mouth. They were paid for their participation. As the summary of demographic data in Table 1 shows, the sample represented a range of current age, age of onset of bilingualism, length of residence in the United States, L1 and L2 proficiency, as well as level of formal education.

A multivariate ANOVA was conducted to compare treatment groups on demographic variables and no significant difference was found, Wilks’s $\Lambda = .82$, $F(18, 159) = 0.64$, $p = .86$. A multivariate $\eta^2 = .06$ indicates that these
Table 1  Descriptive statistics for age and language variables by treatment group

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Variable</th>
<th>+GE+EF (n = 18)</th>
<th>–GE+EF (n = 20)</th>
<th>+GE–EF (n = 16)</th>
<th>–GE–EF (n = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AoT</td>
<td>28.5 (7.8)</td>
<td>31.3 (8.7)</td>
<td>31.1 (7.6)</td>
<td>33.8 (8.1)</td>
</tr>
<tr>
<td></td>
<td>AoA</td>
<td>16.1 (14.0)</td>
<td>21.5 (11.7)</td>
<td>20.8 (11.9)</td>
<td>24.6 (14.0)</td>
</tr>
<tr>
<td></td>
<td>LoR</td>
<td>12.4 (9.5)</td>
<td>9.8 (8.7)</td>
<td>10.3 (8.5)</td>
<td>9.2 (10.2)</td>
</tr>
<tr>
<td></td>
<td>L1 Prof</td>
<td>95.6 (8.3)</td>
<td>97.9 (5.8)</td>
<td>96.9 (7.2)</td>
<td>97.7 (3.9)</td>
</tr>
<tr>
<td></td>
<td>L2 Prof</td>
<td>83.7 (14.6)</td>
<td>84.0 (12.0)</td>
<td>83.3 (14.4)</td>
<td>80.3 (13.6)</td>
</tr>
<tr>
<td></td>
<td>Educ</td>
<td>3.5 (0.9)</td>
<td>3.2 (0.9)</td>
<td>3.3 (0.9)</td>
<td>3.6 (0.8)</td>
</tr>
</tbody>
</table>

Note. AoT = current age (in years), AoA = age (in years) of arrival to the United States, LoR = length (in years) of residence in the United States, L1 Prof = self-rated Spanish proficiency, L2 Prof = self-rated English proficiency, Educ = level of education. Proficiency means are expressed as proportions. For Educ, a response of 3 represented a Bachelor’s degree or the equivalent and a 4 represented a Master’s degree or the equivalent; +/–GE = provision or not of prepractice grammar explanation during treatment, +/–EF = provision or not of metalinguistic feedback during treatment; means are presented with SD in parentheses.

combined variables account for just 6% of global variation, leading us to conclude that the range of experience in the sample would not be problematic for comparing the effects of the four instructional treatments on learning outcomes.

**Target Form**

Given that a primary interest for this study was the examination of how participants connected form and meaning under different instructional conditions, the Competition Model (CM; Bates & MacWhinney, 1982; see also MacWhinney, 2005) was the theoretical framework we adopted. According to the CM, the strength of cues in linguistic input governs the mapping between form and function. A cue’s strength is determined by its validity, which refers to cue availability, or the frequency with which the cue marks a particular function, and cue reliability, or the consistency with which an available cue uniquely marks a function. A language-specific cue hierarchy results from the relative strength of its cues, and learning under the framework of the CM is defined as restructuring cue hierarchies.

The linguistic target for this study was Latin morphosyntax related to the assignment of thematic agent/patient roles to nouns at the sentence level. Latin was chosen because its morphosyntactic structure includes three cues that learners may rely on for thematic role assignment: verb agreement, word
order, and noun case morphology (specifically, nominative and accusative case morphemes). Whereas the first two of these cues are strong cues for assigning semantic function in both Spanish and English, the third represents a new cue—the strongest in the Latin hierarchy—that is the only cue that consistently maps onto thematic roles in Latin. Thus, participants would have to learn to rely on this unfamiliar cue over the strong Spanish/English cues in order to assign thematic roles correctly. A fourth cue that can be used to assign thematic roles is animacy; however, as all nouns in the input were animate, this cue was controlled for in this study.

Practice and critical test items, examples of which are presented in (1)–(3), consisted of two animate nouns and a transitive verb. Verb agreement and word order were manipulated so that they would be inconsistently available. In sentences in which one noun is singular and the other plural, as in (2), verb agreement is an informative cue for determining which noun is the agent. In sentences of this type, agreement and case morphology converge on the same sentence interpretation. In contrast, in sentences in which nouns are either both singular or both plural, as in (3), word order and case morphology cues potentially compete (i.e., suggest different interpretations for the sentence). Verb agreement was varied so that it, too, would contribute to creating competition between cues for assigning thematic roles.

(1) Potentissimus auscultat reginam.
  king-nom. sing. listen-third sing. queen-acc. sing.
  *The king listens to the queen.*

(2) Potentissimi reginam auscultant.
  kings-nom. pl. queen-acc. sing. listen-third pl.
  *The kings listen to the queen.*

(3) Reginas auscultant potentissimi.
  queens-acc. pl. listen-third pl. kings-nom. pl.
  *The kings listen to the queens.*

Target sentences were generated from a list of 35 nouns and 11 verbs, all of which are morphophonologically regular. Each noun function (e.g., feminine singular patient) is marked by a unique inflection (e.g., –*am*) so that morphosyntax is a highly reliable cue to assigning thematic roles correctly.

