Compositional production in Spanish second language conjugation

NORA PRESSON, NURIA SAGARRA, BRIAN MACWHINNEY and JOHN KOWALSKI

Compositional production in Spanish second language conjugation*

NORA PRESSON
Department of Psychology and Program in Interdisciplinary Education Research, Carnegie Mellon University & Learning Research and Development Center, University of Pittsburgh

NURIA SAGARRA
Department of Spanish and Portuguese, Rutgers University

BRIAN MACWHINNEY
Department of Psychology and Program in Interdisciplinary Education Research & Department of Modern Languages, Carnegie Mellon University

JOHN KOWALSKI
Department of Psychology and Program in Interdisciplinary Education Research, Carnegie Mellon University

Dual-route models of second language (L2) morphology (Clahsen & Felser, 2006; Ullman, 2004) argue that adult L2 learners rely on full-form retrieval, and therefore cannot use combination to produce inflected forms. We tested this prediction with learning of Spanish verb conjugations. Beginning (Experiment 1) and intermediate (Experiment 2) learners (total N = 816) completed 80–90 minutes of web-based training, conjugating regular and subregular verbs in present and preterite tense. Tests of generalization items showed that training led to substantial improvement, equally for metalinguistic and analogical feedback. Comparison with an untrained group showed that gains were maintained 18 weeks after training. In contrast with dual-route model predictions, pre-test accuracy and learning gains were strongly predicted by conjugation pattern, showing that full-form retrieval was insufficient to explain learner performance. Results indicate that adult L2 learners apply compositional analysis, and that conjugation patterns are learned on the basis of their relative cue validity.

Adults have difficulty in processing (Sagarra, 2008; Ellis & Sagarra, 2010) and producing (Bardovi-Harlig, 2000; Clahsen & Felser, 2006) second language (L2) verbal morphology. Adult L2 learners initially use lexical cues such as temporal adverbs (e.g., *yesterday*, *today*) and overt subject pronouns (e.g., *I*, *she*; Lee, Cadierno, Glass & VanPatten, 1997; Rossomondo, 2003), only later attending to verbal morphology (e.g., *-s* in *she walks*, *-ed* in *he walked*; Dietrich, Klein & Noyau, 1995; Ellis & Sagarra, 2010; Klein, 1994; Starren, 2001). Even experienced learners still produce uninflected and incorrect forms and detect fewer errors than do native speakers (Clahsen & Felser, 2006; Prévost & White, 2000).

Most adults begin learning an L2 in a classroom, where input is quantitatively and qualitatively different from the input that native speakers receive (Ellis, 2002; Gor & Chernigovskaya, 2004). Quantitatively, classroom learners without an immersion experience receive less input and attend less to verbal morphology than do study abroad learners of the same proficiency level (LaBrozzi, 2009). Qualitatively, classroom input is characterized by the overuse and underuse of various L2 forms (Goddall, 2008; Santilli, 1996; Sanz, 1999; Schinke-Llano, 1986). For example, native Spanish-speaking teachers use more overt subject pronouns when talking with their students in class than when talking with other native Spanish speakers outside of class (Dracos, 2010). Because teachers' overuse of overt subject pronouns may hinder learners' attention to verbal morphology, it is important to find ways to help learners focus on verbal morphology. In order to do this, we must first consider how verbal morphology is processed in first language (L1) and L2.

Theoretical approaches to morphological processing

First-language processing of verbal morphology has been explained with dual-route, single-route, and hybrid

Address for correspondence:
Nora Presson, Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15213, USA
presson@pitt.edu

Table 1. *Predictions of competing processing models.*

	Dual-route		Single-route		Hybrid	
	L1	L2	L1	L2	L1	L2
Can use composition for fully regular default (rule-based generalization)?	Yes[a]	No	No	No[d]	Yes[a]	Yes[b]
Can use composition for non-default regular or subregular (rule-based generalization from minor rules)?	No	No	No	No	Yes	Yes
Can use retrieval of full-form chunks? (predicts token frequency effect – form-based generalization)	Yes	Yes	Yes	Yes	Yes	Yes
Predict a further effect of type frequency	No[c]	No	No[d]	No	Yes	Yes

[a] Procedural.
[b] Initially explicit, proceduralized with practice.
[c] Although defaultness is highly correlated with type frequency, dual-route models attribute the combination/retrieval dissociation to defaultness.
[d] Single-route models can be used to show emergent type frequency effects using phonological similarity (i.e., gang effects).

models (Table 1 shows an overview of contrasts among the three models). In this paper, we argue that the hybrid models provide the best account of data on early L2 learning of verbal morphology. Dual-route models (Anderson, 1992; Clahsen, 2006; Pinker, 1999; Stump, 2001; Ullman, 2004) propose a qualitative, neurologically grounded, mechanistic dissociation between a fully regular, productive default pattern (processed with procedural combination) and all other conjugational patterns (processed with declarative retrieval of inflected forms). In these models, the procedural route for L1 relies on automatic composition of regular default verbs through a combination of base and default inflections. The declarative route relies on full-form, rote retrieval of inflected forms, and can be used to produce both regular and irregular forms. For Spanish conjugation, the default is regular, first conjugation -*ar* verbs, such as *cantar* "to sing" or *bailar* "to dance". Evidence for the dissociation between the two routes in L1 comes from English, Spanish, German, and Italian, from behavioral and neuroimaging differences between regular default verbs and other verbs, the use of regular patterns as default markings, and generalization of patterns to nonce verbs (e.g., Linares, Rodríguez-Fornells & Clahsen, 2006; Marcus, Brinkmann, Clahsen, Wiese & Pinker, 1995; Rodríguez-Fornells, Münte & Clahsen, 2002; Say & Clahsen, 2002). For example, Spanish-speaking children make overregularization errors and treat non-*ar* verbs like -*ar* verbs (Clahsen, Aveledo & Roca, 2002). Also, native Italian speakers generalize the default first conjugation to nonce words more frequently than they do the non-default second or third conjugations (Say & Clahsen, 2002).

In contrast, L1 single-route associative models explain differences in production of regular and predictably irregular (SUBREGULAR) patterns, as well as idiosyncratic fully irregular items, in terms of properties of the input and relations among forms, all within a single integrated system without a fundamental dissociation in processing

mechanism between regularity and irregularity (Bybee, 1995, 2007; Bybee & McClelland, 2005; Eddington, 2000; Hahn & Nakisa, 2000; McClelland & Patterson, 2002; Seidenberg & McClelland, 1989). Unlike dual-route models, single-route models do not postulate compositional production of inflected forms.

Finally, L1 hybrid models accept the dual-route contrast between compositional and associative processing mechanisms, but include a role for minor rule patterns in the compositional route. Within that route, minor rule patterns compete with regular (default and non-default) patterns in terms of frequency and cue reliability (Ellis, 2002; Gor & Cook, 2010; MacWhinney, 1978, 2007; Nicoladis & Paradis, 2012; Stemberger & MacWhinney, 1986; Tkachenko & Chernigovskaya, 2010), and compositional patterns compete with associative retrieval. One important implication of comparing these models is for how verbs are stored in memory. If all verbs are processed by associative full-form retrieval, then it follows that each form of each verb is represented separately in memory. In contrast, if procedural composition of stem and affix generates inflected forms, then conjugations could be formed by storing the root alone along with multiple affixes associated with different forms.

Unlike the dual-route model, the hybrid approach for L1 also envisions ways in which compositional processing can involve specific alternations of phonological patterns, such as the Spanish stem allomorphic variation in *colgar*, which is conjugated as *cuelgo* in the *yo* form instead of *colgo, as in the regular pattern. This predictable irregularity, or *subregularity*, can be characterized as an -*o* → -*ue* stem transformation. In the hybrid account, therefore, compositional processing could involve more than just the concatenation of morphemes (MacWhinney, 1978).

Let us now consider how these models apply to L2 learning. Current dual-route models of L2 learning

(Clahsen & Felser, 2006; Ullman, 2004) claim that adult L2 learners show a deficit in compositional processing, leading them to depend on full-form storage for all inflected forms. In this regard, dual-route models of L2 learning converge with single-route models in terms of the emphasis on the role of retrieval. Because of this convergence, we can treat dual-route and single-route models as equivalent in terms of the predictions they make for L2 learning. The compositional deficit postulated by the dual-route model is held to be the central difference between native speakers and learners. If L2 Spanish learners do have a compositional deficit, then three predictions follow for acquisition of Spanish verb morphology.

First, regular and irregular Spanish verbs should be equally difficult to memorize as whole words. Bowden, Gelfand, Sanz and Ullman (2010) asked native speakers and advanced learners of Spanish to produce a limited item set (1st person present and imperfect forms). They found that native speakers showed frequency effects for all verbs except regular *-ar* verbs. However, advanced learners showed frequency effects for all verbs, including regular *-ar* verbs, though it must be specified that because of the different circumstances of language learning and differences in input exposure, frequency in L2 learners may not fully correspond to frequency in L1 native speakers (e.g., Gollan, Montoya, Cera & Sandoval, 2008). Bowden and colleagues concluded that regular *-ar* verbs are only composed in native speakers, because learners are reliant on full-form declarative representations.

Second, if all Spanish verbs are stored as whole words, there should be no difference in difficulty within a given subregular verb (that is, a verb with a predictable transformation pattern) for forms of that verb that do and do not require an irregular transformation relative to the regular pattern (see section "Target verbs" below for a description of the patterns used in the current study). That is, it should be irrelevant for subregular verbs whether or not a particular conjugated form matches the regular conjugation pattern. Each form must be separately memorized. Third, although Spanish has many more verbs in the default *-ar* conjugation than in non-default *-er* or *-ir* conjugations, the dual-route and single-route accounts for L2 do not expect such differences in type frequency to affect L2 learning.

Hybrid models of L1 and L2 processing do not assume any general compositional deficit, and therefore make different predictions than dual- and single-route models for L2 learning. These models predict that learners will acquire combinatorial patterns in the relative order of their cue validity (predictive value of the pattern; MacWhinney, 2007). First, learners should find regular verbs easier to produce than subregular or irregular verbs. Second, within a subregular verb, particular forms that match the regular pattern should be easier to produce than those that do not.

However, because of competition with the transformed forms in the paradigm for that verb, these forms should still be more difficult for learners than corresponding forms of fully regular verbs. Third, because of its high type frequency, the *-ar* conjugation should be easy for learners. Table 1 summarizes the contrasting predictions of these three models.

Effects of metalinguistic and analogical feedback on L2 learning

Having reviewed models of L2 morphological processing, we now turn to the ways in which instructional treatments can improve that processing. One common instructional treatment involves the introduction of metalinguistic information through formal rules. For example, there is a general spelling rule in Spanish that *c-* changes to *qu-* when followed by either *-e* or *-i* (*banco/banqueta*, *saco/saquito*). For Spanish verbs ending in *-car* this rule forces a change of the *c-* to *qu-* when the ending begins with *-e* (e.g., *sacar → saqué*, but also *sacamos*). There is mounting evidence that providing metalinguistic information regarding rules of this type increases learners' attention to L2 morphology (see Norris & Ortega, 2000; Spada & Tomita, 2010, for meta-analyses). Metalinguistic information is associated with increased learning for many grammatical structures including English dative, French grammatical gender, and Spanish conditional (Nagata & Swisher, 1995; Presson & MacWhinney, in press; Rosa & Leow, 2004). Despite this evidence, feedback containing metalinguistic information is rarely present in classroom practice (Lyster & Ranta, 1997).

