The Synergistic Benefits of Systematic and Random Interleaving for Second Language Grammar Learning

Steven C. Pan[a,b], Jarrett Lovelett[b], Vicky Phun[b1], and Timothy C. Rickard[b]

[a]University of California Los Angeles, USA;
[b]University of California San Diego, USA;
[1]Michigan State University, USA

Abstract

Repeatedly switching between a series of to-be-learned topics, also called interleaved practice (*interleaving*), can improve learning over traditional, one-topic-at-a-time blocked practice (*blocking*). We investigated whether interleaving's benefits for second language learning are facilitated by *random schedules*, wherein training trials follow unpredictable patterns, or *systematically alternating schedules*, wherein trials are predictably sorted. Students learned to conjugate Spanish verbs in the *preterite* and *imperfect* tenses and then took a 48-hr. delayed verb conjugation test. A consistently random (Experiment 1) or systematically alternating schedule (Experiment 2) did not improve learning versus blocking. However, the combination of both types – systematic alternation for study trials and randomization for practice trials – enhanced learning (Experiments 3-4). Thus, neither interleaving schedule alone appears to be sufficient; for verb conjugation skills and likely other materials involving study and problem-solving practice, both are needed. Interleaving's benefits are therefore impacted by the alignment between training schedule and task type.

Keywords: interleaving; random practice; systematically alternating schedules; training sequence order; verb conjugation; foreign language learning; second language learning

Author Note

This manuscript was accepted for publication in *Journal of Applied Research in Memory and Cognition* on July 17, 2019. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. The final article will be available, upon publication, through the aforementioned journal. In a departure from APA style, it has been reformatted here for readability. Content remains unchanged.

General Audience Summary

Although learners traditionally focus on one topic at a time, an approach known as *blocked practice* (i.e., blocking), recent research suggests that switching between multiple topics during learning, or *interleaved practice* (i.e., interleaving), can be more effective. One such example involves learning to conjugate verbs (modifying verbs to reflect grammatical tense). We investigated whether the type of interleaving schedule – *random* or *systematically alternating* – impacts interleaving's benefits for verb conjugation skills. In a random schedule, learners switch between topics in an unpredictable manner, whereas in a systematically alternating schedule, learners switch between topics in a logically-organized, repeating sequence. Undergraduate students learned to conjugate verbs in the Spanish *preterite* and *imperfect* tenses using interleaving or blocking. A delayed test of verb conjugation ability occurred 48 hrs. later. The use of a consistently random or systematically alternating schedule did not enhance learning relative to blocking. However, the use of systematic alternation for study trials (in which learners were presented with information about each grammatical tense in a manner that switched tense on every trial) and randomization for practice exercises (in which learners attempted to conjugate verbs in either tense on any trial) resulted in a substantial interleaving benefit. Thus, in order for interleaving to enhance learning for this task, adopting systematic and random interleaving for studying and practicing, respectively, appears to be ideal. This approach can manifest in various ways, including lessons which follow a predetermined sequence and practice questions that are electronically or manually shuffled.

The Synergistic Benefits of Systematic and Random Interleaving for Second Language Grammar Learning

Instructors and students regularly schedule learning activities in various ways. For instance, they may focus on one subject per day, tackle easier topics first, or review before practice exercises. Many, however, would be surprised to learn that a seldom-used scheduling technique, *interleaved practice* (i.e., interleaving), can be more efficacious than the nearly ubiquitous method of *blocked practice* (i.e., blocking). Whereas blocking involves learning one topic at a time before moving to the next (e.g., given topics A, B, and C, an AAABBBCCC schedule), interleaving involves repeatedly switching between two or more topics as they are learned (e.g., an ABCABCABC schedule). This method commonly results in increased initial difficulty yet improved long-term retention and transfer performance. As such, interleaving may be a "desirable difficulty" (Bjork, 1994; Pan & Bjork, 2019) that can benefit learners. The *interleaving effect* has been compellingly demonstrated for inductive visual category learning (e.g., Kornell & Bjork, 2008), mathematics skill learning (e.g., Rohrer & Taylor, 2007), and a wide range of motor skills (e.g., Shea & Morgan, 1979), among others. For those higher level tasks, the effect often appears to be the most robust when target materials involve categories that are highly similar with one another; in such cases, interleaving appears to facilitate a process of comparison and contrast between temporally adjacent items (for reviews see Carpenter, 2014; Carvalho & Goldstone, 2015; Kang, 2017; Rohrer, 2012).

Despite interleaving's promise in a growing number of research studies to date, no consensus has yet been reached on the generality of its benefits for educationally-relevant materials (for discussion see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Moreover, the issue of exactly how to interleave a series of to-be-learned topics – in other words, the interleaving *schedule type* that should be used – has yet to be resolved. Should topics be arranged in a random or pseudorandom fashion, as occurs in many studies? Or should they follow a systematic pattern, such as alternating back-and-forth between two skills (e.g., Smith, Gregory, & Davies, 2003) or serially rotating among three or more in a predictable manner (e.g., Gagné, 1950)? Yet other types such as serial-to-random *hybrid* scheduling – wherein hybrid refers to combining more than one schedule type – have also been attempted (e.g., Porter & Magill, 2010). A related issue is whether the schedule type should vary according to whether learners are (a) studying or answering questions, or (b) at early or advanced stages of learning. To inform effective uses of interleaving, more research on these issues is needed.

Random Versus Systematically Alternating Schedules

Many interleaving researchers have theorized that the technique's benefits stem from the unpredictability of successive training trials (e.g., Bjork, 1999; Carpenter, 2014), which is an intrinsic characteristic of randomization. Relative to blocking, wherein information for a target category consistently applies across successive trials and can be held in working memory, a random interleaving schedule (e.g., an ABCACBCBA schedule) may generate more elaborate memories due to (a) learners' inability to reuse information in working memory across successive trials, which forces greater processing on each, and (b) more opportunities to compare

the topics being studied (i.e., "discriminative contrast"). Additionally, if training involves randomized practice problems and especially if there is a constraint against any problem category appearing multiple times in succession, then (c) a strategy of reusing solutions across successive trials is rendered ineffective, as is (d) the use of any feedback from a given trial as a hint for the next (Lee & Simon, 2004). Consequently, randomly interleaved practice exercises require more memory retrieval and may yield better learning (Carpenter & Mueller, 2013; Pan, 2015; Rohrer, 2012). However, some researchers have also suggested that randomization can be too confusing and generates cognitive "overload," particularly in novice learners (Wulf & Shea, 2002). This premise underlies suggestions to avoid random interleaving or at least reserve it until after a basic level of proficiency is achieved (e.g., Magill & Hall, 1990; Wulf & Shea, 2002).