Prepractice grammar explanation exposed participants to 22 target sentences and the practice session exposed them to a total of 93 target sentences. Practice items included 39 written sentences and 27 aural sentences; the latter were repeated once each. As a means of controlling time on task, participants who did not receive the grammar explanation repeated the practice session;
therefore, these groups were exposed to a total of 186 target sentences and the groups that received prepractice explanation were exposed to a total of 115 target sentences plus grammar explanation.

**Overview of Procedure**
A Web-based application combining Flash and ColdFusion programming administered treatment and tests and participants’ answers were automatically uploaded to a database. Participants completed the study in three sessions, meeting individually or in pairs with one of the researchers. An outline of the experimental design is displayed as Figure 1.

**Treatment**
The four instructional treatments selected for this study varied in their degree of explicitness as well as in the timing of grammar explanation (when it was provided) in order to investigate whether prepractice grammar explanation
and/or provision of more or less explicit feedback during task-essential practice would facilitate learning of a new cue for assigning thematic roles, namely, noun case morphology. Two of the treatments included prepractice grammar explanation (+GE) and the other two did not (–GE); two treatments provided more explicit feedback (+EF) during practice and two treatments provided less explicit feedback (–EF). Participants in all treatment groups completed input-based, interactive practice activities designed to promote processing of case morphology such that it would be linked with the meaning it encodes. Individual features of the instructional treatments are detailed below.

**Vocabulary Lesson and Quiz**

Before the pretest, participants completed a computer-delivered vocabulary lesson that introduced them to the Latin words they would use throughout treatment and testing. First, two images depicting the word’s meaning appeared onscreen. Next, the Latin word appeared with its pronunciation presented simultaneously through headphones. Finally, the word’s English equivalent appeared. (See Appendix, Figure 1.1 in the online Supporting Information.)

Vocabulary learning was assessed with a multiple-choice quiz and participants answered with mouse clicks. Immediate feedback was provided, including the correct answer for any incorrect response. Participants were required to pass the vocabulary quiz to a criterion of 100% and the computer application cycled back through either the entire vocabulary lesson (if a participant scored 60% or lower on the quiz) or a review of items missed on the quiz (if a participant scored between 60% and 100%) until the 100% criterion score was met. Participants also completed multiple-choice vocabulary reviews before each of the language tests to ensure that errors made during testing were the result of difficulties with the target structure rather than with lexical knowledge.

**Grammar Explanation**

When provided (for the two +EF groups), the grammar explanation was self-paced and participants controlled advancement through its 70 screens with the computer mouse. The lesson provided explicit metalinguistic information in English about how thematic roles are assigned to nouns in Latin. First, the functional difference between nominative (referred to in the explanation as “subject”) and accusative (referred to as “object”) case was explained and participants interacted with several examples. Next, the Latin morphemes used to mark nominative and accusative case of masculine singular, feminine singular, masculine plural, and feminine plural nouns were presented, with individual presentation of each morpheme followed by a series of interactive examples.
Latin’s flexible word order and the difference between singular and plural verb forms were pointed out; within this context, helpful and unhelpful processing strategies were highlighted, similar to the technique employed in processing instruction. Onscreen feedback was provided for responses to all interactive examples included in this portion of instructional treatment. The grammar explanation concluded with a summary table that organized the Latin morphemes by case, gender, and number.

**Practice and Feedback**

Task-essential practice consisted of three written and three aural interpretation tasks. Reading and listening tasks were included in order to provide a range of practice in both of these language modalities. Two of the practice tasks required participants to read or listen to a Latin sentence and select which of two onscreen photographs accurately depicted the sentence’s meaning. (See Appendix, Figure 1.2 in the online Supporting Information.) Two tasks required participants to read or listen to a Latin sentence and interpret it by choosing its English translation from two onscreen options. A third type of interpretation task required participants to read two Latin sentences and decide which one accurately described an onscreen photograph. The final aural interpretation task required participants to listen to a Latin sentence and decide whether it accurately described an onscreen photograph.

Target sentences consisted of two animate nouns and a transitive verb; word order of the sentences varied throughout the practice session. Feedback was provided for both correct and incorrect answers and appeared onscreen immediately after a participant made an answer choice. As noted earlier, feedback was more explicit in two treatments (+EF) and less explicit in the other two (–EF). More explicit feedback indicated whether an answer was correct and included item-specific metalinguistic information designed to reinforce rules of Latin morphosyntax. (See Appendix, Figure 1.2 in the online Supporting Information.) Less explicit feedback indicated only whether an answer was correct or incorrect. Feedback remained onscreen for 5 seconds before the next practice item was presented.