An alternative to presenting metalinguistic information involves presenting analogical information; that is, another example following the same pattern. For example, the first person singular preterite of *buscar* (*busqué*) is analogous to *sacar–saqué*. Learning by analogy is a process that is common to both L1 (Chan, Liven & Tomasello, 2009; Goldberg, 2006; Ninio, 1999) and L2 learning (Kilborn & Ito, 1989). Analogical feedback is relevant because high-frequency exemplars can boost category activation (MacWhinney, 2007, 2011), especially early in learning, which can improve the strength of category representation (e.g., Ellis, 2002). In the current study, we compare feedback with metalinguistic information to feedback with an analogical comparison.

The present study

The current study trained learners of Spanish in a task requiring the production of regular and subregular verbs, for forms of subregular verbs both with and without a transformation (that is, forms of subregular verbs that do or do not match the regular pattern). Training included

both default *-ar* and non-default *-er* and *-ir* verbs in present and preterite tenses. A longitudinal training intervention with immediate and delayed post-tests was used to test three basic research questions.

The first question is whether conjugation pattern (including tense, transformation status, and verb conjugation) systematically predicts production accuracy. We test this question by examining baseline performance before training (i.e., performance at pre-test). If learner accuracy is predicted by properties of the pattern (e.g., tense, conjugation, and regularity), and not merely by the individual inflected form token, it would indicate that full-form retrieval of each conjugated form (affected by token but not type frequency) is insufficient to fully explain conjugation performance, contrary to the predictions of dual- and single-route models. That is, dual-route compositional deficit and single-route models of morphological processing predict that full-form retrieval will fully explain learner performance. Thus, for the current study, dual- and single-route models predict that:

a) Training with feedback will improve learner performance equally for all trained tokens, but improvement will show only form-based and not rule-based generalization to test items.

b) Regular default (*-ar*) and non-default (non-*ar*) verbs will improve equally after training (when all three conjugations are presented with equal frequency in training).

c) Within a subregular verb, whether or not a specific form requires a transformation will not change its difficulty (because all forms are stored and retrieved as wholes). Subregular forms with and without transformations will improve equally.

In contrast, hybrid models predict the following:

a) Training with feedback will improve learner performance for both practiced individual tokens and conjugational patterns, therefore showing rule-based generalization as well as generalization from form similarity.

b) Regular default conjugation (*-ar*) verbs will show the highest baseline (pre-test) accuracy, and therefore the smallest level of improvement. Regular non-default conjugations (non-*ar* verbs) will show lower baseline accuracy and will improve more after practice.

c) Within the paradigm for subregular verbs, forms not requiring a transformation will show higher baseline accuracy than forms requiring a transformation. This effect will be distinguishable from simple overapplication of the regular pattern, because regular verbs will show even higher baseline accuracy.

In addition to these theoretical issues, we wanted to understand whether gains from focused production training could make a meaningful educational impact. To that end, we tested whether trained beginners would outperform a comparison group of their untrained classmates one semester after training. It must be noted that because the population tested here consists of classroom learners, all participants were exposed to similar classroom instruction during the delay interval; however, because only the trained group completed the additional experimental training, the comparison is still highly valuable.

Further, most studies on L2 morphological processing focus on advanced learners and cannot speak to L2 developmental patterns. To fill this gap, we tested beginning (Experiment 1) and intermediate (Experiment 2) learners, and compared the effect of training across the two experiments, testing whether beginning learners would benefit more from the intervention than intermediate learners, who presumably have more experience with conjugation.

The final research question tests the effect of two forms of feedback during training. Specifically, does training with metalinguistic feedback lead to greater improvement than training with analogical feedback? In addition to testing the prediction of an overall advantage of metalinguistic relative to analogical feedback, we also considered whether properties of the conjugational pattern would moderate the effectiveness of metalinguistic feedback. Specifically, we speculated that simpler rule feedback (e.g., for regular verbs) would show a larger advantage of metalinguistic feedback due to the simplicity of the rule statement.

Experiment 1

Method

Participants

One hundred and forty-four English native speakers enrolled in an intensive Spanish course offered across two semesters at a large North American university participated in exchange for extra credit: of those, 32 made up the untrained comparison group (tested in the second semester follow-up only). Four participants did not complete any training or post-tests, and were therefore excluded. The other 108 made up the main sample of trained participants: 52 were randomized to the metalinguistic feedback group, and 56 to the analogical feedback group. The sample came from five course sections: all sections followed the same syllabus and teaching methodology, used the same textbook, completed the same homework, and completed five hours of class per week. Participants were assigned to one of the interventions randomly (not from intact class sections). All participants included in the trained sample met the following criteria: completing the pre-test, at least one

training session, and at least one post-test. The final sample comprised 108 learners at pre-test, 94 for the immediate post-test, 92 for the delayed post-test, and 122 (90 trained, 32 untrained) for the next semester follow-up.

Procedure

Training and testing were delivered in a web-based system using JavaScript and Flash to log learner data to a central server. During the second week of class, participants completed the consent form, and the language background questionnaire (both administered online). During the third week of class, they were exposed to a 40-minute teacher-administered grammar explanation, and completed a 5-minute pre-test. Three weeks later, participants completed the first training session. There were three 30-minute training sessions conducted at three-week intervals. The three 5-minute post-tests were performed immediately after the final training session, one week after, and 18 weeks after (early in the next semester). Finally, learners completed a vocabulary test, in which they typed the English translation of testing verbs presented in Spanish infinitive form, to ensure that a lack of lexical knowledge did not affect test results. Overall, learners correctly translated an average of 45% (SD = 15%) of the words used in testing.[1] All tests were administered in class and proctored by teachers to increase the validity of the measure.

Materials
Target verbs
Spanish verbs are distributed into three conjugations (*-ar*, *-er*, *-ir*), and, according to compositional accounts, are formed by a root (stem) and a suffix that indicates tense, aspect, mood, person and number: [root + thematic vowel] + tense-aspect-mood suffix + agreement (person-number) suffix. Spanish grammar traditionally divides verbs into regular and irregular: the main difference lies in regular verbs having one root and irregular verbs at least two variants of the root, at least two suffixes for the same morphosyntactic information, or radically different forms. Irregular verbs can be further divided into "subregular" (one variant of the root different from the basic variant with a regular suffix, or the opposite, the basic variant of the root with an irregular suffix) and genuinely "irregular" (idiosyncratic or fully suppletive).

The current study tested verbs in all three conjugations, in present and preterite tense of the indicative mood,

because these are the first tenses covered in basic Spanish courses. We included all persons except the 2nd person plural because most Spanish dialects use the 3rd person plural form with both 2nd and the 3rd person plural subjects. We will focus on these tenses and persons when describing Spanish regular verbs. Spanish regular *-ar* verbs are formed by adding the suffixes *-o*, *-as*, *-a*, *-amos*, *-an* to the root (e.g., *hablar* "to talk" → *hablo* "I talk") for the present, and the suffixes *-é*, *-aste*, *-ó*, *-amos*, *-aron* to the root (e.g., *hablé* "I talked") for the preterite. According to Hualde, Olarrea, Escobar & Travis (2010), 90% of Spanish verbs belong to the *-ar* conjugation, and, in general, *-ar* verbs are more completely regular than *-er* or *-ir* verbs. Spanish regular *-er* and *-ir* verbs are formed by adding the suffixes *-o*, *-es*, *-e*, *-emos* for *-er* or *-imos* for *-ir*, *-en* to the root (e.g., *vivir* "to live" → *vivo* "I live") for the present, and the suffixes *-í*, *-iste*, *-ió*, *-imos*, *-ieron* to the root (e.g., *viví* "I lived") for the preterite.

Subregular and irregular verbs can have irregularities in the root, suffix, prosody, or a combination of the above. Irregularities in the root use the same inflectional endings as regular verbs of the same conjugation (Zollo, 1993) and comprise:

(i) INSERTION OF A VELAR OR CORONAL CONSONANT (*poner* "to put" → *pongo* "I put", *conducir* "to drive" → *conduzco* "I drive", *salir* "go go out" → *saldré* "I will go out"). Verbs of this type were not included in the current study.
(ii) DIPHTHONGIZATION: changes of /e/ to /ie/ (*cerrar* "to close" → *cierro* "I close"), /o/ to /ue/ (*dormir* "to sleep" → *duermo* "I sleep"), /u/ to /ue/ (*jugar* "to play" → *juego* "I play").
(iii) VOWEL RAISING (substitution of a semi-close vowel with a close vowel): changes of /e/ to /i/ (*vestir* "to dress" → *visto* "I dress"), /o/ to /u/ (*morir* "to die" → *murió* "(s)he died").

For (ii) and (iii), in the present indicative, the irregular configuration arises in all singular forms (e.g., *visto* "I dress", *vistes* "you dress", *viste* "(s)he dresses") and in the 3rd person plural (e.g., *visten* "they dress"), but not in the 1st person plural (e.g., *vestimos* "we dress") or 2nd person plural (e.g., *vestís* → "you (plural) dress"). In the preterite, the changes occur not in first person (*vestí* "I dressed"), but rather in both singular and plural of the 3rd person (e.g., *vistió* "(s)he dressed", *vistieron* "they dressed"). Because there is only one verb with the change of /u/ to /ue/ (*jugar* "to play": *juego* "I play", *jugamos* "we play"), this verb was not included in the study. Vowel-raising affects all *-ir* verbs whose infinitive has /e/ in the last syllable of the root.

Irregularities in the suffix include:

(i) Suffix different from the conjugation corresponding to the verb under consideration (*dar* "to give" →

---

[1] Translation accuracy was not used to exclude items from analysis. One learner was an outlier on the translation test, showing only 8% accuracy. Removing students in the bottom 10% of translation test scores did not change any statistical results. In addition, the untrained comparison group did not score not significantly better or worse on the translation test than the trained group either with or without eliminating low performers (*t*(114) and *t*(101) < 1, *p*s > .5).

diste "you gave", *ir* "to go" → *vas* "you go", *vamos* "we go"). Verbs of this type were not included in the current study.

(ii) Suffix different from that of any regular conjugations (*poder* → *pude* "I could", *pudo* "(s)he could").

Irregularities in the prosody show an accent scheme different from the norm (e.g., *poder* → *pude*, *pudo*). Mixed irregularities consist of an alternation in the root, the suffix and the accent scheme (*poner* "to put" → *puso* "(s)he put (past)").

In addition to these irregularities, some verbs require changes in orthography, while preserving a consistent phonology. These changes involve consonants, rather than vowels. For example, verbs ending in *-car* change *c-* to *qu-* whenever the regular ending begins in *-e*, because in Spanish the *c* in *ce* is soft, whereas the *c* in *ca*, *co*, or *cu* is hard (e.g., *saco* "I take out" and *saqué* "I took out"). Although these patterns may not be processed morphologically in native speakers, they are taught to learners as rule-based, predictable irregular transformations, making them a possible source of rule-based generalization.

The present study included regular verbs of the three conjugation types, subregular verbs with irregularities in the root (subregular stem change verbs: irregular root + regular affix), and subregular verbs with irregularities in the orthography (subregular spelling change verbs).

Treatment and testing materials

Trained participants were randomly assigned to one of two groups: metalinguistic feedback and analogical feedback. In addition, an untrained comparison group not enrolled in the experimental sections during training was tested at the next semester follow-up to measure the effect of maturation throughout the course. In tests and training, all learners read prompts like "*acertar* / present / *Yo* ___" and typed the conjugated verb. Feedback was provided during training but not during testing.