Systematically alternating interleaving schedules (e.g., an ABABAB or ABCABC schedule) may also foster conditions (a) and (b), although the predictability of the training sequence may impact (c) and (d). Such schedules have another potential advantage in that materials can be presented in a logical or hierarchical manner, which itself may improve learning (e.g., $A_1B_1A_2B_2A_3B_3$, wherein subscripts refer to category exemplars being ordered in a meaningful way). Studies of systematically alternating schedules, largely involving motor skills, have shown benefits relative to blocking (e.g., Bortoli, Robazza, Durigon, & Carra, 1992), but not universally (e.g., Smith et al., 2003). Direct comparisons of systematically alternating versus random schedules have also found that both yield similar benefits relative to blocking (e.g., Bortoli et al., 1992; Lee & Magill, 1983; cf. Travlos, 2010). Those results have prompted theorizing that both interleaving methods enhance learning by forcing learners to reconstruct solution strategies across trials (Lee & Magill, 1983; Travlos, 2010).

The two most prominent accounts of interleaving effects, namely the discriminative contrast (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2014; Kang & Pashler, 2012) and distributed practice (e.g., Carpenter, 2014; Kornell & Bjork, 2008) hypotheses, are generally compatible with random and systematically alternating schedules. These accounts ascribe interleaving's benefits to a process of comparing category exemplars and the spacing effect (Ebbinghaus, 1885), respectively. Indeed, both schedule types afford opportunities to compare and contrast categories, as well as distribute learning (i.e., spacing) across nonadjacent trials. If those two factors are the most critical to interleaving's benefits, then the two schedule types are likely to have similar efficacy. However, with respect to spacing and the unpredictability of not knowing what comes next – as may be especially important for learning with practice problems – random interleaving always provides both, systematic interleaving always provides spacing, and blocked practice provides neither.

The Interleaving Effect for Learning Grammar Skills

Interleaving has recently been discovered to be a potent enhancer of second language (L2) verb conjugation skills. Pan, Tajran, Lovelett, Osuna, and Rickard (2018) trained undergraduate students to conjugate verbs (i.e., modify verbs to reflect grammatical tense) in the *preterite* and *imperfect* Spanish past tenses. Both tenses have defining rules (see Table 1), but L2 learners require substantial training and practice to be able to tease apart surface-level

ambiguities and discriminate between them (Castañeda, 2011; Westfall & Foerster, 1996). Accordingly, Pan et al. designed a three-phase training paradigm that begins with *study of the rules* for when to use a given tense, followed by *study of the verb suffixes* that are used to conjugate verbs in that tense, and then *practice of the entire verb conjugation process* (i.e., which requires identification of tense, recall of the suffix, and conjugation of the verb). The initial training phases always involved blocking. However, when practice trials were randomly interleaved between tenses (as opposed to blocked by tense), and additional randomized practice trials were administered in a second session, greater improvements and better retention versus a purely blocked schedule were observed. In other words, a blocked-to-interleaved hybrid schedule (i.e., blocking first, interleaving second) enhanced learning. The authors attributed this result to joint contributions of discriminative contrast and spacing, as well as the possibility that a blocked-to-interleaved schedule is easier than a purely interleaved schedule (Pan et al., 2018; Sorensen & Wolz, 2016; Wulf & Shea, 2002; Yan, Soderstrom, Seneviratna, Bjork, & Bjork, 2017), which involves separate issues from those examined in the current work.

[Table 1 around here]

Although Pan et al. (2018) discovered a substantial interleaving benefit in the domain of L2 learning, among questions left unresolved are whether such benefits stem specifically from the use of a random schedule, and whether those benefits would be maintained if (a) study phase trials were also interleaved and if (b) interleaving was implemented consistently throughout a single training session (i.e., entirely interleaved rather than a blocked-to-interleaved hybrid schedule). The current study addressed these issues.

The Current Study

Across four experiments, we investigated the effects of interleaving in the form of random versus systemically alternating schedules for the acquisition of Spanish verb conjugation skills. Crucially, unlike Pan et al. (2018), we compared both types of interleaving in their "pure" form (Experiments 1 and 2), with consistent implementation across an entire training session, as well as a hybrid variant that combined systematic and random interleaving schedules (Experiments 3 and 4), relative to a blocked condition. In all cases we assessed retention and transfer via a 48-hr. delayed test of verb conjugation ability.

Experiment 1

The first experiment investigated the effects of a *fully randomized interleaving schedule* wherein study trials (involving tense rules and verb suffixes) and practice trials (involving conjugating verbs) switch between grammatical tenses in an unpredictable manner within each training phase, and once verb suffixes are introduced, switch between them unpredictably as well. If interleaving's benefits stem from unpredictability across all stages of learning, then this schedule should improve learning. We compared randomization against a blocked schedule that also incorporated randomized study and practice trials, but only for one tense at a time.

Method

Participants. In this and subsequent experiments, undergraduate students recruited from

the University of California, San Diego participated in exchange for course credit. All were fluent English speakers with no prior Spanish language experience. The target sample size, following Pan et al. (2018), was 42 per group. One-hundred and eleven students participated in Experiment 1. Data from 8 students that did not complete the second session were excluded, yielding a final sample of 103 participants (*interleaved* group, $n = 49$; *blocked* group, $n = 54$).

Materials. All materials, including instructions, examples, practice trials, and delayed test questions, were drawn from Pan et al. (2018). Selected materials are presented in Table 1 and a more extensive discussion of those materials is available in the source article. These materials included 16 sentences as Phase 1 practice trials, 6 sentences as Phase 2 copy trials, 18 fill-in-the-blank questions as Phase 3 practice trials, and 21 short answer and 21 multiple-choice questions for the delayed test. The materials were balanced to have equal numbers representing each tense for Phase 1 practice trials, one sentence per tense-suffix combination for Phase 2 copy trials, three practice questions per tense-suffix combination in Phase 3, and equal numbers of each rule-suffix combination within each tense on the delayed test.

Design. The design was adapted from Pan et al. (2018) but with (a) differently ordered sequences of study and training trials and (b) interleaving or blocking throughout the entire training session. Each participant was randomly assigned to an *interleaved* or *blocked* group (between-participants). Both groups completed a training session and a 48-hr. delayed test, both of which were self-paced. The dependent measure was proportion correct on the short answer and multiple-choice portions of the delayed test.

Procedure. The three-phase training procedure for all experiments is overviewed in Table 2 and is schematized, for Experiment 1, in the upper section of Figure 1 (including examples of study and practice trial sequences). Details of both experiment sessions are summarized as follows.

[Table 2 and Figure 1 around here]

Training session. All participants in this and subsequent experiments completed the following three training phases. In Phase 1 (*tense rules*), the four defining rules per tense (Table 1) were presented for study, one at a time. Rule study was followed by 2 cycles of 8 practice trials per tense (i.e., on the basis of the rules, making a yes or no judgment as to whether each presented sentence exemplified that tense or not). In Phase 2 (*verb suffixes*), participants learned the suffixes that are used to conjugate a subset of regular verbs wherein the root verb (i.e., infinitive) has the common "-ar" ending and the pronoun is the equivalent of "I", "you", or "we" (Table 1). There were 3 suffixes per tense. Each suffix was presented individually on a single study trial and immediately followed by a single copy trial (i.e., retyping a root verb with that suffix appended). In Phase 3 (*verb conjugation practice*), participants completed 9 practice trials per tense (conjugating new "-ar" root verbs into new sentences that had an "I", "you", or "we" pronoun). A summary slide was presented after each phase and correct answer feedback was provided after each practice or copy trial.