**Latin Test Battery**

Four Latin language tests were included in the test battery to enable examination of a variety of aspects of initial language learning. Two interpretation tests—one written and one aural—were similar in format to the practice tasks completed during treatment (see Appendix, Figure 1.2 in the online Supporting Information) and performance would thus be indicative of how successfully participants applied what they had learned during treatment in a familiar
context. A grammaticality judgment test was included as a different approach to assessing the degree to which form-meaning connections were made during interaction with the treatment. For this test, participants read a Latin sentence and decided whether it was well formed or not. Finally, a production test was included in order to assess to what degree knowledge of Latin morphosyntax acquired through interaction with the input-based treatment could be transferred to a guided production task. On this test, participants constructed Latin sentences to describe an onscreen photograph. They were provided with noun and verb stems as well as unlabeled endings, and they used the mouse to combine these elements to form a sentence. (See Appendix, Figure 1.3 in the online Supporting Information.) The three word stems were presented onscreen in random order to avoid biasing a particular word order in production.

As with practice sentences included in instructional treatment, critical items on the four language tests consisted of two animate nouns and a transitive verb and assessed participants’ ability to correctly assign thematic roles to Latin nouns. Distractor items containing just one noun and a verb were included on the production test, and NVN distractors that focused on lexicon rather than morphosyntax were included on the interpretation and grammaticality judgment tests. Three versions of the test battery were created and participants were assigned at random to complete a different version as their pretest, immediate posttest, and delayed posttest. The order of completion of the interpretation and grammaticality judgment tests was randomized, but all participants completed the production test last. Of the 75 items included in a test battery, 46 were critical items and 29 were distractors. The layout of interpretation test items was the same as for practice items, but in order to avoid the 50% chance of a dichotomous choice design, a third “I don’t know” option was included. Participants answered via key presses and completion of the tests was self-paced.

Scoring
There were 12 critical items and 8 distractors on the interpretation and grammaticality judgment tests. Participants earned one point for each correct response to a critical item for a total possible score of 12 on each of these tests. There were 10 critical items and 5 distractors on the written production test. For each critical item, participants earned one point each for correctly marking both nominative and accusative case on the two nouns, number of both nouns, and verb agreement. Partial credit was not awarded, so, for example, if just one of an item’s two nouns was marked correctly for case, the item was scored 0 for case. Following these criteria, each critical production item had a total possible score of 3, making 30 the maximum score possible on the production test.
Results

This study investigated the following two research questions: First, does prepractice grammar explanation influence initial learning of Latin morphosyntax associated with thematic role assignment? Second, does explicit corrective feedback during practice influence initial language learning? Initial language learning was operationalized as follows: (a) overall accuracy on tests of written interpretation, aural interpretation, grammaticality judgment, and written production; (b) accuracy of sentence interpretation when different combinations of morphosyntactic cues—verb agreement, SVO word order, and noun case morphology—were available to guide assignment of thematic roles; and (c) patterns of use of these three cues in written production.3

To investigate our research questions, we conducted 3 (Time) × 4 (Treatment) repeated-measures ANOVAs. Partial eta squared ($\eta^2$) is reported as an estimation of effect size for ANOVAs; according to Green and Salkind (2005), values of partial $\eta^2$ in the .01-.05 range indicate small effect sizes, .06-.13 indicate medium effect sizes, and values greater than or equal to .14 indicate large effect sizes. The standardized effect size, $d$, is reported for $t$-tests; according to Green and Salkind, positive or negative $d$ values of 0.2, 0.5 and 0.8 indicate small, medium, and large effect sizes, respectively. Finally, except where noted, observed power for all ANOVA results fell within the acceptable range of 0.8 to 1.0.

Overall Test Performance

Descriptive statistics for overall performance on the four Latin tests are displayed in Table 2. As the table shows, there was posttreatment improvement across the sample in overall performance on all measures, with partial regression to the mean in most cases at the delayed test 3 weeks after treatment.

Pretest scores were first submitted to one-way ANOVAs and results revealed a significant between-groups difference in performance on the written production pretest, $F(3, 55) = 4.15, p = .01$, partial $\eta^2 = .19$. Post hoc Bonferroni tests revealed that pretest scores in the $+GE+EF$ group were significantly lower than those in the $+GE–EF$ group. Pretest scores were thus entered as a covariate in subsequent analyses of production test performance. There was no other significant between-groups difference in overall pretest performance on the Latin tests, so differences in groups’ posttest and delayed test scores (apart from those in written production) may be attributed to effects of treatment.

Next, separate repeated-measures ANOVAs were conducted for each of the four Latin tests. Results revealed significant main effects for Time on accuracy on all tests, written interpretation, $F(2, 122) = 38.86, p < .0001$, partial
Table 2  Latin Test Scores by Treatment Group