In training, when learners conjugated a verb correctly (in this case by typing *acierto*), they received the prompt "correct". For incorrect answers, learners saw the correct target. In addition, the analogical group received a model of a highly frequent verb of the same type (e.g., *acertar* → *acierto*/*defender* → *defiendo*) and the metalinguistic group received a short grammar explanation of the rule (e.g., "*acertar* → *acierto*/*e* → *ie* when ending is unstressed" for subregular verbs, or "*limpiar* → *limpio*/This verb is regular" for regular verbs). Grammar explanations were drawn from the learners' Spanish textbook (Castells, Guzmán, Lapuerta & García, 2002) and a verb handbook (Zollo, 1993) to make them similar to current educational materials and reduce extraneous cognitive load.

Verbs used in testing were not used in training, such that all test scores reflect generalization to untrained items.

Across testing and training, there were 60 regular verbs (half *-ar*, half *-er*/*-ir*), 48 stem-change subregular verbs with four patterns (*e* → *ie* like *defender* "to defend", *o* → *ue* like *encontrar* "to find", *e* → *i* like *repetir* "to repeat", and *e* → *ie* and *i* like *preferir* "to prefer"), and 52 spelling-change subregular verbs with five patterns (*c* → *qu* like *practicar* "to practice", *z* → *c* like *abrazar* "to hug", *g* → *gu* like *obligar* "to force to", *g* → *j* like *escoger* "to choose"). Subregular verb forms were selected such that half required a transformation and half did not, with the overall distribution of 33% regular verbs, 33% subregular verbs without transformation, and 33% subregular verbs with transformation. Verb conjugation was also distributed such that 33% of items were *-ar* verbs, 33% were *-er* verbs, and 33% were *-ir* verbs. Half of the regular, spelling-change subregular, and stem-change subregular verbs appeared in the present tense and half in the preterite tense. Finally, the director of the basic Spanish program (and two English learners of Spanish) looked at the verbs in the *Berlitz Spanish Verb Handbook* (Zollo, 1993) and labeled those that a student could be expected to encounter in class. They then determined the most common verbs to be used as models for the analogy training. Verbs used as models in analogy feedback were never presented as target items in training, or in pre- or post-tests. Therefore, both training groups conjugated the same set of verbs, though the analogical group saw an additional model verb in feedback.

Scoring

Accuracy was coded as a binary outcome (0 = incorrect, 1 = correct). Written accents were not considered in scoring. Any other deviation between the typed student answer and the target was scored as an error.

Results

Descriptive statistics

Descriptive statistics are reported for accuracy (average proportion correct). Because of the large sample size, all standard errors up to three-way interactions were .01 after rounding (range: .005–.014), unless otherwise noted (Table 2 shows the lowest-level cells of the design). Overall accuracy across all verb types and training conditions was .25 at pre-test, .53 at immediate post-test, and .44 at one-week delayed post-test. Note that, because learners were required to type the inflected form, it is unlikely that they could produce a correct answer by chance. The following semester (18 weeks after training), accuracy for trained participants was .49, compared to .41 for the untrained comparison group. Considering conjugation as a two-level variable (*-ar* and non-*ar*), accuracy was .50 for *-ar* verbs and .30 for non-*ar* verbs. When conjugation was taken as a three-level variable (*-ar*,

Table 2. *Average observed proportion correct (standard error) over time of testing broken down by Transformation Status, Tense, and Verb Conjugation for the 108 participants included in the final analysis in Experiment 1.*

	Regular verb				Subregular without transformation				Subregular with transformation			
	-ar		Non -ar		-ar		Non -ar		-ar		Non -ar	
	Present	Preterite	Present	Preterite	Present	Preterite	Present	Preterite	Present	Preterite	Present	Preterite
Pre-test		.28	.61	.14	.71	.28	.45	.18	.01	.02	.03	.01
		(.01)	(.02)	(.02)	(.02)	(.02)	(.03)	(.02)	(.01)	(.01)	(.01)	(.01)
Immediate	.89	.73	.78	.59	.84	.72	.69	.59	.13	.30	.33	.23
post-test	(.02)	(.02)	(.03)	(.03)	(.02)	(.03)	(.05)	(.04)	(.02)	(.03)	(.02)	(.03)
Delayed	.85	.66	.68	.40	.80	.68	.69	.45	.10	.19	.27	.16
post-test	(.02)	(.02)	(.04)	(.04)	(.03)	(.03)	(.05)	(.04)	(.01)	(.02)	(.02)	(.02)
Follow-up	.88	.61	.79	.38	.85	.63	.67	.35	.06	.19	.25	.09
	(.01)	(.02)	(.02)	(.02)	(.01)	(.01)	(.02)	(.02)	(.01)	(.02)	(.01)	(.01)

-*er*, and -*ir*), accuracy was .34 for -*er* verbs, and .29 for -*ir* verbs (compared to .50 for -*ar* verbs).

Model selection

For inferential tests, the data were modeled using PROC GLIMMIX in SAS with a binary logistic hierarchical linear model. Trials were nested within participants, with the correctness of the typed verb form as the outcome. Using a trial-level binary outcome was necessary because each participant completed a different randomly selected sample of trials per test, in order to sample from all cells across participants within time constraints. The hierarchical structure was imposed to capture the within-subjects nature of the data, and because treatment and individual differences variables were measured at the level of the participant. There were six variables initially considered in model selection at the trial level: (i) Test Time (pre-test, immediate post-test, one-week delayed post-test, 18-week follow-up), (ii) Transformation Status (regular, subregular without transformation, subregular with transformation), (iii) Verb Tense (present, preterite), (iv) Verb Irregularity Type (regular, spelling change, stem change), (v) Verb Conjugation Collapsed (-*ar*, non-*ar*), and (vi) Verb Conjugation Separated (-*ar*, -*er*, -*ir*). In addition, there were three variables considered at the participant level: (i) Feedback Type (rule, analogical), (ii) Standardized (*z*-score) Amount of Training (operationalized as number of training sessions completed), and (iii) Participant Gender. Participant gender was considered on the basis of prior claims that sex hormones influence dependence on the declarative memory system (Ullman, 2004); however, there was no evidence that gender predicted production accuracy for any verb type, so it was not included in the final model.

The model selection procedure was adapted from Snijders and Bosker (1999) and consisted of a random intercept for a reference category of pre-test, present tense, regular -*ar* verbs, estimated as .93 (*SE* = .15) on the log odds scale, with a random slope for immediate post-test (.51, *SE* = .12), delayed post-test (.45, *SE* = .11), and the smallest slope at the spring semester follow-up (.37, *SE* = .08).[2] Covariances between slopes and the intercept were allowed to vary, and showed a typical negative correlation between the random intercept and slopes (−.27 for immediate post-test, −.26 for delayed post-test, −.34 for follow-up), such that higher initial performance was correlated with lower levels of improvement. There were also positive correlations between slopes (covariances ranged from .33 to .35).

Because there were no differences between the metalinguistic and analogical feedback groups at any post-test (largest $t(25333) = -0.87$, $p = .383$), and no differences between metalinguistic and analogical feedback in amount of improvement after training, as well as no interactions between feedback condition and conjugation patterns, the final model collapsed across metalinguistic and analogical feedback. Separate models were estimated for Transformation Status (regular, subregular with transformation, subregular without transformation) and Verb Irregularity Type (regular, spelling change, stem change), because a design that would have randomized trials by both factors would have imbalanced the sampling of other factors. Because randomization was based on transformation status, the final model (see Table 3) used Transformation Status rather than Irregularity Type, but the coefficient direction and statistical inferences for the effects of other verb type

---

[2] This procedure involves first performing forward selection of random effects on a rich fixed effects model, then using backward selection to remove fixed effects that do not improve the model fit, and finally using backward selection on the random effects to confirm that these are necessary for the model.

Table 3. *Tests of fixed effects for the final model in Experiment 1.*

Model term	F(df)	p
Test Time (pre-test/immediate post-test/delayed post-test/follow-up)	149.1 (3, 25572)	<.001
Tense (present/preterite)	398.3 (1, 25572)	<.001
Transformation Status (regular verb/subregular without transformation/subregular with transformation)	1505.4 (2, 25572)	<.001
Verb Conjugation (-*ar*/not –*ar*)	130.9 (1, 25572)	<.001
Number of Training Sessions (z-score)	11.4 (1, 25572)	<.001
Test Time × Number of Training Sessions	4.2 (3, 25572)	.006
Test Time × Tense	17.5 (3, 25572)	<.001
Test Time × Transformation Status	17.4 (6, 25572)	<.001
Tense × Transformation Status	132.1 (2, 25572)	<.001
Transformation Status × Verb Conjugation	127.2 (2, 25572)	<.001
Test Time × Verb Conjugation	0.8 (3, 25572)	.504
Test Time × Tense × Transformation Status	2.8 (6, 25572)	.011
Tense × Transformation Status × Verb Conjugation	60.9 (2, 25572)	<.001
Test Time × Tense × Verb Conjugation	4.5 (3, 25572)	.004

variables (tense and conjugation), as well as for learning effects, were similar for both models. Removing spelling change verbs from the model also did not mitigate or remove the effects of the factors in the final model. In addition, the final model used the collapsed verb conjugation variable rather than separating -*er* and -*ir* verbs because of the greater parsimony of the two-level variable and because the separated model showed similar patterns for -*er* and -*ir* verbs, though there was a consistent main effect such that -*er* verbs were easier than -*ir* verbs. Non-significant lower-order interactions and main effects were included in the final model only if they were constituents of significant higher-order interactions.

Pre-test

At pre-test (see Table 2 for descriptive statistics), there were significant main effects of Verb Tense, Transformation Status, and Verb Conjugation (smallest $F(1$ or $3, 25337) = 130.10$, all $p < .001$). Present tense (.31) was more accurate than preterite tense (.16, $t(25337) = 21.75, p < .001$); regular verbs (.47) were more accurate than both subregular forms with transformation (.02, $t(25337) = 33.56, p < .001$) and subregular forms without transformation (.40, $t(25337) = 2.39, p = .017$); subregular forms without transformation were more accurate than subregular forms with transformation ($t(25337) = 7.90, p < .001$); and -*ar* verbs (.34) were easier than non-*ar* verbs (.15, $t(25337) = -7.07, p < .001$).

These main effects were qualified by interactions of Transformation Status × Tense, Transformation Status × Verb Conjugation, and Transformation Status × Tense × Verb Conjugation. The two-way interaction of Transformation Status and Tense showed that regular verbs were easier than both subregular verbs with and without transformation (both $t(25337) > 6.10$, $p < .001$) in the present tense; however, in the preterite tense, regular verbs were easier than subregular forms with transformation ($t(25337) = 25.17, p < .001$) but not easier than subregular forms without transformation ($t(25337) = 1.11, p = .266$). The two-way interaction of Transformation Status and Verb Conjugation showed that -*ar* verbs were easier than non-*ar* verbs for regular verbs ($t(25337) = 13.86, p < .001$), and subregular forms without transformation $t(25337) = 16.05, p < .001$), whereas non-*ar* verbs were more difficult than -*ar* verbs for subregular forms with transformation $t(25337) = 6.46$, $p < .001$), for which performance was close to floor.

There was also a three-way interaction of Tense × Verb Conjugation × Transformation Status (see Figure 1) such that, for regular verbs, moving from present to preterite tense and from -*ar* to non-*ar* verbs had an additive effect on difficulty. Present tense non-*ar* verbs were easier than preterite tense -*ar* verbs for regular verbs ($t(25337) = 9.96, p < .001$), but not for subregular forms without transformation ($t(25337) = 1.26, p = .209$). Performance for subregular forms with transformation was close to floor.