All participants within the blocked and interleaved groups in this and the subsequent experiments viewed the same instructional content (e.g., identical rules, verb suffixes, examples, and so on) and completed the same total number of training trials. The trials were also identical

for each tense. Further, as noted earlier, the materials were balanced such that the examples and trials that each participant viewed represented each tense and each suffix with equal frequency. The arrangement of instructional content and trials within each phase, however, differed between groups. This arrangement is summarized in Table 2 and described next.

Blocked group. Participants in the blocked group first completed Phases 1-3 for one tense (wherein they learned to conjugate verbs in that tense), with each phase proceeding exactly as previously described. Next, they repeated Phases 1-3 but for the other tense. To enable comparisons of the blocked and interleaved groups, the blocked group experienced a form of randomization within each phase, but without any alternation between tenses. That is, each phase featured random sequences (i.e., random trial order without replacement) such that for Phase 1, the four rules for the tense being learned were presented once each in any order, with subsequent practice trials for that tense also presented in any order; for Phase 2, the three suffixes for that tense were presented in any order; and in Phase 3, the practice trials for that tense were presented in any order. Crucially, despite trial-level randomization, the tense remained constant in each phase in the blocked group.

Interleaved group. Participants in the interleaved group learned to conjugate verbs in both tenses simultaneously (i.e., Phases 1-3 involved both tenses combined rather than each occurring separately) and with unconstrained randomization that enabled repeated switching between tenses (and, where applicable, verb suffixes) throughout each phase. The same total number of study and training trials were allotted per tense as in the blocked group. In Phase 1, participants studied the full set of 8 defining rules for the two tenses, followed by two cycles of 16 practice trials (8 per tense). In Phase 2, they learned the full set of 6 verb suffixes (one study and copy trial per suffix) for the two tenses. In Phase 3, they completed 18 practice trials (9 per tense). Crucially, random sequences (i.e., random trial order without replacement) were used within each phase such that for Phase 1, the 8 rules across two tenses were presented once each in any order, with subsequent practice trials involving either tense and presented in any order; for Phase 2, the 6 suffixes across two tenses were presented in any order; and in Phase 3, the practice trials for both tenses were presented in any order. Unlike the blocked group, the random sequences in the interleaved group allowed for alternation between tenses within each phase, and within the second and third phases, alternation between any of the 6 suffixes that are used to conjugate verbs in relation to three pronouns across the two tenses.

Metacognitive judgments. After training, participants answered two metacognitive questions. As in Pan et al. (2018), these questions involved global judgments of learning ("How well did you learn Spanish verb conjugation today?") and difficulty ("How easy was it to learn Spanish verb conjugation?"), respectively, and were answered using a five-level scale. Reporting of all metacognitive data will occur after Experiment 4.

Delayed test. Two days after training, participants completed a test that was identical for both groups and entailed conjugating Spanish verbs in the preterite and imperfect tenses. Participants first answered 21 short answer questions, one at a time and in random order. These questions resembled those from Phase 3 of the training session and involved the presentation of a Spanish fill-in-the-blank question, its English translation, and a Spanish root verb. On each trial,

participants were required to type the conjugated verb into a textbox. Next, participants answered 21 multiple-choice questions. These questions were largely identical to the short answer questions except for the provision of 6 answer choices (drawn from 2 tenses x 3 pronouns). No feedback was provided.

The use of two test formats served to combine the methods featured in Pan et al. (2018), in which short answer tests, which are more ecologically valid, are more sensitive to detecting any interleaving effects, and multiple-choice tests, which are less difficult, avoid floor effects.

Results

Training. Descriptive statistics (mean proportion correct and *SE*) for Phase 1 practice trials, Phase 2 copy trials, and Phase 3 practice trials are reported for all experiments in Table 3. Consistent with a large body of research on interleaving (e.g., Carpenter, 2014; Schmidt & Bjork, 1992), performance in the interleaved group was lower than the blocked group throughout the training session. This apparent deficit was the most pronounced during practicing of the target skill in Phase 3.

[Table 3 around here]

Delayed test. Short answer and multiple-choice results for all experiments are displayed in Panels A and B, respectively, of Figure 2. Each result was analyzed using a one-way Analysis of Variance (ANOVA) with a factor of Group (Interleaved vs. Blocked). In this and all subsequent analyses, α was set at .05. For short answer questions, there was no significant effect of Group, $F(1,101) = 0.30$, $MSE = 0.03$, $p = .58$, $\eta_p^2 < 0.01$. For multiple-choice questions, there was also no significant effect of group, $F(1,101) = 1.00$, $MSE = 0.07$, $p = .32$, $\eta_p^2 = 0.01$. Those analyses confirm a pattern that is evident upon inspection of Figure 2: for both test formats, performance in the blocked and interleaved groups was not statistically different.

[Figure 2 around here]

Discussion

The results of Experiment 1 challenge an unpredictability-focused account of interleaving effects as well as, more broadly, the premise that random interleaving between tenses, and between suffixes involving those tenses, enhances verb conjugation skills. When randomization was deployed throughout all training phases, learning in the interleaving group was no better, and numerically worse, than that in the blocked group. That pattern held despite learners' inability to predict which tense or verb suffix was being studied or practiced on each successive trial. Such unpredictability might have yielded more distributed practice and better Spanish verb conjugation skills relative to the blocked group, and yet no improvements consistent with those premises were found.

Experiment 2

The second experiment investigated the effects of a *systematically alternating interleaving schedule* wherein grammatical tense switches on every successive study or practice trial. This schedule (a) maximizes opportunities for comparison beyond that of a purely random schedule and (b) enables materials to be logically ordered within each phase. As an example of

(a), presentation of each tense's rules in an alternating sequence (e.g., preterite rule #1: "For past actions that had a specific and clear beginning"; imperfect rule #1: "For past actions that lack a specific and clear beginning") enables a more direct comparison of those contrasting rules than a random schedule with other rules interspersed. Regarding (b), logically ordered study trials might yield more integrated learning of tense rules and suffixes. Additionally, by eliminating the need to identify tense, systemically alternating practice trials might improve focus on the mechanics of verb conjugation (although, as previously noted, tense discrimination is a crucial component of the process). We compared systematically alternating interleaving against a blocked schedule that also featured systematically ordered trials, but only for one tense at a time.

Method

Participants. Ninety-four students participated in Experiment 2. Data from eleven students were excluded due to noncompliance with directions or not completing both sessions, yielding a final sample of 83 participants (*interleaved* group, $n = 43$; *blocked* group, $n = 40$).