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>n</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written interpretion (Max = 12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>18</td>
<td>6.5</td>
<td>10.2</td>
<td>8.8</td>
</tr>
<tr>
<td>-GE+EF</td>
<td>20</td>
<td>6.9</td>
<td>8.5</td>
<td>8.0</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>16</td>
<td>6.2</td>
<td>10.0</td>
<td>8.9</td>
</tr>
<tr>
<td>-GE–EF</td>
<td>11</td>
<td>6.1</td>
<td>8.2</td>
<td>8.0</td>
</tr>
<tr>
<td>Aural interpretion (Max = 12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>17</td>
<td>5.6</td>
<td>9.9</td>
<td>6.9</td>
</tr>
<tr>
<td>-GE+EF</td>
<td>20</td>
<td>6.1</td>
<td>9.1</td>
<td>7.1</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>15</td>
<td>5.6</td>
<td>9.6</td>
<td>7.7</td>
</tr>
<tr>
<td>-GE–EF</td>
<td>11</td>
<td>4.9</td>
<td>7.7</td>
<td>7.5</td>
</tr>
<tr>
<td>Grammaticality judgment (Max = 12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>18</td>
<td>5.4</td>
<td>8.2</td>
<td>6.2</td>
</tr>
<tr>
<td>-GE+EF</td>
<td>20</td>
<td>5.6</td>
<td>6.6</td>
<td>6.7</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>16</td>
<td>4.9</td>
<td>9.2</td>
<td>6.3</td>
</tr>
<tr>
<td>-GE–EF</td>
<td>11</td>
<td>5.1</td>
<td>6.6</td>
<td>6.4</td>
</tr>
<tr>
<td>Written production (Max = 30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>18</td>
<td>12.3</td>
<td>24.3</td>
<td>21.2</td>
</tr>
<tr>
<td>-GE+EF</td>
<td>16</td>
<td>15.4</td>
<td>17.8</td>
<td>18.3</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>15</td>
<td>16.9</td>
<td>24.7</td>
<td>19.3</td>
</tr>
<tr>
<td>-GE–EF</td>
<td>10</td>
<td>16.7</td>
<td>19.0</td>
<td>19.7</td>
</tr>
</tbody>
</table>

Note. Ns for the 4 tests differ due to missing data for some participants; means are given with SD in parentheses.

$\eta^2 = .39$, aural interpretation, $F(1.8, 105.4) = 58.21, p < .0001$, partial $\eta^2 = .50$, grammaticality judgment, $F(2, 122) = 25.60, p < .0001$, partial $\eta^2 = .30$, and written production, $F(2, 108) = 32.17, p < .0001$, partial $\eta^2 = .37$. Results of post hoc Bonferroni tests\(^4\) showed that the main effects for Time in written and aural interpretation were due to significant improvements in accuracy from pretest to both posttest (mean difference = 2.78, $p < .0001$ for written interpretation; mean difference = 3.48, $p < .0001$ for aural interpretation) and delayed test (mean difference = 1.96, $p < .0001$ for written interpretation; mean difference = 1.78, $p < .0001$ for aural interpretation), meaning that significant gains that resulted from interaction with any of the four treatments were still evident at least 3 weeks later.

The main effects for Time on the grammaticality judgment and written production tests as well as a main effect for Treatment observed for written
production, \( F(3, 54) = 4.89, p < .01, \) partial \( \eta^2 = .21 \), must be viewed in light of significant Time \( \times \) Treatment interactions, \( F(6, 122) = 3.53, p < .01, \) partial \( \eta^2 = .15 \) for grammaticality judgment and \( F(6, 108) = 5.39, p < .0001, \) partial \( \eta^2 = .23 \) for written production. The interaction for grammaticality judgment is illustrated in Figure 2. As the figure shows, the +GE+EF and +GE–EF groups improved their scores from pretest to posttest and results of paired-samples \( t \)-tests indicated that improvement in these two groups was statistically significant, \( t(17) = 4.70, p < .0001, d = 1.10 \) and \( t(15) = 6.51, p < .0001, d = 1.63 \), respectively. However, significant losses in accuracy in these groups from posttest to delayed test, \( t(17) = -4.01, p = .001, d = -0.95, \) and \( t(15) = -3.37, p = .004, d = -0.84, \) rendered any pretest-to-delayed test improvement nonsignificant. In contrast, the –GE+EF’s grammaticality judgment test scores did not change significantly as a result of instructional treatment. Finally, participants in the –GE–EF group improved their grammaticality judgment scores significantly from pretest to delayed test, \( t(10) = 3.55, p = .005, d = 1.07. \)

The significant Time \( \times \) Treatment interaction for written production is illustrated in Figure 3. As the figure shows, the +GE+EF and +GE–EF groups’

---

**Figure 2** Grammaticality judgment by treatment group.
performed on the written production test followed a pattern similar to their grammaticality judgment test performance, with significantly improved scores from pretest to posttest, $t(17) = 7.27, p < .0001, d = 1.72$, and $t(14) = 5.29, p < .0001, d = 1.37$, respectively, and significant losses in accuracy from posttest to delayed test, $t(17) = -2.68, p = .016, d = -0.63$ and $t(14) = -4.70, p < .0001, d = -1.22$. Nevertheless, contrary to their grammaticality judgment results, improvement in written production scores in the +GE+EF group from pretest to delayed test was statistically significant, $t(17) = 5.77, p < .0001, d = 1.36$. Improved accuracy in overall written production from pretest to delayed test in the –GE+EF group approached statistical significance, $t(15) = 2.64, p = .018, d = 0.66$. There was no significant posttreatment change in written production accuracy in the –GE–EF group.