Learning and retention

Participants completed post-tests immediately after, one week after, and 18 weeks after training. Across all verb types, accuracy increased from the pre-test (.25, $SE = .02$) to the immediate post-test (.53), and this improvement was retained at one week (.44) and 18 weeks (.49) after the training (all $t(25337) > 3.00, p < .003$). Also, there was a significant interaction of Test Time × Amount of Training (see Table 3). The more training sessions a student completed, the greater the learning gains at
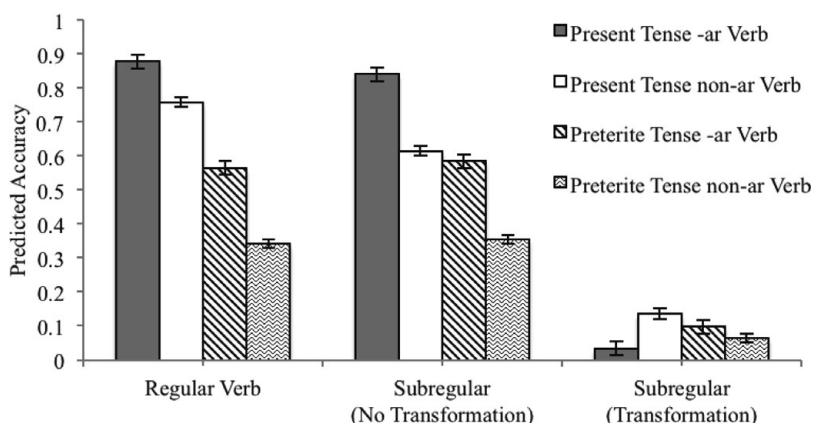
Figure 1. Interaction among Tense, Transformation Status, and Verb Conjugation in Experiment 1. Error bars represent ±1 *SE*.

both the one-week ($t(25337) = 2.26$, $p = .024$) and 18-week ($t(25337) = 3.68$, $p < .001$) post-tests. Amount of Training did not predict learning at immediate post-test ($t(25337) = 0.65$, $p = .517$). Although the test of fixed effects showed a significant main effect of Amount of Training (see Table 3), this was not reflected in a significant coefficient in the regression ($t(25337) = 0.64$, $p = .523$), and thus cannot be reliably interpreted.

Test Time interacted with Verb Tense (see Table 3). Present tense verbs showed higher accuracy than preterite tense at all tests (pre-test difference = .18; immediate post-test = .12; one-week post-test = .18; 18-week post-test = .27; all $t(25337) > 5.60$, all $p < .001$). Both present and preterite tense verbs improved from the pre-test (present = .31; preterite = .16) to the immediate post-test (present = .52; preterite = .55; both $t(25337) > 13.50$, $p < .001$), and the improvement was maintained at both delayed post-tests (one-week post-test: present = .44; preterite = .44; both $t(25337) > 11.30$, $p < .001$; 18-week post-test: present = .53; preterite = .44; both $t(25337) > 14.8$, $p < .001$). However, there was a difference in the magnitude of improvement, such that each post-test showed greater improvement from pre-test for preterite than for present tense verbs (all $t(25337) > 5.60$, $p < .001$).

As illustrated in Figure 2, this pattern was qualified by the interaction of Test Time × Verb Tense × Verb Conjugation. Accuracy for *-ar* verbs declined for both present and preterite tenses from a peak at immediate post-test (present = .58, preterite = .61, both *SE*s = .02) to the one-week post-test (present = .52, preterite = .53, both *SE*s = .02; $t(25337) > 2.80$, $p < .004$). Then, performance was maintained from the one-week to the 18-week post-test (present = .64; preterite = .56; present $t(25337) = 0.74$, $p = .460$; preterite, $t(25337) = 1.59$, $p = .111$). Non-*ar* verbs showed a different pattern. Although present tense non-*ar* verbs showed a similar lack of significant decline from the one-week (.36, *SE* = .02) to the 18-week (.42) post-tests ($t(25337) = 1.57$, $p = .116$; see

Figure 3), preterite tense non-*ar* verbs showed a significant decline from the one-week (.32, *SE* = .02) to the 18-week (.28) post-test ($t(25572) = -2.70$, $p = .007$). Finally, performance declined from the immediate to the 18-week post-test for preterite tense non-*ar* verbs ($t(25337) = -8.37$, $p < .001$), but was maintained for present tense non-*ar* verbs ($t(25337) = -0.95$, $p = .342$).

There was a significant three-way interaction of Test Time × Transformation Status × Tense (illustrated in Figure 3). For present tense verbs, accuracy was higher for regular verbs than subregular forms without transformation at all test times: at pre-test (regular = .71; subregular without transformation = .64, *SE* = .02; $t(25337) = 4.15$, $p < .001$), at immediate post-test (regular = .85, subregular without transformation = .80, *SE* = .02; $t(25337) = 3.07$, $p = .002$), at one-week post-test (regular = .80, subregular without transformation = .77, both *SE*s = .02; $t(25337) = 2.20$, $p = .028$), and at the 18-week follow-up (regular = .85, subregular without transformation = .79; $t(25337) = 4.57$, $p < .001$). In contrast, for the preterite tense, there was no significant difference between regular verbs and subregular forms without transformation at any test time (all $t(25572) < 1.25$, $p > .210$). For both tenses, all three levels of Transformation Status show improvement from pre-test to all post-tests, forgetting from immediate to one-week and 18-week post-tests, and no decline from one week to 18 weeks. However, improvement was larger for regular verbs and subregular verbs without a transformation than for subregular verbs with a transformation.

Trained participants compared to the untrained comparison group

To separate the effects of the training intervention from overall classroom exposure, at the 18-week follow-up the trained learners were compared to their untrained classmates, who had been enrolled in non-experimental classrooms the previous semester. Across all verb types,
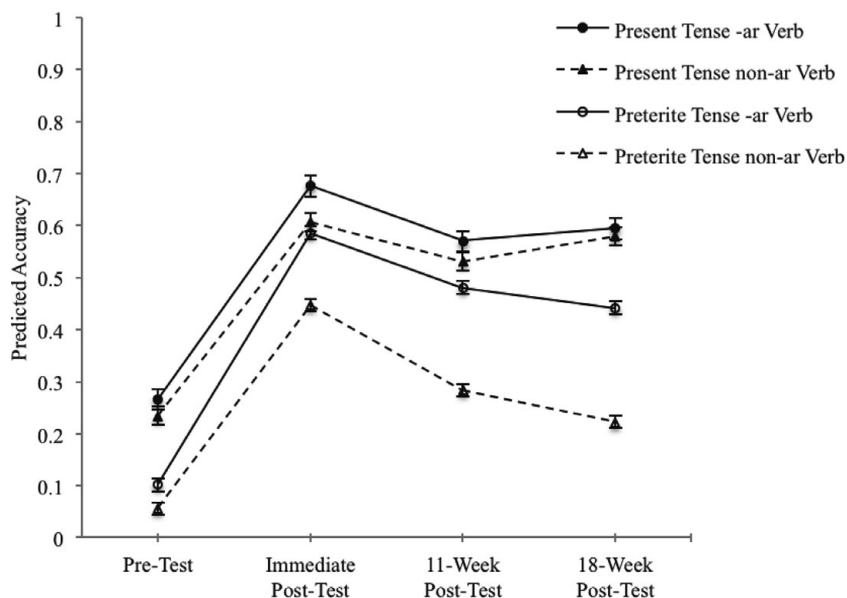
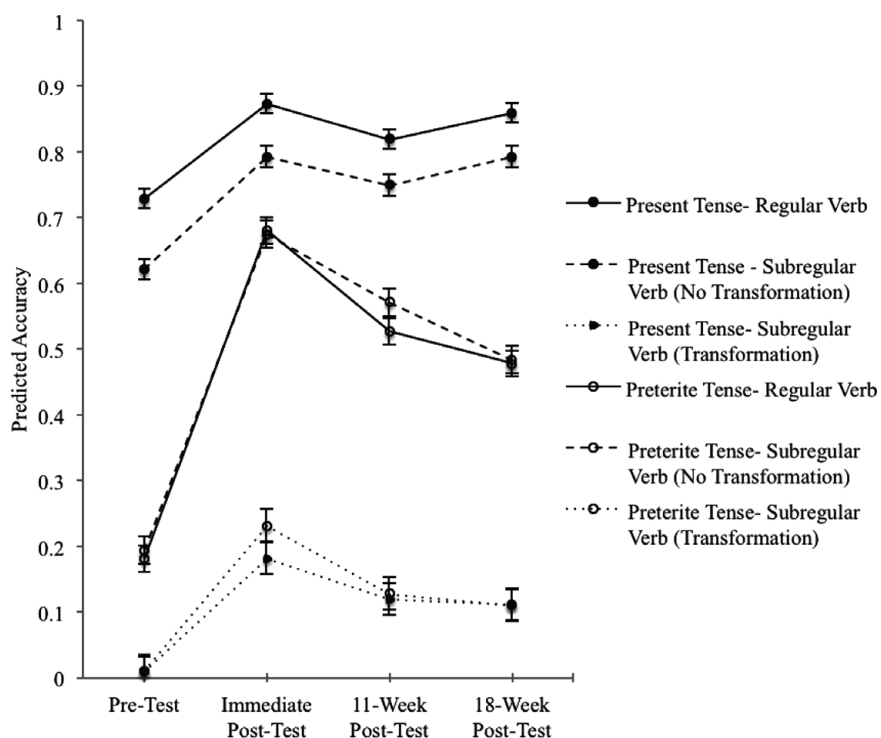Figure 2. Interaction among Test Time, Tense, and Verb Conjugation in Experiment 1. Error bars represent ±1 *SE*.

Figure 3. Interaction among Test Time, Tense, and Transformation Status in Experiment 1. Error bars represent ±1 *SE*.

trained participants were more likely to produce a correct conjugation than were untrained participants (See Table 4). Estimated across all verb types, predicted proportion correct for trained participants was .45 (95% confidence interval [.40, .50]); predicted proportion correct for untrained participants was .34 (95% confidence interval [.26, .42]). With one exception, the pattern of results for the untrained group was the same as for the

trained group at pre-test. The exception was that, for *-ar* verbs, regular (.68) and subregular forms without transformation (.61) were equally difficult ($t(13393) = 0.30$, $p = .762$) for the untrained participants. Because all other patterns were replicated across the two groups, we will not report those analyses here for the untrained participants. The results of the model contrasting the two groups can be found in Table 4. Here, we focus on

Table 4. *Tests of fixed effects of the model including the untrained comparison group in Experiment 1.*

Model term	F(df)	p
Training (trained/untrained)	5.5 (1, 13393)	.020
Tense (present/preterite)	319.6 (1, 13393)	<.001
Transformation Status (regular verb/subregular without transformation/subregular with transformation)	697.2 (2, 13393)	<.001
Verb Conjugation (-*ar*/not –*ar*)	104.2 (1, 13393)	<.001
Training × Verb Conjugation	3.6 (1, 13393)	.057
Training × Transformation Status	6.1 (2, 13393)	.002
Training × Tense	5.0 (1, 13393)	.025
Tense × Verb Conjugation	87.2 (1, 13393)	<.001
Tense × Transformation Status	79.6 (2, 13393)	<.001
Training × Transformation Status × Verb Conjugation	7.4 (2, 13393)	.001
Training × Transformation Status × Tense	2.4 (2, 13393)	.090
Tense × Transformation Status × Verb Conjugation	52.2 (2, 13393)	<.001

interactions involving the contrast between the trained and untrained groups.