Materials, design, and procedure. These were identical to the prior experiment except for systematic trial ordering within each training phase (see middle section of Figure 1). For the blocked group, all 4 rules per tense were presented in sequential order in Phase 1; every other Phase 1 practice trial complied with the rules for that tense; verb suffixes were presented and copied in "I", "you," and "we" order in Phase 2; and Phase 3 practice trials were grouped by pronoun. A corresponding pattern was used for the interleaved group: all 4 rules per tense were presented in sequential order and on every other study trial in Phase 1; subsequent practice trials continuously alternated between tense; the "I", "you," and "we" suffixes per tense were learned in Phase 2 in alternating fashion (e.g., via the pattern of "I"-preterite, "I"-imperfect, "you"-preterite, etc.); and Phase 3 trials switched tense on every trial (and followed the same pattern as the preceding phase). As such, with respect to each dimension being learned (i.e., tenses and the 6 verb suffixes used to conjugate for sentences involving three pronouns), all three phases in the interleaved group involved systematically alternating interleaving between tense, whereas Phases 2 and 3 also featured interleaving between verb suffixes. At the conclusion of training, interleaved group participants were asked whether they had noticed the systematic alternation between tenses, and if they had, at what point in the session it became evident.

Results

Training. Mirroring the patterns observed in the preceding experiment, performance in the interleaved group was lower than the blocked group throughout the training session, and again this difference was the most pronounced in Phase 3 (Table 3). Ninety percent of participants in the interleaved group reported that they were aware of the systematic alternation between tenses; of those participants, 47% became aware during Phase 1, 34% during Phase 2, and 18% during Phase 3.

Delayed test. Two ANOVAs identical to those performed for Experiment 1 revealed no significant effect of Group for short answer questions, $F(1,81) = 0.14$, $MSE = 0.02$, $p = .71$, $\eta_p^2 < 0.01$, as well as for multiple-choice questions, $F(1,81) = 1.05$, $MSE = 0.07$, $p = .31$, $\eta_p^2 = 0.01$. These results are remarkably similar to those of the preceding experiment (compare panels A and

B of Figure 2). For both test formats, performance in the interleaved group was not statistically different from that of the blocked group, and again slightly worse numerically.

Discussion

The results of Experiment 2 reveal that an interleaving schedule that incorporates systematic alternation for all study and practice trials confers no advantage over blocking. This lack of an interleaving benefit occurred despite increased opportunities for comparison and the use of logically ordered sequences. It would therefore appear that these supposed advantages of systematic interleaving are less potent than previously hypothesized, at least in the current task domain.

Experiment 3

The prior experiments illustrate instances of consistently random or systematic interleaving schedules that do not enhance learning. For the third experiment we switched to a hybrid schedule wherein the type of interleaving reflects whether participants are engaged in study or practice (for related discussions see Magill & Hall, 1990; Wulf & Shea, 2002). Specifically, we used systematic alternation for study trials and randomization for practice trials. We theorized that it may be helpful to present information from multiple categories in a logical and predictable pattern on study trials (emphasizing an organized process of comparison and contrast, which randomization can degrade), whereas an unpredictable order might be helpful on practice trials (to prevent the reuse of solutions according to a fixed practice trial pattern and to force learners to engage in memory retrieval processes). We compared this combination of the two schedule types – henceforth described as a *systematic-to-random hybrid interleaving schedule* – against a blocked schedule that featured systematically ordered study and randomized practice trials, but only for one tense at a time.

Method

Participants. One-hundred and three students participated in Experiment 3. Data from 10 participants were excluded due to not completing the second session or technical difficulties, yielding a final sample of 93 participants (*interleaved* group, $n = 44$; *blocked* group, $n = 49$).

Materials, design, and procedure. These were identical to the preceding experiments excepting study and practice trial ordering within each training phase (see lower section of Figure 1). Specifically, all study trials were presented using systematic alternation, as in Experiment 2, and all practice trials were presented using randomization, as in Experiment 1. Due to a program glitch, one delayed test trial was not presented correctly to 9 participants; those trials were dropped from the analyses.

Results

Training. As in the prior experiments, performance in the interleaved group was lower than the blocked group throughout the training session, and again this difference was the most pronounced in Phase 3 (Table 3).

Delayed test. Two ANOVAs identical to those performed for the preceding experiments

yielded a significant effect of Group for short answer questions, $F(1,91) = 7.89$, $MSE = 0.63$, $p = .006$, $\eta_p^2 = 0.08$, as well as for multiple-choice questions, $F(1,91) = 7.75$, $MSE = 0.41$, $p = .007$, $\eta_p^2 = 0.079$. These results reflect a substantial interleaving effect for both types of delayed test questions (see panels A and B of Figure 2).

Discussion

The results of Experiment 3 suggest that systematic alternation and randomization impacts study and practice trials differently. Achieving an optimal alignment between schedule and trial type, as apparently occurred with the adoption of a systematic-to-random hybrid schedule, appears to be necessary for an interleaving advantage to manifest after a single training session.

Experiment 4

For the final experiment we attempted to replicate Experiment 3's results but with one design change: the number of Phase 3 trials was doubled. This experiment served to investigate whether the interleaving advantage would replicate under extended training conditions.

Method

Experiment 4 was preregistered at AsPredicted.org (https://aspredicted.org/9j9ra.pdf).

Participants. Data for both groups was simultaneously collected at two sites (the same university as previously and a sister institution (University of California, Los Angeles) and combined for analysis. The participant pools at both sites drew from students taking the same types of classes and represented comparable academic backgrounds and aptitude levels. Ninety-four students participated in Experiment 4. Data from three participants were excluded due to experimenter error or technical difficulties, yielding a final sample of 91 participants (*interleaved* group, $n = 42$; *blocked* group, $n = 49$).

Materials, design, and procedure. These were identical to Experiment 3 (see lower section of Figure 1 for examples) excepting the use of twice as many Phase 3 practice trials (18) as in prior experiments (9). The additional 9 practice items resembled the originally used set of practice items (e.g., drawing from the same set of verb suffixes), but featured new sentences and different root verbs.

Results

Training. Similar to prior experiments, performance in the interleaved group was lower than in the blocked group throughout much of the training session (that performance gap was eliminated in Phase 2), and again this difference was the most pronounced in Phase 3 (Table 3).

Delayed test. Two ANOVAs identical to those performed for the preceding experiments revealed a significant effect of Group for short answer questions, $F(1,89) = 4.71$, $MSE = 0.34$, $p = .03$, $\eta_p^2 = 0.05$, but not for multiple-choice questions, $F(1,89) = 1.51$, $MSE = 0.08$, $p = .22$, $\eta_p^2 = 0.02$. These results reflect a significant interleaving advantage for short answer but not multiple-choice questions on the delayed test (see panels A and B of Figure 2), although there was a numerical interleaving advantage for that test type as well.

Discussion

The results of Experiment 4 reaffirm the conclusion that a systematic-to-random hybrid schedule can generate a substantial interleaving advantage on a 48-hr. delayed test. Increasing the number of Phase 3 trials did not appear to alter the interleaving benefit for short answer questions, which remained potent, although a comparable advantage for multiple-choice questions was not observed. It is possible that extended practice may have enabled the blocked group to retain more knowledge than before, thus reducing the performance gap on the easier portion of the test. Overall, these results provide further evidence that using different schedules for study and practice trials can substantially impact the resulting learning benefits.