Summarizing the results for overall Latin test performance in the context of our two research questions, we found that although all treatment groups exhibited significant and sustained improvement in both written and aural interpretation, the absence of a significant Time × Treatment interaction indicates no independent or combined benefit for either prepractice explanation or metalinguistic feedback.
For grammaticality judgment, a significant Time × Treatment interaction and post hoc t-tests indicate that prepractice explanation had an immediate positive effect, but this effect was not sustained to the delayed test in either the +GE+EF or the +GE–EF group. In contrast, the group that received neither prepractice explanation nor metalinguistic feedback (–GE–EF) exhibited significant and sustained improvement in the area of grammaticality judgment. No independent benefit for metalinguistic feedback obtained.

Finally, for written production, a significant Time × Treatment interaction and post hoc t-tests show that prepractice explanation promoted immediate gains in productive abilities but that sustained improvement required a combination of prepractice explanation and metalinguistic feedback.

Interpretation by Sentence Type
In order to investigate processing accuracy for sentences in which different combinations of cues to thematic role assignment were available, interpretation accuracy for the following three sentence types was examined: SVO items, in which SVO word order was always available and the verb agreement cue was available approximately half of the time; AGR items, in which the SVO word order cue was not available but the verb agreement cue was always available; CASE items, in which neither the SVO order nor verb agreement cue was available. The noun case morphology cue was always available in all three sentence types, making it the most highly reliable cue to assigning thematic roles.

Accuracy for each of these sentence types was calculated combining items from both the written and aural interpretation tests. Descriptive statistics are shown in Table 3. Overall, interpretation accuracy for all three sentence types showed marked posttreatment improvement with partial regression to the mean at the delayed test in the majority of cases.

Pretest scores were submitted to one-way ANOVAs, with results revealing no significant between-groups difference prior to instructional treatment (all $F$s < 1).

Next, separate repeated-measures ANOVAs were performed on interpretation accuracy data for each of the three sentence types. There was no significant result for accuracy of interpretation of SVO items, which was already above 80% at pretest, indicating a ceiling effect. There was a significant main effect for Time on interpretation accuracy for AGR items, $F(2, 114) = 103.13$, $p < .0001$, partial $\eta^2 = .64$, and post hoc Bonferronis showed significant improvement in accuracy on this sentence type from pretest to both posttest (mean difference = 50.83, $p < .0001$) and delayed test (mean difference = 33.91,
Table 3 Interpretation accuracy by sentence type and treatment group

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>+GE+EF</td>
<td>84.5(24.0)</td>
<td>92.9(14.0)</td>
<td>88.4(13.6)</td>
</tr>
<tr>
<td>–GE+EF</td>
<td>88.5(16.3)</td>
<td>78.5(26.4)</td>
<td>82.4(20.5)</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>84.4(14.8)</td>
<td>92.6(13.6)</td>
<td>90.0(13.9)</td>
</tr>
<tr>
<td>–GE–EF</td>
<td>81.0(16.1)</td>
<td>89.4(20.6)</td>
<td>90.9(13.8)</td>
</tr>
<tr>
<td>SVO items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>23.5(16.3)</td>
<td>85.4(17.0)</td>
<td>60.2(28.2)</td>
</tr>
<tr>
<td>–GE+EF</td>
<td>28.1(18.1)</td>
<td>72.1(20.5)</td>
<td>59.5(24.0)</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>27.0(19.7)</td>
<td>82.5(23.9)</td>
<td>64.4(24.5)</td>
</tr>
<tr>
<td>–GE–EF</td>
<td>26.8(12.6)</td>
<td>70.5(18.4)</td>
<td>58.2(29.4)</td>
</tr>
<tr>
<td>AGR items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>55.3(18.1)</td>
<td>78.8(19.7)</td>
<td>49.4(16.0)</td>
</tr>
<tr>
<td>–GE+EF</td>
<td>55.6(18.5)</td>
<td>68.9(20.8)</td>
<td>53.9(20.6)</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>49.3(17.9)</td>
<td>79.3(19.5)</td>
<td>62.7(24.9)</td>
</tr>
<tr>
<td>–GE–EF</td>
<td>40.9(18.1)</td>
<td>50.9(15.8)</td>
<td>56.4(11.2)</td>
</tr>
<tr>
<td>CASE items</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Group sample sizes: +GE+EF = 17, –GE+EF = 18, +GE–EF = 15, –GE–EF = 11; means are given with SD in parentheses; scores are expressed as proportions due to uneven numbers of items of each sentence type on the three versions of the tests.

There was no significant main effect for Treatment and no Time × Treatment interaction for accurate interpretation of AGR items.

There were main effects for both Time and Treatment for accuracy on CASE items, $F(1.8, 101.6) = 16.71, p < .0001$, partial $\eta^2 = .23$ and $F(3, 57) = 3.74, p = .016$, partial $\eta^2 = .17$, respectively, as well as a significant Time × Treatment interaction, $F(5.4, 101.6) = 2.68, p = .023$, partial $\eta^2 = .12$. Results of paired-samples t-tests indicated that the +GE+EF and +GE–EF groups made significant pretest-to-posttest improvements in accurate interpretation of CASE items, $t(16) = 4.19, p = .001, d = 1.02$ and $t(14) = 3.67, p = .003, d = 0.95$, respectively. Results additionally indicated that the +GE+EF and –GE+EF groups experienced significant losses in accuracy from posttest to delayed test, $t(16) = –6.93, p < .0001, d = –1.68$ and $t(18) = –3.43, p = .003, d = –0.79$, respectively. No treatment group’s improved accuracy of interpretation of CASE items was maintained to the delayed test session and the –GE–EF group’s performance on this measure did not change significantly as a result of treatment.