There was an interaction of Training × Transformation Status × Verb Tense (see Table 4). Trained participants were more accurate across all preterite tense verbs: for preterite tense subregular verbs with a transformation (trained = .13, untrained = .09; $t(13393) = 2.44$, $p = .015$), for preterite tense subregular forms without a transformation (trained = .51, untrained = .40; $t(13393) = 2.63$, $p < .01$), and for regular verbs (trained = .53, untrained = .42; $t(13393) = 2.40$, $p = .016$). The interaction arises from the fact that the size of the training advantage was smaller for the preterite tense subregular verbs with transformation ($t(13393) = -2.19$, $p = .029$) than for preterite tense regular verbs and subregular verbs without transformation.

The interaction of Training × Transformation Status × Verb Conjugation arose from the fact that trained participants outperformed the untrained comparison group 18 weeks after training for non-*ar* subregular verbs with transformation (trained = .21, untrained = .08; $t(13393) = 5.05$, $p < .001$), and marginally for -*ar* subregular verbs without transformation (trained = .78, untrained = .71; $t(13393) = 1.66$, $p = .097$).

Finally, there was no significant difference between trained and untrained participants for present tense regular -*ar* verbs ($t(13393) = 0.19$, $p = .848$), present tense regular non-*ar* verbs ($t(13393) = -1.36$, $p = .173$), present tense subregular -*ar* verbs with transformation ($t(13393) = 1.06$, $p = .287$), or present tense subregular -*ar* without transformation ($t(13393) = 0.39$, $p = .694$).

Discussion

Beginning Spanish students showed evidence of learning Spanish conjugational patterns in their production of regular and subregular present and preterite tense verbs before and immediately after training with metalinguistic or analogical feedback, as well as after delays of one and 18 weeks from the end of training. By comparing the trained and untrained participants, and by showing improvement from pre-test to three post-tests, we demonstrated that 90 minutes of training produced a substantial improvement in performance that was retained 18 weeks after training. The largest advantage for trained participants compared to the untrained comparison group was for preterite tense and subregular forms not requiring a transformation. We found no difference between metalinguistic and analogical feedback, suggesting that the key mechanism underlying the substantial improvement from pre-test to post-tests may be practice that provides correctness feedback along with the correct answer. The difference between the two instructional methods, if it exists, produces a small enough effect that practice with feedback and correct input may be the most important component.

Both cross-sectional and longitudinal data showed that pre-test performance and learning gains were meaningfully predicted by combinations of tense, verb conjugation, and transformation status. The results suggest that beginning learners performance cannot be fully explained by full-form retrieval. This result goes against the dual-route model's prediction of reliance on full-form retrieval in L2 learners, and is consistent with the hybrid model prediction of a combinatorial strategy gradually proceduralized with extensive practice. However, one potential objection to this interpretation is that learners may shift quickly from a compositional strategy, at the very early stages of instruction (as in the Experiment 1 sample), to an emphasis on full-form retrieval, after additional exposure to conjugated verbs. We do not predict such a shift. However, to rule out this alternative explanation, we need to generalize the findings of Experiment 1 by testing learners at later stages

of classroom instruction. In Experiment 2, we examine this question by testing learners in the third semester of the classroom sequence.

Experiment 2

Experiment 2 used the same intervention and training materials from Experiment 1 with a sample of intermediate learners. Unlike Experiment 1, there was no 18-week follow-up and no untrained comparison group.

Method

Participants

Participants were 862 students (439 female, 387 male, 36 missing data) enrolled in third-semester Spanish at a large university who began at least one testing session for extra credit. Eight hundred and six were native English speakers, 25 were native speakers of another language or bilingual, and 32 were missing native language data. For the 778 participants who provided data about their age of first exposure to Spanish (86 missing or unsure responses), the average age of first exposure was 12.4 years (SD = 3.3 years).

Of the 862 participants who began at least one test or training session, 61 were excluded because they were missing pre-test data (of whom 42 did not complete any training sessions), and 83 were excluded because they did not complete any post-test (of whom 23 did not complete any training sessions). Seven participants were excluded because they identified as native or heritage Spanish speakers, and three were excluded for producing no correct responses or gaming the system (typing random strings or single letters).

The final sample consisted of the 708 participants who took the pre-test and at least one post-test, completed at least one training session, and were not excluded based on language background or gaming behavior. From this pool, 345 were randomly assigned to the metalinguistic feedback group and 363 to the analogical feedback group. Of those, 420 participants completed all four assigned training sessions. These 420 participants did not differ in pre-test accuracy from the 288 participants who missed one or more training sessions (t(706) = 1.48, p = .14).

Design and materials

The materials and design were identical to Experiment 1.

Procedure

Participants completed a pre-test, then one training session per week for four weeks, with the final training session taking place in class before the immediate post-test. Three weeks after the immediate post-test, a delayed post-test was administered in class. Note that this is a longer delay than the one-week post-test in Experiment 1.

Each training session lasted for 20 minutes (the number of completed trials varied for each learner), for a total training time of 80 minutes (compared to 90 in Experiment 1), and each test consisted of 25 trials (average time to complete each test was approximately five minutes). Unlike in Experiment 1, no vocabulary data were available. However, because the Experiment 2 sample was enrolled in a more advanced course, it is reasonable to assume that they would not have substantially less vocabulary knowledge than the sample tested in Experiment 1.

Results

Descriptive statistics

Descriptive statistics for the full sample of 708 participants are presented in Table 5. All results are presented as mean proportion correct. Because of the large sample size, all standard errors were .01 after rounding (i.e., between .005 and .014) unless otherwise noted. Overall mean accuracy was .51 at pre-test, .65 at immediate post-test, and .63 at delayed post-test (all SEs = .001).

Model selection

As in Experiment 1, the data were modeled using PROC GLIMMIX in SAS with a binary logistic hierarchical model of trials nested within participants, with the correctness of the typed verb form as the outcome and using the same model selection procedures and the same initial predictors as Experiment 1.

Final model

As in Experiment 1, the final model (see Table 6) included a random intercept (1.20, SE = .08) and random slopes for the immediate (.62, SE = .06) and three-week post-test (.72, SE = .07), which were negatively associated with the intercept (immediate post-test = −.41, SE = .06; delayed post-test = −.43, SE = .06) and positively associated with each other (.72, SE = .07). The predictors retained in the final model were Test Time, Transformation Status, Tense, Verb Conjugation, and Amount of Training, as well as the significant interactions among those variables that improved the model fit. As in Experiment 1, Transformation Status could not co-exist with Verb Irregularity Type; however, a model with Irregularity Type also revealed that spelling change verbs were easier than stem change verbs at all test times (all t(53720) > 5.00, p < .001). This model also replicated the main effects of Test Time, Tense, Amount of Training, and Verb Conjugation, as described below.

As in Experiment 1, there was no consistent advantage for metalinguistic or analogical feedback in learning from pre-test to immediate or delayed post-test, no difference between metalinguistic and analogical feedback at any test

Table 5. *Average observed proportion correct (standard error) over time of testing broken down by Transformation Status, Tense, and Verb Conjugation for the 708 participants included in the final analysis in Experiment 2.*

	Regular verb				Subregular without transformation				Subregular with transformation			
	-ar		Non *-ar*		*-ar*		Non *-ar*		*-ar*		Non *-ar*	
	Present	Preterite	Present	Preterite	Present	Preterite	Present	Preterite	Present	Preterite	Present	Preterite
Pre-test	.87	.53	.78	.40	.82	.50	.70	.37	.06	.18	.13	.07
	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)
Immediate	.91	.77	.82	.61	.85	.73	.69	.58	.14	.39	.35	.16
post-test	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.02)	(.01)	(.01)	(.02)	(.01)	(.01)
Delayed	.91	.76	.78	.58	.86	.71	.74	.54	.10	.30	.29	.17
post-test	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)	(.02)

Table 6. *Tests of fixed effects for the final model in Experiment 2.*

Model term	F(df)	p
Test Time (pre/immediate post/delayed post)	190.3 (2, 53984)	<.001
Tense (present/preterite)	797.8 (1, 53984)	<.001
Verb Conjugation (*-ar*/not-*ar*)	366.4 (1, 53984)	<.001
Transformation Status (regular verb/subregular verb without transformation/subregular verb with transformation)	3221.8 (2, 53984)	<.001
Number of Sessions (*z*-score)	27.0 (1, 53984)	<.001
Test Time × Number of Sessions	9.6 (2, 53984)	<.001
Test Time × Tense	35.7 (2, 53984)	<.001
Test Time × Transformation Status	15.6 (4, 53984)	<.001
Tense × Transformation Status	285.6 (2, 53984)	<.001
Transformation Status × Verb Conjugation	93.1 (2, 53984)	<.001
Test Time × Verb Conjugation	0.8 (2, 53894)	.455
Tense × Verb Conjugation	171.1 (1, 53984)	<.001
Number of Sessions × Test Time × Verb Conjugation	2.9 (3, 53984)	.034
Number of Sessions × Test Time × Tense	3.1 (3, 53984)	.026
Test Time × Tense × Transformation Status	14.3 (4, 53984)	<.001
Tense × Transformation Status × Verb Conjugation	177.8 (2, 53984)	<.001
Test Time × Transformation Status ×Verb Conjugation	3.8 (4, 53984)	.004

time, and no interaction of feedback type with any other predictor. Therefore, in the final analysis, we collapsed across training conditions. Table 6 presents the factors included in the final best-fit model with tests of fixed effects.

Performance at pre-test

At pre-test, there were main effects of Transformation Status, Tense, and Verb Conjugation (see Table 6). This pattern was consistent with the results for Experiment 1. Regular verbs (.70) were easier than subregular forms without a transformation (.64; $t(53720) = 6.56, p < .001$). As predicted by hybrid models, both regular verbs ($t(53720) = 46.20, p < .001$) and subregular forms without a transformation ($t(53720) = 44.68, p < .001$)

were easier than subregular forms with a transformation (.11).

Consistent with our predictions, across all verb types, present tense (.60) was easier than preterite tense (.42, $t(53720) = 21.58, p < .001$), and *-ar* verbs (.61) were easier than non-*ar* verbs (.38, $t(53720) = 9.57, p < .001$). When the same model was run with Verb Conjugation as a separated variable (*-ar*/-*er*/-*ir*), the difference between *-er* (.45) and *-ir* (.35) verbs was also significant at each test ($p < .001$).