Training Duration and Learning Rate Analyses

Despite having the same total number of study and training trials, the interleaved group took longer to complete the training session in all four experiments. That disparity was not unexpected given the greater difficulty of repeatedly switching between tenses (or, conversely, the greater ease of learning one tense at a time in a blocked schedule). The mean time (*SE*) in min for the interleaved and blocked groups, respectively, was 20.1 (0.54) vs. 16.7 (0.30) in Experiment 1, 18.2 (0.55) vs. 16.3 (0.34) in Experiment 2, 19.9 (0.52) vs. 16.2 (0.47) in Experiment 3, and 24.8 (0.70) vs. 20.2 (0.41) in Experiment 4. To determine whether that greater training duration could in principle account for differences on the delayed test, we computed retention rate estimates for each participant wherein delayed test proportion correct was divided by the corresponding training duration (cf. Pan et al., 2018). That analysis, conducted separately for short answer and multiple-choice questions, revealed a significantly lower retention rate for interleaved group participants on both question types in Experiment 1 (Ms ≥ 0.02 vs. Ms ≥ 0.03, ts ≥ 2.11, ps $\leq .04$, ds ≥ 0.42) and on multiple-choice questions in Experiment 2 ($M = 0.03$ vs. $M = 0.04$, $t(81) = 2.26$, $p = .03$, $d = 0.49$), plus no significant rate differences for both question types in Experiments 3 and 4 (ps ≥ 0.08). Thus, in terms of learning per unit of time, the results indicate that a consistently random or systematic schedule was less efficient than blocking, whereas for hybrid schedules, the learning rate was similar to the blocked group and yielded better delayed test performance. These results raise an issue, not just for the current study but for the interleaving literature more generally, in that although it seems likely that systematic-to-random hybrid interleaving is responsible for the superior performance of the interleaved groups in Experiments 3 and 4, we cannot rule out the possibility that the results are due in part to increased time on task. However, our finding that systematic-to-random hybrid interleaving is superior to non-hybrid (entirely systematic or random) interleaving appears to be immune to that possibility.

Metacognitive Judgments of Difficulty and Learning

In all four experiments, participants in the interleaved group gave higher difficulty ratings to the training session (67-80% chose “moderately difficult” to “very difficult”) than the blocked

group (61-72% chose "easy" to "very easy"). This difference was statistically significant in each case (X^2 test for independence, ps < .001). Further, participants in the interleaved group rated their mastery of verb conjugation skills at the end of the training session as lower (66-78% chose "poor" to "average") than participants in the blocked group (49-55% chose "good" to "excellent"). This difference was also statistically significant in each case (X^2 test for independence, ps ≤ .003). Thus, in spite of the divergent delayed test results, all forms of interleaving in this study caused participants to experience a heightened sense of difficulty and to more harshly assess their learning. These findings are consistent with prior evidence that learners rate blocked categories as better learned than interleaved categories (Yan, Bjork, & Bjork, 2016) and judge blocking as more effective than interleaving (e.g., Kornell & Bjork, 2008), despite empirical evidence to the contrary. Moreover, the improved accuracy for judgments of learning following interleaving is analogous to a similar pattern for retrieval practice versus restudying (e.g., Tullis, Finley, & Benjamin, 2013).

Analysis of Delayed Test Errors

For potential insights into the learning of verb conjugation skills, in a post-hoc analysis we examined the number and types of delayed test errors (Figure 3). The three error types that were analyzed involved incorrect verb conjugations that (a) correctly match the given pronoun but not the tense (*tense* errors), (b) correctly match the tense but not the given pronoun (*suffix* errors), or (c) incorrectly match both the pronoun and tense (*both* errors). All multiple-choice and correctly spelled short answer errors could be categorized as such (misspellings introduced ambiguity that could not be clearly categorized due to the similarity of different conjugations and were not analyzed). Inspection of the figure and standard errors reveals that the most obvious between-group differences involved tense errors. Specifically, in Experiments 3 and 4, interleaving yielded fewer tense errors, a finding that is consistent with the premise that systematic-to-random hybrid interleaving improves the ability to identify tense. The entirely systematic interleaving schedule in Experiment 2 was less effective for learning tense, at least according to the short answer data. Interleaving may also have yielded poorer learning of suffixes, particularly in the earlier experiments, but any between-group differences were relatively slight, as was the case for both errors.

[Figure 3 around here]

General Discussion

Rather than enhancing learning as a result of the unpredictability of randomization or the organization and maximal contrast afforded by systematic alternation, the current study reveals that interleaving's efficacy – and, by implication, the cognitive mechanisms that interleaving engages – vary as a function of schedule type and whether learners are engaged in study or practice. Consistent with this conclusion, a homogenous approach of complete randomization (Experiment 1) or systematic alternation (Experiment 2) yielded learning that was no better than

that following blocking. Instead, a systematic-to-random hybrid schedule was necessary for an interleaving benefit to emerge (Experiments 3-4). The generalizability of that finding to other tasks that involve studying and then practicing should be explored in future work.

Building on earlier research, the results of Experiments 3 and 4 indicate that multi-session interleaving and a transition from initial blocked to interleaved practice, as occurred in Pan et al. (2018), is not required to yield better learning of verb conjugation skills relative to blocking. Rather, a systematic-to-random hybrid schedule generated learning benefits on a 2-day delayed test that were comparable to those found on a one-week delayed test after multi-session training with blocked-to-interleaved hybrid schedules (short answer results, *d*s = 0.46, 0.58 vs. 0.79 in Pan et al., 2018). Experiments 3 and 4 also address prior concerns (e.g., Wulf & Shea, 2002) that interleaving from the onset of training may be detrimental: with systematic-to-random hybrid interleaving, no such deficits are observed (see also Yan et al., 2017). That observation further establishes the importance of achieving an alignment of schedule with the type of task—that is, study or practice—that is being performed.

Implications for Theories of Interleaving Effects and Schedule Types

The current study reveals that random and systematically alternating interleaving schedules can have an inverse relationship: randomization is a boon to practice trials and the bane of study trials, whereas the reverse occurs for systematic alternation. Further, it appears that interleaving's effects cannot be explained using a single cognitive mechanism. For studying, the ability to engage in processes involving discriminative contrast across successive category exemplars in a logical and organized fashion can be vital. Such processes would appear to be crucial when a series of opposing rules need to be learned and are presented in an alternating manner, as occurred in the present experiments (with a caveat that not all learning materials have these characteristics). The benefits of systematic alternation for studying may be comparable to that of text coherence for learners' memory and comprehension of unfamiliar prose (e.g., Kintsch, 1994). In comparison, for practicing, both spaced retrieval and unpredictability are likely to be crucial factors. Randomized interleaving provides both.

In multiple domains, a prominent precondition for interleaving effects is that the to-be-learned categories have high between-category similarity wherein it is difficult to discriminate one category from the other (e.g., Carvalho & Goldstone, 2014; Rohrer, 2012; Sana et al., 2017; but see also Foster, Mueller, Was, Rawson, & Dunlosky, 2019; Rohrer, Dedrick, & Burgess, 2014). That high similarity held for the materials in this study. If not, a benefit of interleaving might not have been observed (e.g., Hausman & Kornell, 2014). The present results also suggest other preconditions, including (a) an alignment between interleaving schedule and trial type (i.e., study vs. practice), and for systematically alternating schedules, (b) the aforementioned ability to arrange category materials in a directly opposed manner. These preconditions may be related in that manipulations which increase the efficacy of practice, such as randomization, impede the acquisition of prerequisite knowledge, for which systematic schedules can be more beneficial.