Summarizing these results in the context of our research questions, a significant Time × Treatment interaction and post hoc t-tests indicate that participants
who received prepractice explanation improved significantly more than other participants in the ability to accurately interpret sentences in which only the new case morphology cue was available to guide assignment of thematic roles. However, this improvement was not maintained to the delayed test. Results do not suggest any benefit for metalinguistic feedback.

Cue Use in Written Production

In order to investigate how participants chose to mark thematic roles, we coded production test responses according to which of the three cues they used, and the number of times (Max = 10) they used each cue at each test session was tallied. Descriptive statistics are shown in Table 4. Overall patterns of performance showed reduced frequency after treatment in use of SVO word order along with increased accurate use of both verb agreement and case morphology cues. There was partial regression to the mean in accurate use of case morphology at the delayed test session.

One-way ANOVAs were conducted to determine whether groups performed differently at pretest. The result for accurate marking of case morphology

<table>
<thead>
<tr>
<th>Treatment group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVO word order</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+GE+EF</td>
<td>7.2 (3.1)</td>
<td>5.7 (2.5)</td>
<td>6.1 (3.4)</td>
</tr>
<tr>
<td>-GE+EF</td>
<td>7.6 (2.7)</td>
<td>4.8 (3.4)</td>
<td>3.4 (2.4)</td>
</tr>
<tr>
<td>+GE–EF</td>
<td>8.1 (3.0)</td>
<td>8.2 (2.6)</td>
<td>6.7 (3.4)</td>
</tr>
<tr>
<td>-GE–EF</td>
<td>6.1 (3.2)</td>
<td>4.9 (3.9)</td>
<td>6.0 (3.7)</td>
</tr>
</tbody>
</table>

| Verb agreement  |         |          |         |
| +GE+EF          | 6.2 (2.0) | 9.8 (0.4) | 9.3 (1.0) |
| -GE+EF          | 8.1 (2.3) | 7.8 (2.2) | 8.6 (2.0) |
| +GE–EF          | 7.4 (2.4) | 9.5 (1.1) | 8.7 (1.6) |
| -GE–EF          | 8.0 (2.1) | 9.0 (1.3) | 9.1 (1.3) |

| Noun case morphology |         |          |         |
| +GE+EF              | 1.1 (1.2) | 7.1 (3.0) | 4.7 (3.0) |
| -GE+EF              | 1.8 (0.9) | 4.3 (3.1) | 3.8 (2.2) |
| +GE–EF              | 2.5 (1.6) | 7.5 (3.3) | 4.9 (2.9) |
| -GE–EF              | 2.6 (1.8) | 2.4 (1.7) | 2.8 (1.9) |

Note. Group sample sizes: +GE+EF = 18, -GE+EF = 16, +GE–EF = 15, -GE–EF = 10; means are given with SD in parentheses; Max = 10 for all measures; given that scoring one point for noun case morphology required accurate marking of two morphemes, a score of 2.5 represents chance performance on that measure.
was significant, $F(3, 55) = 4.18, p = .01$, partial $\eta^2 = .19$, due to less accurate pretreatment marking of that cue among participants in the +GE+EF treatment group compared with participants in the +GE–EF and –GE–EF groups. Pretest performance was therefore entered as a covariate in subsequent analyses.

Next, repeated-measures ANOVAs were conducted for each of the three cues. Results revealed significant main effects for both Time, $F(2, 110) = 6.14, p < .01$, partial $\eta^2 = .10$, and Treatment, $F(3, 55) = 3.48, p = .022$, partial $\eta^2 = .16$, on marking of SVO word order. At 0.75, observed power was somewhat low for the main effect for Treatment. Results of post hoc Bonferroni tests indicated that participants marked SVO word order significantly less at both posttest (mean difference = –1.33, $p = .011$) and delayed test (mean difference = 1.67, $p = .005$) than at pretest; the main effect for Treatment was accounted for by the fact that the –GE+EF group marked SVO order significantly less frequently than the +GE–EF group.

The ANOVA results indicated a main effect for Time, $F(1.6, 88.2) = 21.43, p < .001$, partial $\eta^2 = .28$, as well as a significant Time $\times$ Treatment interaction, $F(4.8, 88.2) = 5.86, p < .0001$, partial $\eta^2 = .24$, for accurate use of verb agreement in production. Results of paired-samples $t$-tests revealed that participants in the +GE+EF group significantly improved their accurate use of verb agreement from pretest to both posttest, $t(17) = 7.67, p < .0001, d = 1.81$, and delayed test, $t(17) = 5.58, p < .0001, d = 1.32$, whereas those in the +GE–EF group improved in accurate use of agreement from pretest to posttest only, $t(14) = 3.26, p = .006, d = 0.84$. Among those in the –GE+EF and –GE–EF groups, accurate use of the verb agreement cue did not change significantly as a result of treatment.