These main effects were qualified by several significant interactions. The two-way interactions of Tense × Verb Conjugation and Tense × Transformation Status (see Table 6) were qualified by a three-way interaction among Tense × Transformation Status × Verb Conjugation,
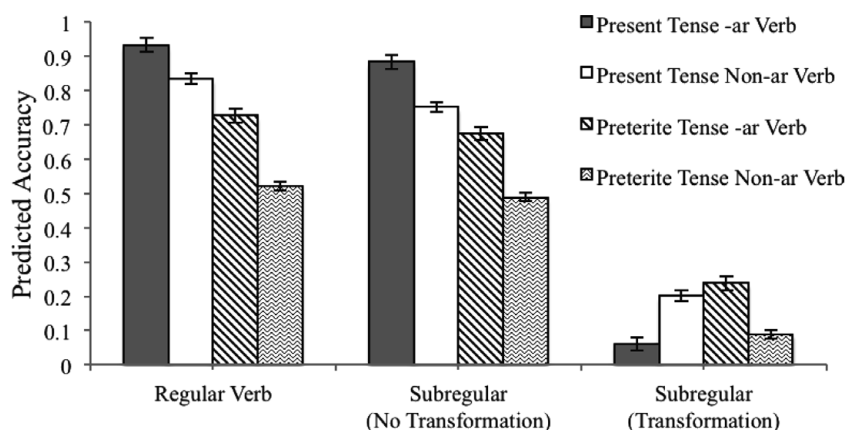
Figure 4. Interaction among Test Time, Transformation Status, and Verb Conjugation in Experiment 2. Error bars represent ±1 *SE*.

illustrated in Figure 4. Each mean in the three-way interaction was significantly different from all other means (smallest $t(53720) = 2.81$, $p = .005$). Figure 4 shows that, for regular verbs and for subregular forms without a transformation, the effects of Verb Conjugation and Tense were similar and additive (that is, present tense -*ar* verbs > present tense non–*ar* verbs > preterite tense -*ar* verbs > preterite tense non-*ar* verbs). In contrast, for subregular forms with a transformation, preterite -*ar* verbs (.18) were easiest, followed by present non-*ar* verbs (.13), preterite non-*ar* verbs (.07), and present -*ar* verbs (.06). This pattern likely reflects the fact that subregular transformations in the preterite tense were more likely to be spelling-change patterns, whereas transformations in the present tense were more likely to be stem-change patterns.

Improvement after training
Across all verb types, there was a significant effect of Test Time such that performance improved significantly after training (see Table 6). Learners improved from pre-test (.51) to immediate post-test (.65, $t(53720) = 4.45$, $p < .001$) and to delayed post-test (.63, $t(53720) = 3.84$, $p < .001$). As expected, across all verb types, there was significant forgetting from the immediate to the three-week delayed post-test, ($t(53720) = -3.02$, $p < .001$). The amount of improvement from pre-test to immediate and three-week delayed post-tests was qualified by interactions of Test Time × Transformation Status and Test Time × Tense, but there was no significant Test Time × Verb Conjugation interaction (see Table 6). These two-way interactions were further qualified by three-way interactions of Test Time × Transformation Status × Verb Conjugation and Test Time × Tense × Transformation Status (see Table 6).

The three-way interaction of Test Time × Transformation Status × Verb Conjugation is illustrated in Figure 5. All six patterns improved from pre-test

to immediate and delayed post-test, and the three-way interaction reflects differences in forgetting from the immediate to the three-week delayed post-test. Both -*ar* and non-*ar* subregular verbs without a transformation did not decline from immediate (-*ar* = .78, non-*ar* = .61) to three-week delayed post-test (-*ar* = .78, $t(53720) = -0.54$; non-*ar* = .60, $t(53720) = -0.09$, both $p > .500$). However, both -*ar* and non-*ar* subregular verbs with transformation showed forgetting (-*ar* $t(53720) = -3.84$; non-*ar* $t(53720) = -2.57$, both $p < .011$). Non-*ar* regular verbs showed forgetting ($t(53720) = -2.39$, $p = .017$), unlike -*ar* regular verbs ($t(53720) = -0.64$, $p = .523$).

The three-way interaction of Test Time × Tense × Transformation Status is shown in Figure 6. There was significant learning from pre-test to immediate post-test for all patterns except present tense subregular verbs without a transformation ($t(53720) = 0.55$, $p = .582$; all other $t(53720) > 4.00$, $p < .001$). Although present tense subregular verbs without transformation did not improve from pre-test to immediate post-test, they did improve at the three-week delayed post-test ($t(53720) = 2.69$, $p = .007$), as did all other conjugation patterns (smallest $t(53720) = 2.55$, $p = .011$). As illustrated in Figure 6, improvement was largest for preterite tense regular verbs and preterite tense subregular forms without a transformation.

One indicator of the success of the intervention was greater improvement from pre-test to the post-tests for learners who completed more training sessions (i.e., a significant Test Time × Amount of Training interaction; see Table 6). This was true for both immediate post-test ($t(53720) = 2.29$, $p = .022$) and three-week delayed post-test ($t(53720) = 3.51$, $p < .001$). Those learners who completed more sessions did not score significantly higher at pre-test ($t(53720) = 1.41$, $p = .160$), suggesting that this effect was not because learners who completed more training were initially better at Spanish. This interaction was further qualified by three-way interactions with Verb
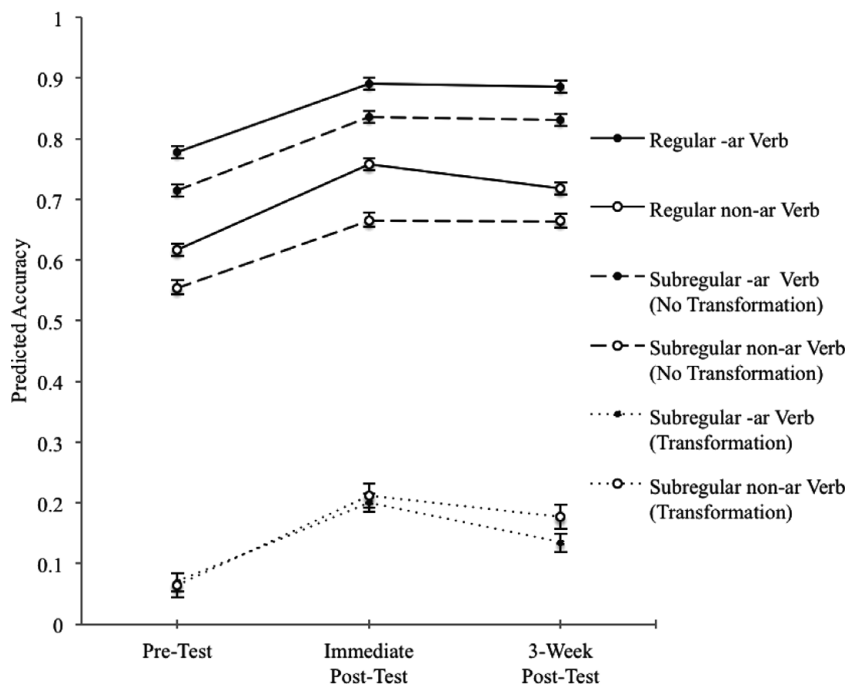
Figure 5. Interaction among Tense, Transformation Status, and Verb Conjugation in Experiment 2. Error bars represent ±1 *SE*.
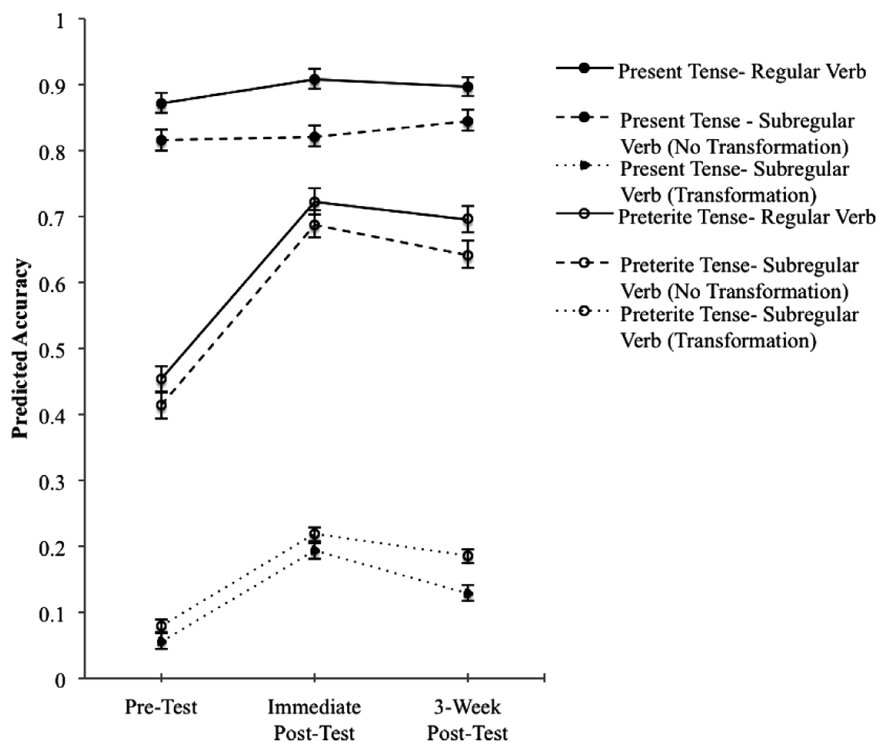

Figure 6. Interaction among Test Time, Tense, and Transformation Status in Experiment 2. Error bars represent ±1 *SE*.

Conjugation and with Tense (see Table 6). The size of the learning advantage for higher doses of training was larger at delayed post-test for -*ar* verbs than non-*ar* verbs ($t(53720) = 2.99$, $p = .003$), although it was equal for -*ar* and non-*ar* verbs at immediate post-test ($t(53720) =$

$1.53, p = .127$). The increased learning with more training sessions was also larger for preterite tense than present tense, though only at immediate ($t(53720) = 2.54$, $p = .011$) and not delayed ($t(53720) = 1.24$, $p = .213$) post-test.

Combined model

To compare the effect of training for beginners in Experiment 1 with the effect for intermediates in Experiment 2, we also ran a model combining both datasets and coding Class Level as a two-level between-subjects predictor (beginning/intermediate). Here we report only the effects involving Class Level, because all other effects are described above. The model reflects the overall high level of similarity between the two samples, with both groups improving after training and Tense, Transformation Status, and Verb Conjugations each significantly and similarly predicting accuracy before and after training in both groups. Note that the interval between immediate and delayed post-test was two weeks larger for intermediates in Experiment 2, which may result in more forgetting for intermediate than beginner learners.

At pre-test, across all patterns, intermediate learners were significantly more accurate than beginners (beginner = .14, intermediate = .40; $t(69117) = 11.14$, $p < .001$). This difference was not significant immediately after training (beginner = .58, intermediate = .62; $t(69117) = 1.50$, $p = .133$). However, by the delayed post-test, intermediates again out-performed beginners across verb patterns (beginner = .48, intermediate = .58; $t(69117) = 3.38$, $p < .001$). That is, intermediates forgot less in three weeks than beginners forgot in one week.

There were three significant three-way interactions. In each case, the difference between beginners and intermediates was eliminated at the immediate post-test for the pattern that was less accurate at pre-test. In each case, the difference then re-emerged at the delayed post-test.

There was an interaction of Class Level × Test Time × Verb Conjugation such that for *-ar* verbs, intermediate learners were more accurate than beginners at pre-test (beginner = .18, intermediate = .46; $t(69117) = 10.26$, $p < .001$) at immediate post-test (beginner = .63, intermediate = .68; $t(69117) = 2.08$, $p = .037$) and at delayed post-test (beginner = .53, intermediate = .64; $t(69117) = 3.44$, $p = .001$). In contrast, for non-*ar* verbs, intermediates were more accurate at pre-test (beginner = .11, intermediate = .34; $t(69117) = 10.73$, $p < .001$), but the difference disappeared at immediate post-test (beginner = .53, intermediate = .55; $t(69117) = 0.67$, $p = .500$), and re-appeared at delayed post-test (beginner = .42, intermediate = .51; $t(69117) = 2.75$, $p = .006$). Both beginners and intermediates showed the overall pattern of improvement for both *-ar* and non-*ar* verbs.