The choice of randomization or systematic alternation is also likely to more substantially impact interleaving than blocking. Specifically, random or organized reshuffling within a single tense (blocking) yields comparatively minimal change in trial organization or predictability than

does shuffling between categories (interleaving). Consequently, similar levels of learning may have occurred, overall, across the blocked groups. That fact may account for the narrower range of delayed test proportion correct over experiments for the blocked versus interleaved groups (a range of 0.07 for the blocked groups vs. 0.21 for the interleaved groups).

Limitations and Future Work

It remains to be determined whether the current results will generalize to other tasks that have study and practice components and require learning rules. Although this study was arguably focused on category learning, the potential effects of rule memorization warrant further consideration. It has been theorized that interleaving may be less helpful for explicit rule learning (Sorensen & Wolz, 2016). Followup studies may shed more light on this issue.

As in prior work, we restricted our use of root verbs and suffixes to only common instances in Spanish and only recruited undergraduate student L2 learners; future studies might involve other materials (e.g., French as in Carpenter & Mueller, 2013) and involve learners of different abilities or stages of language acquisition (e.g., children as in Jones et al., 2015). Followup work might also incorporate random schedules that prevent any category from repeating on successive trials (which requires more than two categories), better controls for time-on-task, interleaving of only one "dimension" (e.g., tenses only), and comparisons of systematic-to-random, random-to-systematic, and blocked-to-interleaved schedules.

Although the error analyses suggest that interleaving primarily by tense can improve the ability to identify tense, the ordering of pronouns and suffixes also warrants consideration. One possibility, for example, is that the presentation of three suffixes within each tense in Phase 2 for the blocked group yielded a form of "interleaving by suffix within tense," and that ordering might have influenced the learning of suffixes relative to the schedules used by the interleaved groups (although the post-hoc error analyses were inconclusive). Research designs that specifically aim to disambiguate verb conjugation errors may provide further insights into the effects of interleaving for tenses vs. suffixes and other dimensions.

Practical Implications

Happily, the chief takeaways of the present research for verb conjugation training are immediately actionable. When using interleaving, studied information should be presented in a logically organized sequence that predictably alternates between to-be-learned tenses. That systematic alternation will maximize opportunities to make comparisons. In contrast, practice exercises should be randomly shuffled such that it is impossible to predict the tense on any given trial. That randomization will keep learners guessing on each attempt and require them to engage in greater amounts of retrieval. In a variety of circumstances, it should be feasible for instructors and students alike to implement these recommendations with traditional (e.g., paper-and-pencil) and other (e.g., computerized) methods.

References

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*(3), 392-402.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance. Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA, US: The MIT Press.

Bortoli, L., Robazza, C., Durigon, V., & Carra, C. (1992). Effects of contextual interference on learning technical sports skills. *Perceptual and Motor Skills*, *75*(2), 555-562.

Carpenter, S. K. (2014). Spacing and interleaving of study and practice. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying the science of learning in education: Infusing psychological science into the curriculum* (pp. 131-141). Washington, DC: Society for the Teaching of Psychology.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*(5), 671-682. doi:http://dx.doi.org/10.3758/s13421-012-0291-4

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481-495.

Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: what can sequencing effects tell us about inductive category learning?. *Frontiers in Psychology*, *6*, 505.

Castañeda, D. A. (2011). The effects of instruction enhanced by video/photo blogs and wikis on learning the distinctions of the Spanish preterite and imperfect. *Foreign Language Annals*, *44*(4), 692-711.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58. doi:http://dx.doi.org/10.1177/1529100612453266

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot.

Foster, N. L., Mueller, M. L., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition*, 1-14.

Gagné, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired associates. *Journal of Experimental Psychology*, *40*(1), 61.

Hausman, H., & Kornell, N. (2014). Mixing topics while studying does not enhance learning. *Journal of Applied Research in memory and Cognition*, *3*(3), 153-160.

Jones, A. C., Wardlow, L., Pan, S. C., Zepeda, C., Heyman, G. D., Dunlosky, J., & Rickard, T. C. (2016). Beyond the rainbow: Retrieval practice leads to better spelling than does rainbow writing. *Educational Psychology Review*, *28*(2), 385-400.

Kang, S. H. K. (2017). The benefits of interleaved practice for learning. In J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79-93). New York: Routledge.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97-103. doi:http://dx.doi.org/10.1002/acp.1801

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction?". *Psychological Science*, *19*(6), 585-592. doi:http://dx.doi.org/10.1111/j.1467-9280.2008.02127.x

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 730.

Lee, T. D., Simon, D. A. (2004) *Contextual interference*. In Williams, A. M., Hodges, N. J. (Eds.), Skill acquisition in sport: Research, theory and practice. London: Routledge. Pp. 29–44.

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science, 9*(3-5), 241-289. doi:http://dx.doi.org/10.1016/0167-9457(90)90005-X

Pan, S. C. (2015). *The interleaving effect: mixing it up boosts learning.* Available at: http://www.scientificamerican.com/article/the-interleaving-effect-mixing-it-up-boosts-learning/

Pan, S. C., & Bjork, R.A. (1994). Acquiring an Accurate Model of Human Learning: Toward an Owner's Manual. In M. Kahana and A. Wagner (Eds.), *Oxford Handbook of Memory, Vol. II: Applications*. Oxford, England, UK: Oxford University Press.

Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (2018). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology* (online first publication).

Porter, J. M., & Magill, R. A. (2010). Systematically increasing contextual interference is beneficial for learning sport skills. *Journal of Sports Sciences*, *28*(12), 1277-1285.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*(3), 355-367. doi:10.1007/s10648-012-9201-3

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*(5), 1323-1330.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*(6), 481-498. doi:http://dx.doi.org/10.1007/s11251-007-9015-8

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology, 109*(1), 84-98.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217. doi:http://dx.doi.org/10.1111/j.1467-9280.1992.tb00029.x

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*(2), 179-187. doi:http://dx.doi.org/10.1037/0278-7393.5.2.179

Sorensen, L. J., & Woltz, D. J. (2016). Blocking as a friend of induction in verbal category learning. *Memory & Cognition*, *44*(7), 1000-1013.

Smith, P. J., Gregory, S. K., & Davies, M. (2003). Alternating versus blocked practice in learning a cartwheel. *Perceptual and Motor Skills*, *96*(3), 1255-1264.

Travlos, A. K. (2010). Specificity and variability of practice, and contextual interference in acquisition and transfer of an underhand volleyball serve. *Perceptual and Motor Skills, 110*(1), 298-312.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, *41*(3), 429-442.

Westfall, R., & Foerster, S. (1996). Beyond aspect: New strategies for teaching the preterite and the imperfect. *Hispania, 79*(3), 550-560.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, *9*(2), 185-211.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*(7), 918.

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied, 23*(4), 403-416.