Significant main effects for Time, $F(2, 108) = 64.59, p < .0001$, partial $\eta^2 = .55$, and Treatment, $F(3, 54) = 6.20, p = .001$, partial $\eta^2 = .26$, as well as a Time $\times$ Treatment interaction, $F(6, 108) = 4.63, p < .0001$, partial $\eta^2 = .21$, were also revealed for accurate use of noun case morphology. Results of paired-samples $t$-tests showed that the +GE+EF and –GE+EF groups significantly improved accurate use of noun case morphology from pretest to posttest, $t(17) = 6.72, p < .0001, d = 1.58$ and $t(15) = 2.94, p = .01, d = 0.74$, respectively, and from pretest to delayed test, $t(17) = 4.41, p < .0001, d = 1.04$ and $t(15) = 3.35, p = .004, d = .84$, respectively, whereas the +GE–EF group made significant improvements in accurate use of case morphology from pretest to posttest only, $t(14) = 4.82, p < .0001, d = 1.24$. Finally, accurate use of case morphology among those in the –GE–EF group did not change significantly as a result of instructional treatment.
To sum up in the context of our research questions, ANOVA results indicate no independent advantage for prepractice grammar explanation as provided in our computer-based language lesson. The two groups that received prepractice grammar explanation made significant pretest-to-posttest improvements in overall written production accuracy, accurate use of noun case morphology in production, and interpretation of sentences in which case morphology was the only available cue to guide thematic role assignment. However, the prepractice explanation was not sufficient for successful retention of these improvements, as evidenced by a lack of statistically significant differences between pretest and delayed test scores except in the \( +GE+EF \) group, which exhibited significant improvements from pretest to delayed test on the production measures. Given that this group’s grammar explanation was coupled with metalinguistic feedback during practice, we cannot conclude that prepractice grammar explanation alone facilitated learning.

In contrast, our results suggest that explicit metalinguistic feedback provided during practice was beneficial, although the benefits appeared to be limited to accurate production of Latin case morphology. This is not a trivial result, however, given that of the three morphosyntactic cues involved, Latin case morphology was entirely new for the participants in our study.

Discussion and Conclusion

In this study we investigated participants’ initial learning of Latin with computer-delivered instructional treatments that differed in the amount and timing of grammar explanation provided. We found significant improvement across the sample in the ability to interpret written and aural Latin sentences, and gains in these areas were maintained for a period of at least 3 weeks. Closer examination of interpretation test performance further revealed that all treatment groups improved significantly in their ability to interpret AGR items (i.e., sentences not presented in SVO word order but that had both verb agreement and case morphology cues available to guide assignment of thematic roles). A final significant result observed across the sample was posttreatment reduction in use of the strong SVO word order cue to mark thematic roles in production. In light of these across-the-board results, we conclude with Sanz and Morgan-Short (2004) that task-essential practice coupled with feedback—the common denominators among the four instructional treatments—were sufficient to trigger sustained language learning in these areas.

Prepractice grammar explanation as it was provided in this study had no clear benefit for language learning over and above what participants garnered
from practice and feedback. The learning outcomes of the +GE groups differed from those of the –GE groups in that the former improved significantly at posttest in the areas of grammaticality judgment, accurate marking of verb agreement and noun case in written production, and interpretation of Latin sentences in which noun case morphology was the only available cue. However, what was learned was not retained over the 3-week interval between the immediate and delayed posttests, except by the +GE+EF group in accurate marking of agreement and noun case in production. Interestingly, the –GE+EF group patterned with the +GE+EF group in improving significantly from pretest to delayed test in accurate use of noun case morphology in production. These results suggest to us that metalinguistic feedback provided concurrently with practice—the treatment component common to the +GE+EF and –GE+EF groups—was a key factor in gains in our participants’ productive abilities.

Notice, however, that what the +EF groups learned about noun case morphology was reflected in sustained improvement in case marking in production but not in interpretation of sentences in which case morphology was the only cue available to guide assignment of thematic roles (i.e., CASE items). It is likely that the nature of the interpretation and production tests accounts for this difference. On the one hand, the interpretation tests required participants to infer meaning that was already encoded for them, but given that the availability of the three cues for assigning thematic roles was experimentally manipulated, this encoding varied from item to item with different combinations of converging and competing cues. When cues are in competition, the strength of misleading cues (e.g., word order and verb agreement in CASE items) is likely to overshadow (Ellis, 2006) a weaker, albeit more reliable, cue (e.g., case morphology) such that weaker cues are passed over during sentence processing. On the other hand, the production test required participants to encode meaning themselves by whatever means they chose. In this do-it-yourself context, the presence of competing cues would not be a stumbling block in the same way that it appeared to be with interpretation.