Next, there was an interaction of Class Level × Test Time × Transformation Status. Both beginners and intermediates showed the overall pattern of regular verbs being produced more accurately than subregular verbs without a transformation, each of which was produced more accurately than subregular verbs with a transformation. For regular verbs, intermediates were

more accurate at all test times: at pre-test (beginner = .43, intermediate = .70; $t(69117) = 8.98$, $p < .001$), at immediate post-test (beginner = .79, intermediate = .84; $t(69117) = 2.29$, $p = .022$), and at delayed post-test (beginner = .70, intermediate = .82; $t(69117) = 4.67$, $p < .001$). In contrast, for subregular forms without a transformation, intermediates were more accurate at pre-test (beginner = .40, intermediate = .64; $t(69117) = 7.54$, $p < .001$), but the difference disappeared at immediate post-test (beginner = .74, intermediate = .76; $t(69117) = 0.98$, $p = .328$), and re-emerged at the delayed post-test (beginner = .69, intermediate = .82; $t(69117) = 2.27$, $p = .023$). Like subregular forms without a transformation, for subregular forms with a transformation, intermediates were more accurate than beginners at pre-test (beginner = .01, intermediate = .07; $t(69117) = 11.31$, $p < .001$), but the two groups were equal immediately after training (beginner = .20, intermediate = .21; $t(69117) = 0.56$, $p = .576$). Unlike for subregular forms without a transformation, intermediates were not more accurate at the delayed post-test, although the trend was in that direction (beginner = .13, intermediate = .15; $t(69117) = 1.67$, $p = .096$).

Finally, there was an interaction of Class Level × Test Time × Verb Tense. Both beginners and intermediates improved after training for both present and preterite tense, and present tense was easier than preterite tense at all time points. For present tense verbs, intermediates were significantly more accurate than beginners at pre-test (beginner = .26, intermediate = .55, $t(69117) = 9.38$, $p < .001$), marginally more accurate at immediate post-test (beginner = .64, intermediate = .69; $t(69117) = 1.77$, $p = .076$), and significantly more accurate at delayed post-test (beginner = .57, intermediate = .66; $t(69117) = 2.95$, $p = .003$). For preterite tense verbs, intermediates were significantly more accurate than beginners at pre-test (beginner = .07, intermediate = .27; $t(69117) = 11.59$, $p < .001$), but this advantage disappeared immediately after training (beginner = .64, intermediate = .69; $t(69117) = 0.97$, $p = .331$). At the delayed post-test, intermediates were again more accurate than beginners (beginner = .39, intermediate = .49; $t(69117) = 3.23$, $p = .001$).

Discussion

The results of Experiment 2 replicated Experiment 1, showing that beginning and intermediate learners demonstrated an almost identical pattern of baseline performance and learning after training for all conjugational patterns, although the magnitude of the difference between beginners and intermediates before and after training varied by conjugation pattern (as shown in the interactions with Class Level). This suggests that the evidence for a compositional strategy (transfer within

a conjugational pattern and better baseline performance and greater learning for more predictable patterns) is generalizable across beginning and intermediate learners. The most important difference between Experiments 1 and 2 was that beginning learners improved substantially more after the training intervention. Intermediates showed improvement from .51 to .62, compared to improvement from .25 to .53 in beginners, suggesting that this type of targeted practice could be most effective early in learning, when familiarity with conjugational patterns is lower.

General discussion

Overall, the current study indicates that learners are able to make use of conjugational pattern information beyond whole word form retrieval of each token, and that practice with difficult patterns can improve performance across an instructionally relevant delay of one semester. Let us examine these two findings more closely.

Do learners use a compositional strategy?

Learner performance at pre-test showed that properties of the conjugational pattern predicted conjugation accuracy. This provides evidence against the complete reliance on full-form retrieval predicted by the dual-route model early in learning and by single-route models. Regular verbs showed the highest accuracy before training, which cannot be wholly attributed to overregularization, because learners were less accurate for subregular forms not requiring a transformation (which match the regular pattern) than for regular verbs. Furthermore, subregular forms without a transformation (e.g., *podemos* "we can") were easier than subregular forms with a transformation (e.g., *puedo* "I can"). This pattern is not consistent with the dual-route compositional deficit model prediction that reliance on full-form retrieval for regular verbs is a central deficit in adult L2 learners.

All three models we have considered – dual-route, single-route, and hybrid – recognize that token frequency predicts accuracy. However, only the hybrid model predicts an effect for type frequency beyond the effects of token frequency. In fact, type frequency effects were evident in the findings. First, the improvement from pre-test to post-test on untrained generalization items cannot be attributed to token frequency alone. Second, there was better baseline performance and less decay in performance for -*ar* verbs than for non -*ar* verbs, perhaps because the -*ar* verb conjugation type accounts for 90% of all Spanish regular verbs (Hualde et al., 2010). Third, instructional sequence predicted baseline accuracy, as evidenced by the fact that present tense was more accurate than preterite tense at pre-test across all verbs. Fourth, within a subregular verb, forms without a transformation were much easier than forms with a transformation. Finally,

the drop in accuracy from present tense -*ar* verbs to preterite tense and non-*ar* verbs was additive, suggesting that these difficulty factors act independently to decrease the probability of correctly conjugating the verb.

Patterns in improvement after training also provide evidence of the use of a compositional strategy. Properties of the conjugation pattern predicted the size of the learning gains, as well as the effects of amount of training and the size of the advantage for trained participants compared to untrained participants. That is, learning gains were larger for regular verbs and subregular forms without a transformation than for subregular forms with a transformation, and gains were larger for preterite tense than for present tense verbs. Gains were largest for preterite tense regular verbs in both experiments, for all non-*ar* verbs in Experiment 1, and for all -*ar* verbs in Experiment 2. This difference between the two experiments could reflect the difference between initial learning of a pattern (i.e., beginners may not be comfortable conjugating non-*ar* verbs at all) compared to refining a known pattern (i.e., intermediate learners may have surpassed some initial threshold of familiarity). In both experiments, the decline from the immediate to the delayed post-test (forgetting) was largest for non-*ar* subregular verbs without a transformation, and larger for preterite tense than for present tense. In interpreting these results, we can speculate that the differences in forgetting may correspond with differences in practice opportunities in the classroom after the end of training.

If learners were using whole word form recall rather than a compositional strategy, then each verb form should only improve as a function of how often the individual form was practiced. This was not the case in the current data, because testing data came from verbs not used in training, so that improvements reflect transfer across the conjugation pattern, rather than mere exposure to specific tokens. Moreover, this result cannot be attributed to ceiling effects, because no condition showed post-test accuracy greater than .90. These results again suggest that learners show effective transfer within conjugational patterns – which does not support the hypothesis that all inflected-form tokens are retrieved independently as whole lexical forms. It is important to note the difference between the results of the current study and those of Bowden et al. (2010), who argued that learners showed full-form retrieval of all verbs. This difference may arise from two methodological differences between the two studies. First, the current study measured accuracy rather than response time (as measured in Bowden et al., 2010), and accuracy. Therefore, it is more reflective of initial strategic processing than of automaticity. This outcome is more appropriate for a learner population who are in the very early stages of classroom instruction. Second, because Bowden et al. (2010) used a much smaller pool of stimuli, drawing only from imperfect and present tense

first and third person forms and not presenting any present tense stem-change verbs without a transformation, some contrasts detected in the current data could not be reflected in that analysis. Therefore, the data presented here can be interpreted as measuring the very early stages of learning. However, it is unknown how the response profile demonstrated here might evolve into that shown by the advanced L2 speakers tested by Bowden et al. (2010).

Is training effective at improving L2 verb conjugation?

The second research question was whether training with feedback can improve the production of inflected verb forms and, if it does, whether this improvement can be maintained over time. The strongest evidence for a training effect is that trained participants outperformed untrained participants a semester after training. Moreover, overall improvement was significant and robust 18 weeks after the end of training for beginners in Experiment 1 (from .25 to .53), and three weeks after training for intermediates in Experiment 2 (from .51 to .63). The amount of improvement depended on amount of training for both samples, also showing that training had a meaningful effect on learner ability to produce correct conjugations at both levels. This conclusion was also supported by the finding that pre-test performance for intermediate learners was .51 across all patterns, and performance after training for beginners was .53, although this comparison is cross-sectional rather than longitudinal. This finding shows how much training can speed up learning, especially for difficult patterns.

It is important to note that we could not control classroom exposure, because the training intervention supplemented existing instructional time in intact classrooms. Therefore, it is likely that patterns more often practiced in class would be better retained than those that are less often practiced. Indeed, this hypothesis is the basis of our interpretation of the finding that for some verb types but not others, conjugation accuracy did not decline or actually increased from immediate to delayed post-tests, which could reflect increased exposure to verbs that are taught first or are more frequently practiced in class (particularly present tense, -*ar* verbs).

Intermediates forgot less three weeks after training than beginners did one week after training, perhaps because they had received a higher cumulative level of exposure to all patterns, permitting a higher level of consolidation. Forgetting was also greater for patterns with low pre-test accuracy, possibly because of less frequent exposure in the curriculum leading to weaker consolidation. We speculate that the amount of forgetting from the immediate to the delayed post-test varied by conjugational pattern and class level due to variability in the amount of practice opportunities in the curriculum. However, hypothesized differences in classroom exposure do not

explain improvement after training, as the difference between earlier and later instructed patterns (e.g., present versus preterite tense) narrowed with training rather than expanding (which would be predicted if classroom exposure were fully responsible for the results).

The findings suggest that training, particularly early in the classroom sequence, is especially helpful for difficult patterns. In accordance with these findings, we propose that practice should focus on less-frequent structures, and that a relatively small amount of practice on targeted grammatical structures can have a substantial effect on production accuracy.

Is metalinguistic feedback more effective than analogical feedback?

The major pedagogical implication of this study is that practice, combined with correctness feedback and the correct answer, leads to effective learning. This result is important in relation to the debate regarding the role of decontextualized practice in L2 learning. The training in this study involved "focus on forms" rather than "focus on meaning" or a "focus on form" (Long & Robinson, 1998). The current results indicate that a focus on forms is valuable in learning Spanish conjugation, particularly for beginners. However, these results should not be interpreted as indicating that either "focus on meaning" or "focus on form" are not also important for instruction.

Across two experiments, we found no difference between metalinguistic and analogical feedback. This suggests that the difference between metalinguistic and analogical feedback is not the key mechanism underlying the substantial improvement from pre-test to post-tests. Instead, the key mechanism is learning through practice with correctness feedback. The failure to find an advantage for metalinguistic feedback may arise from the fact that students in both training groups already had received metalinguistic in-class instruction regarding the relevant patterns. If this is true, then the lack of an advantage for analogy can be interpreted as indicating that the familiar examples did not add value beyond reminding students of the rules they had been taught. A good way to test this prediction would be to compare practice with metalinguistic feedback to practice without feedback, practice with correctness (right/wrong) feedback only, and practice with correctness feedback and the target response.

Conclusion

In sum, the results of the current study demonstrate that learners behave as if they are treating conjugation as a compositional task rather than a retrieval task for individual inflected forms. Crucially, this approach applies both to regular verbs and to verbs with subregular

patterns, and cannot be fully explained as arising from overregularization. In addition, providing a total of 80–90 minutes of practice with correctness feedback led to substantial gains that were maintained up to one semester.