Table 1

Stimuli and Training Example Materials

Phase	Trial type	Tense	Detail or example (answer)
1	Study	Preterite	1. For past actions that had a specific and clear beginning and/or end.
			2. To specifically state the beginning and end of a past action.
			3. For past actions that were repeated a specific number of times.
			4. For past actions that occurred during a specific period of time.
		Imperfect	1. For past actions that lack a specific and clear beginning or end.
			2. For past actions that were repeated habitually.
			3. For stating one's age in the past.
			4. For past actions that "set the stage" for another action.
	Training	Preterite	Is the following sentence *preterite*? "On Tuesday I ate four tacos." (*Yes*)
			Is the following sentence *preterite*? "I used to walk in the park." (*No*)
		Imperfect	Is the following sentence *imperfect*? "I used to read in my free time." (*Yes*)
			Is the following sentence *imperfect*? "We slept for eight hours." (*No*)
2	Study	Preterite	If the pronoun is "I" ("*yo*"), replace "*-ar*" with "*-e*"
			If the pronoun is "you" ("*tu*"), replace "*-ar*" with "*-aste*"
			If the pronoun is "we" ("*nosotros*"), replace "*-ar*" with "*-amos*"
		Imperfect	If the pronoun is "I" ("*yo*"), replace "*-ar*" with "*-aba*"
			If the pronoun is "you" ("*tu*"), replace "*-ar*" with "*-abas*"
			If the pronoun is "we" ("*nosotros*"), replace "*-ar*" with "*-abamos*"
	Copy	Preterite	Type the proper form of *bailar* into: "I ____ with my friend last month." (*baile*)
		Imperfect	Type the proper form of *bailar* into: "I used to ____ with my friend." (*bailaba*)
3	Training	Preterite	Conjugate *hablar* into: "We ____ with two doctors last week." (*hablamos*)
		Imperfect	Conjugate *llamar* into: "I used to ____ with my teacher." (*llamabas*)

Note. Suffixes were limited to those for the "I", "you [singular]," and "we" pronoun equivalents only. Removal of accent marks and some simplifications were used to ensure consistency.

Table 2

Summary of Training Procedures in Experiments 1-4

Group		Phase 1: Tense rules		Phase 2: Verb suffixes	Phase 3: Verb conj. practice
		Study	Training		
Blocked	Task description	Study rules for one tense	Identify each presented sentence as matching that tense or not	Study and copy suffixes for "I", "you," and "we" pronouns in that tense	Practice verb conjugation for that tense
	Trial count	4 study trials (one for each of 4 rules for one tense)	16 practice trials (2 cycles of 8 trials for one tense)	3 study+copy trials (one for each of 3 suffixes for one tense)	Exps. 1-3: 9 practice trials (all involving one tense); Exp. 4: 18 practice trials (all involving one tense)
	Trial pattern (systematic examples shown*)	P_1-P_2-P_3-P_4 or I_1-I_2-I_3-I_4	P_{Yes}-P_{No}-P_{Yes}-P_{No}-P_{Yes}-P_{No}-P_{Yes}-P_{No} or I_{Yes}-I_{No}-I_{Yes}-I_{No}-I_{Yes}-I_{No}-I_{Yes}-I_{No}	P_I-P_Y-P_W or I_I-I_Y-I_W	P_I-P_Y-P_W-P_I-P_Y-P_W-P_I-P_Y-P_W or I_I-I_Y-I_W-I_I-I_Y-I_W-I_I-I_Y-I_W
Interleaved	Task description	Study rules for each of two tenses	Identify each presented sentence as matching one of the two tenses or not; one tense specified per trial	Study and copy suffixes for "I", "you," and "we" pronouns in the two tenses	Practice verb conjugation for both tenses
	Trial count	8 study trials (one for each of 4 rules per tense)	32 practice trials (2 cycles of 16 trials; in each cycle, 8 trials per tense)	6 study+copy trials (one for each of 3 suffixes per tense)	Exps. 1-3: 18 practice trials (9 trials per tense); Exp. 4: 36 practice trials (18 trials per tense)
	Trial pattern (systematic examples shown*)	P_1-I_1-P_2-I_2-P_3-I_3-P_4-I_4 or I_1-P_1-I_2-P_2-I_3-P_3-I_4-P_4	P_{Yes}-I_{Yes}-P_{No}-I_{No}-P_{Yes}-I_{Yes}-P_{No}-I_{No}- P_{Yes}-I_{Yes}-P_{No}-I_{No}-P_{Yes}-I_{Yes}-P_{No}-I_{No} or I_{Yes}-P_{Yes}-I_{No}-P_{No}-I_{Yes}-P_{Yes}-I_{No}-P_{No}-I_{Yes}- P_{Yes}-I_{No}-P_{No}-I_{Yes}-P_{Yes}-I_{No}-P_{No}	P_I-I_I-P_Y-I_Y-P_W-I_W or I_P-P_P-I_Y-P_Y-I_W-P_W	P_I-I_I-P_Y-I_Y-P_W-I_W-P_I-I_I-P_Y-I_Y- P_W-I_W-P_I-I_I-P_Y-I_Y-P_W-I_W or I_I-P_I-I_Y-P_Y-I_W-P_W-I_I-P_I-I_Y-P_Y-I_W- P_W-I_I-P_I-I_Y-P_Y-I_W-P_W

Note. All participants completed Phases 1-3 in sequential order. Blocked group participants completed Phases 1-3 twice, once for each tense. Each capital letter with a subscript refers to a presentation or practice trial, wherein the capital letter indicates tense (P = *preterite*; I = *imperfect*) and the subscripted numbers and letters denote rules (1-4) and pronouns ("I", "you", "we"), respectively. Conj. = Conjugation; Exp. = Experiment. (*) Trial pattern examples are systematic, as in Experiment 2, for simplicity; for examples of the trial patterns used in each experiment, see Figure 1.

Table 3

Training Session Means (SE)

Exp.	Schedule type	Group	Phase 1: Tense rules		Phase 2: Verb suffixes	Phase 3: Verb conj. practice
			First cycle	Second cycle		
1	Random	Blocked	0.88 (0.01)	0.93 (0.01)	0.87 (0.02)	0.94 (0.01)
		Interleaved	0.77 (0.02)	0.80 (0.03)	0.82 (0.03)	0.51 (0.03)
2	Systematically alternating	Blocked	0.84 (0.02)	0.94 (0.01)	0.91 (0.02)	0.95 (0.01)
		Interleaved	0.73 (0.02)	0.82 (0.02)	0.84 (0.03)	0.68 (0.04)
3	Systematic-to-random	Blocked	0.81 (0.02)	0.89 (0.01)	0.87 (0.03)	0.88 (0.02)
		Interleaved	0.75 (0.02)	0.83 (0.02)	0.85 (0.03)	0.67 (0.03)
4	Systematic-to-random	Blocked	0.81 (0.02)	0.91 (0.02)	0.84 (0.03)	0.89 (0.01)
		Interleaved	0.74 (0.02)	0.82 (0.02)	0.84 (0.03)	0.60 (0.03)

Note. Data are collapsed across tenses. Exp. = Experiment; Conj. = Conjugation.