In contrast with significant losses in accuracy suffered between the immediate and delayed posttests for interpretation of CASE items, improvement was sustained across the sample for interpretation of AGR items, in which agreement and case morphology cues converged on thematic role assignment while word order competed with those two cues. In other words, for this sentence type, participants seem to have selectively ignored the strong but misleading word order cue in favor of the more reliable—although presumably weaker—agreement and case morphology cues. It is likely that, similar to what Henry et al. (2009) found, the relative simplicity of verb agreement as it was presented...
in this study (i.e., exclusive use of third-person singular or plural in the input) made encoding and retention of this cue fairly straightforward. Nevertheless, the fact that our participants maintained accuracy for the AGR sentence type 3 weeks after treatment suggests to us that they successfully reached the leading edge of the restructuring of cue hierarchies that characterizes learning under the CM. This incipient restructuring was reflected not only in maintained gains on interpretation of sentences using verb agreement cues but also in the across-the-board reduced frequency in marking of SVO word order in production, a notable result when one considers what a strong cue word order is for assigning semantic function to nouns in Spanish and English, our participants’ two languages.

Interestingly, only the –GE–EF group exhibited sustained improvement in the area of grammaticality judgment. The two +GE groups made significant improvements on this test from pretest to posttest, but improvement was not maintained over time as it was among those in the –GE–EF group. A possible explanation of this somewhat surprising result is that, as Rosa and O’Neill (1999) suggested, different types of instruction led to qualitatively different input processing, with more explicit instruction triggering concept-driven (top-down) processing and less explicit instruction promoting data-driven (bottom-up) processing. We conclude that data-driven processing in which participants were left maximally to their own devices was the best way to help them induce patterns and develop intuitions on their own, whereas providing metalinguistic information in the form of prepractice grammar explanation or more explicit feedback or both may have muddied the waters for the more intuitive type of language learning reflected in grammaticality judgments.

Another noteworthy result of the present study is that the learning that occurred as a result of our input-based treatments was successfully applied in production. As Wong (2004) noted, successful productive use of language after instruction that did not include any output practice provides evidence of nonnegligible interlanguage change. It is within this context of improvement in productive abilities that the +GE+EF group fared better than all other treatment groups. This result suggests, in line with VanPatten and Oikkenon’s (1996), Wong’s, Farley’s (2004), and Rosa and Leow’s (2004) findings, that metalinguistic information provided as part of input-based instructional treatment facilitated gains in productive abilities. In all other areas of language learning considered in this study, however, our results are consistent with those of Benati (2004), Sanz and Morgan-Short (2004), and Morgan-Short et al. (2010) in that prepractice grammar explanation did not have any appreciable effect on learning over and above what participants derived from task-essential practice.
with feedback. With regard to questions of the explicitness of feedback, our findings depart slightly from Sanz and Morgan-Short’s. Although, similar to what they found, provision of metalinguistic feedback did not facilitate gains in interpretive abilities or in marking verb agreement in production, it was a contributing factor in promoting learning of noun case morphology, a particularly complex structure.

We found no clear long-term advantage for our prepractice grammar explanation in initial language learning as it was assessed in this study; however, it is not our intention to suggest that there is no place for preemptive instruction in language teaching. Our grammar explanation was highly structured and explicit but relatively brief, which leaves open the question of whether repeated exposure to grammar explanation over time facilitates nonprimary language learning. Existing evidence suggests otherwise, with the results of both Benati’s (2004) and Morgan-Short et al.’s (2010) studies, for example, indicating no advantage for repeated exposure to explicit instruction. However, given the methodological differences in these studies, their findings are not strictly comparable, so replication with longitudinal designs is needed.

The central findings of our study join previous work in suggesting little role for prepractice grammar explanation in the context of computer-assisted, nonprimary language instruction that includes task-essential practice. A role was identified, however, for provision of metalinguistic feedback in promoting learning of a novel and complex form. Our results align with those of other studies that have examined different areas of morphosyntax in a variety of languages and offer a useful springboard for continued research on the effects of instruction on nonprimary language learning.

Revised version accepted 7 January 2011

Notes

1 Language proficiency was self-reported, and we agree with an anonymous reviewer that lack of inclusion of independent proficiency measures represents a limitation.

2 A subset of participants \( n = 26 \) verbalized their thoughts during treatment, and as these data were not the focus of the present study, we do not report on them further. We conducted a series of ANOVAs to determine whether verbalization introduced an intervening variable, and results revealed no significant main effect for verbalization and no significant interaction with Time and/or Treatment (all \( ps > .05 \), with negligible to small effect sizes as indicated by values of partial \( \eta^2 \) between 0 and .06).
3 As one anonymous reviewer correctly noted, our operationalizations are not easily reconciled with the CM’s definition of learning as the restructuring of a cue hierarchy from that of a known language to that of the target language. The CM additionally assumes that restructuring takes amounts of both input and time that are beyond the capabilities of this study’s instructional treatment. It is for that reason that we refer to (and operationalize) initial language learning, which we take to be a definitive move in the direction of long-term language development but not synonymous with learning in the CM sense.

4 For all post hoc tests and paired-samples t-tests reported in the results, Bonferroni-corrected significance levels were set at $p = .017$ to adjust for multiple comparisons.

5 Although we did purposefully manipulate cue availability so that we could compare participants’ processing of sentences in which different combinations of cues were available, the resulting variety of structures can be considered to reflect that of Latin outside of the experimental context.

References


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure 1.1. Vocabulary lesson item.
Figure 1.2. Practice item with feedback.
Figure 1.3. Written production test item.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.