Both pre-test performance and gains after training tested the predictions of three common models of morphological processing. Dual-route models propose that L1 compositional processing occurs for regular default verbs and full-form retrieval occurs for non-default and non-regular verbs (Marcus et al., 1995). The L2 compositional deficit extension of the dual-route model holds that L2 learners process all verbs through full-form retrieval (Clahsen & Felser, 2006; Ullman, 2004). Single-route models also claim that all verbs are processed as whole words (Bybee & McClelland, 2005; Eddington, 2000), while allowing for analogical effects. Finally, hybrid models combine full-form processing with compositional processing of multiple patterns weighted based on the predictability and frequency of each pattern (Ellis, 2002; MacWhinney, 2007; Tkachenko & Chernigovskaya, 2010). The pattern of findings both before training and for improvement after practice cast doubt on models postulating only full-form retrieval and instead support the predictions of the hybrid models. In learners such as those tested in the current study, reliable patterns should be easiest after classroom exposure (reflected in pre-test performance) and should improve more with practice (reflected in learning data).

The current data suggest that, at the beginning of learning, composition is a strategic process that requires a greater degree of attention and depends to a large degree on the amount of prior exposure and practice with a given pattern. Over the course of learning, the composition of regular and subregular patterns may become increasingly automatized, as described in other domain-general models of skill and procedural learning (Anderson & Fincham, 1994; Shiffrin & Schneider, 1977). Dual-route models also accept this possibility. For example, Bowden et al. (2010, p. 48) note that, "with increasing experience (and accompanying proficiency), it is predicted that the grammar will undergo proceduralization, and thus will become increasingly L1-like in its neurocognitive mechanisms". The hybrid view and the dual-route view do not differ in regard to this issue. Where they differ is in regard to the timing of the learner's use of compositional analysis. Dual-route models envision a long period during which the learner only acquires rote chunks without engaging in compositional analysis. The hybrid model envisions a much earlier onset of compositional analysis, which is consistent with the results of the current study. In both accounts, proceduralization increases through learning.

Future studies should test whether indeed learners can proceduralize grammatical patterns. The hybrid model

predicts that the greater the type frequency and reliability of a pattern, the easier it should be to automatize. This predicted effect could help explain not only the pattern of acquisition in adult learners, but also the difference in the degree of procedural composition found between regular and subregular or between default and non-default regular verbs in native speakers. In this effort, future studies would also gain from comparing the current dataset to performance with fully idiosyncratic verbs; that is, verbs that are difficult to map onto predictable patterns. This comparison would allow us to address, for example, whether forms of subregular verbs with transformations are treated like forms of fully irregular verbs. It would also be valuable to collect native speaker baseline data within these specific task constraints, in order to assess how close or far away learners are from achieving a nativelike ceiling in typing production.

The contrast between dual-route and hybrid models has interesting consequences for second language pedagogy. In the case of Spanish, the dual-route model holds that the combinatorial default pattern applies only to regular -*ar* verbs, leaving all other verbs to full-form retrieval. This means that there should be two methods of instruction for Spanish verbs. The first method would involve training on the default -*ar* pattern, which is the only pattern that can be proceduralized in the dual-route model. The second method would involve rote memorization of forms, independent of their conjugational status. Because non-default conjugational patterns cannot be composed, teaching those patterns would be ineffective and inefficient, relative to time spent in rote memorization. On the other hand, the hybrid model suggests that instruction should focus on all reliable patterns, including the minor rules for stem and spelling alterations, as well as the regular patterns for the non-default second and third conjugations. In fact, the approach supported by the hybrid model is currently being followed in most Spanish language curricula.

The current results suggest that, especially for L2 pedagogy, the debate between single-route and dual-route models of learning has failed to provide a sharp focus on the role of learning of subregular and non-default patterns. However, hybrid models that enhance the dual-route approach with competition between patterns of varying reliability can help explain much of these data. By expanding the study of L2 morphological processing to beginning classroom learners, the current study also demonstrates the importance of an explicitly developmental approach to models of L2 learner processing. These results also suggest that learners may need deliberate production practice, particularly with difficult patterns. Without such practice, learners may not be able to achieve full mastery of the various regular, subregular, and irregular patterns involved in conjugating Spanish verbs.

References

Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills form examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1322–1340.

Anderson, S. R. (1992). *Amorphous morphology*. Cambridge: Cambridge University Press.

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.

Bowden, H. W., Gelfand, M. P., Sanz, C., & Ullman, M. (2010). Verbal inflectional morphology in L1 and L2 Spanish: A frequency effects study examining storage versus composition. *Language Learning, 60,* 44–87.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10,* 425–455.

Bybee, J. (2007). *Frequency of use and the organization of language*. New York: Oxford University Press.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22,* 381–410.

Castells, M. d., Guzmán, E., Lapuerta, P., & García, C. (2002). *Mosaicos: Spanish as a world language*. Upper Saddle River, NJ: Prentice Hall.

Chan, A., Lieven, E., & Tomasello, M. (2009). Children's understanding of the agent– patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English. *Cognitive Linguistics, 20,* 267–300.

Clahsen, H. (2006). Dual-mechanism morphology. In K. Brown (ed.), *Encyclopedia of language and linguistics* (vol. 4), pp. 1–5. Oxford: Elsevier.

Clahsen, H., Aveledo, F., & Roca, I. (2002). The development of regular and irregular verb inflection of Spanish child language. *Journal of Child Language, 29,* 591–622.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3–42.

Dietrich, R., Klein, W., & Noyau, C. (eds.) (1995). *Acquisition of temporality in a second language*. Amsterdam: John Benjamins.

Dracos, M. (2010). Overrepresentation of subject personal pronouns in instructed Spanish. Presented at the Second Language Research Forum, October 14–17, 2010, University of Maryland, College Park, MD.

Eddington, D. (2000). Analogy and the dual-route model of morphology. *Lingua, 110,* 281–298.

Ellis, N. C. (2002). Frequency effects in language processing: review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24,* 143–188.

Ellis, N. C., & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning, 60(2),* 85–108.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58,* 787–814.

Goodall, G. (2008). Poverty/richness of the stimulus in instructed L2 acquisition of Spanish. Presented at the Hispanic Linguistics Symposium, Université de Laval, Quebec.

Gor, K., & Chernigovskaya, T. (2004). Formal instruction and the acquisition of verbal morphology. In A. Housen & M. Pierrard (eds.), *Investigations in instructed second language acquisition*, pp. 131–163. Berlin: Walter de Gruyter.

Gor, K., & Cook, S. (2010). Nonnative processing of verbal morphology: In search of regularity. *Language Learning, 60,* 88–126.

Hahn, U., & Nakisa, R. (2000). German inflection: Single route or dual route? *Cognitive Psychology, 41,* 313–360.

Hualde, J. I., Olarrea, A., Escobar, A. M., & Travis, C. (2010). *Introducción a la lingüística hispánica*. New York: Cambridge University Press.

Kilborn, K., & Ito, T. (1989). Sentence processing strategies in adult bilinguals. In B. MacWhinney & E. Bates (eds.), *The crosslinguistic study of sentence processing*, pp. 257–291. Cambridge: Cambridge University Press.

Klein, W. (1994). *Time in language*. London: Routledge.

LaBrozzi, R. (2009). Processing of lexical and morphological cues in a study abroad context. Ph.D. dissertation, Pennsylvania State University.

Lee, J. F., Cadierno, T., Glass, W. R., & VanPatten, B. (1997). The effects of lexical and grammatical cues on processing past temporal reference in second language input. *Applied Language Learning, 8,* 1–23.

Linares, R., Rodríguez-Fornells, A., & Clahsen, H. (2006). Stem allomorphy in the Spanish mental lexicon: Evidence from behavioral and ERP experiments. *Brain and Language, 97,* 110–120.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. Doughty & J. Williams (eds.), *Focus on form in classroom second language acquisition*, pp. 15–41. Cambridge: Cambridge University Press.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19,* 37–66.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development, 43,* 1–123.

MacWhinney, B. (2007). A unified model. In N. C. Ellis & P. Robinson (eds.), *Handbook of cognitive linguistics and second language acquisition*, pp. 341–371. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (2011). The logic of the Unified Model. In S. Gass & A. Mackey (eds.), *Handbook of second language acquisition*, pp. 85–112. New York: Routledge.

Marcus, G., Brinkmann, U., Clahsen, H., Wiese, R., & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29,* 189–256.

McClelland, J. L., & Patterson, K (2002). Rules or connections in past tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6,* 465–472.

Nagata, N., & Swisher, M. V. (1995). A study of consciousness-raising by computer: The effect of metalinguistic feedback on SLA. *Foreign Language Annals, 28,* 336–347.

Nicoladis, E., & Paradis, J. (2012). Acquiring regular and irregular past tense morphemes in English and French:

Evidence from bilingual children. *Language Learning, 62,* 170–197.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language, 26,* 619–653.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50,* 417–528.

Pinker, S. (1999). *Words and rules*. New York: Basic Books.

Presson, N., & MacWhinney, B. (in press). Learning grammatical gender: The use of rules by novice learners. *Applied Psycholinguistics*.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16,* 103–133.

Rodríguez-Fornells, A., Münte, T., & Clahsen, H. (2002). Morphological priming in Spanish verb forms: An ERP repetition priming study. *Journal of Cognitive Neuroscience, 14,* 443–454.

Rosa, E. M., & Leow, R. P. (2004). Computerized task-based instruction in the L2 classroom: The effects of explicitness and type of feedback on L2 development. *Modern Language Journal, 88,* 192–217.

Rossomondo, A. E. (2003). The role of lexical temporal indicators in the incidental acquisition of the Spanish future tense. Ph.D. dissertation, Indiana University.

Sagarra, N. (2008). Working memory and L2 processing of redundant grammatical forms. In Z. Han (ed.), *Second language processing and instruction: Broadening the scope of inquiry*, pp. 142–159. Clevedon: Multilingual Matters.

Santilli, M. (1996). Teacher talk and written materials in the Spanish as a second language classroom and their importance as input for second language acquisition. Ph.D. dissertation, University of Illinois at Urbana–Champaign.

Sanz, C. (1999). What form to focus on? Linguistics, language awareness, and the education of L2 teachers. In J. Lee & A. Valdman (eds.), *Meaning and form: Multiple perspectives*, pp. 3–23. Boston, MA: Heinle & Heinle.

Say, T., & Clahsen, H. (2002). Words, rules and stems in the Italian mental lexicon. In S. Nooteboom, F. Weerman & F. Wijnen (ed.), *Storage and composition in the language faculty*, pp. 93–130. Dordrecht: Kluwer.

Schinke-Llano, L. (1986). Foreigner talk in joint cognitive activities. In R. Day (ed.), *Talking to learn: Conversation in second language acquisition*, pp. 99–117. Rowley, MA: Newbury House.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing, II: Perceptual learning, automatic attending and a general theory. *Psychological Review, 84,* 127–190.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60,* 263–308.

Starren, M. (2001). *The second time: The acquisition of temporality in Dutch and French as a second language.* Utrecht: LOT.

Stemberger, J., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition, 14,* 17–26.

Stump, G. (2001). *Inflectional morphology*. Cambridge: Cambridge University Press.

Tkachenko, E., & Chernigovskaya, T. (2010). Input frequencies in processing of verbal morphology in L1 and L2: Evidence from Russian. *Oslo Studies in Language, 2,* 281–318.

Ullman, M. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92,* 231–270.

Zollo, M. (1993). *Berlitz Spanish verb handbook*. Princeton, NJ: Berlitz.