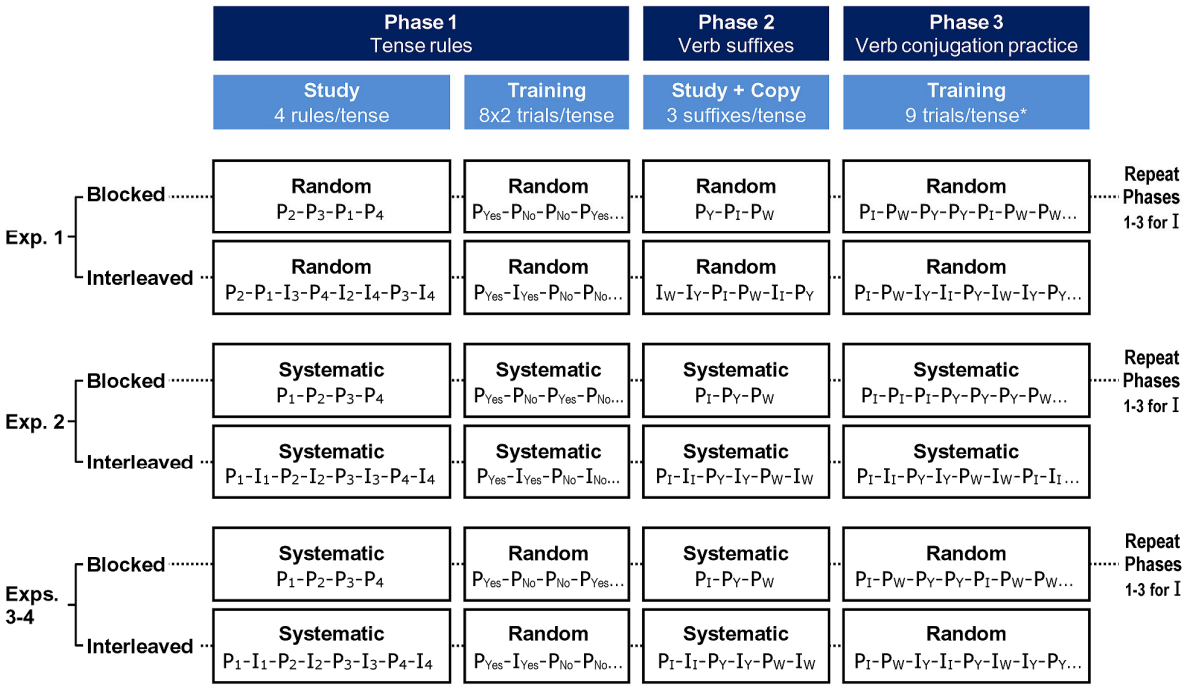

Figure 1. Timelines and examples of the schedule types used in Experiments 1-4 (please refer to Table 1 for stimuli; for a general summary of the procedure, please refer to Table 2). Each capital letter with a subscript refers to a presentation or practice trial, wherein the capital letter indicates tense (P = *preterite*; I = *imperfect*) and the subscripted numbers and letters denote rules (1-4) and pronouns ("I", "you", "we"), respectively. One of two counterbalanced orders (i.e., P first) is displayed. (*) in Experiment 4, the number of Phase 3 trials was doubled. Exp. = Experiment.

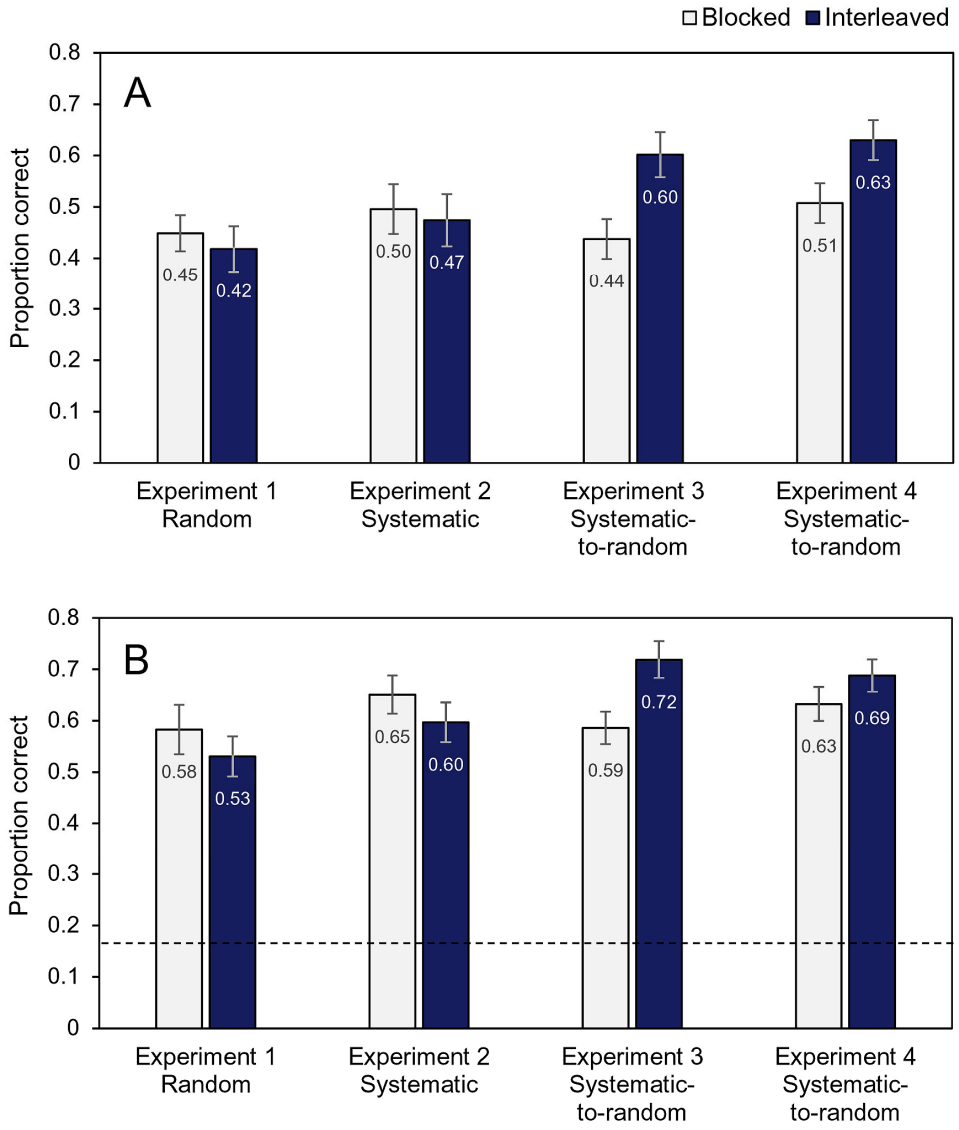

Figure 2. Results from the delayed test of Experiments 1-4. Panel A: short answer. Panel B: multiple-choice. In each experiment, participants answered short answer before multiple-choice questions. In panel B, the dotted line indicates the expected accuracy rate for guessing. Error bars = *SEM*.

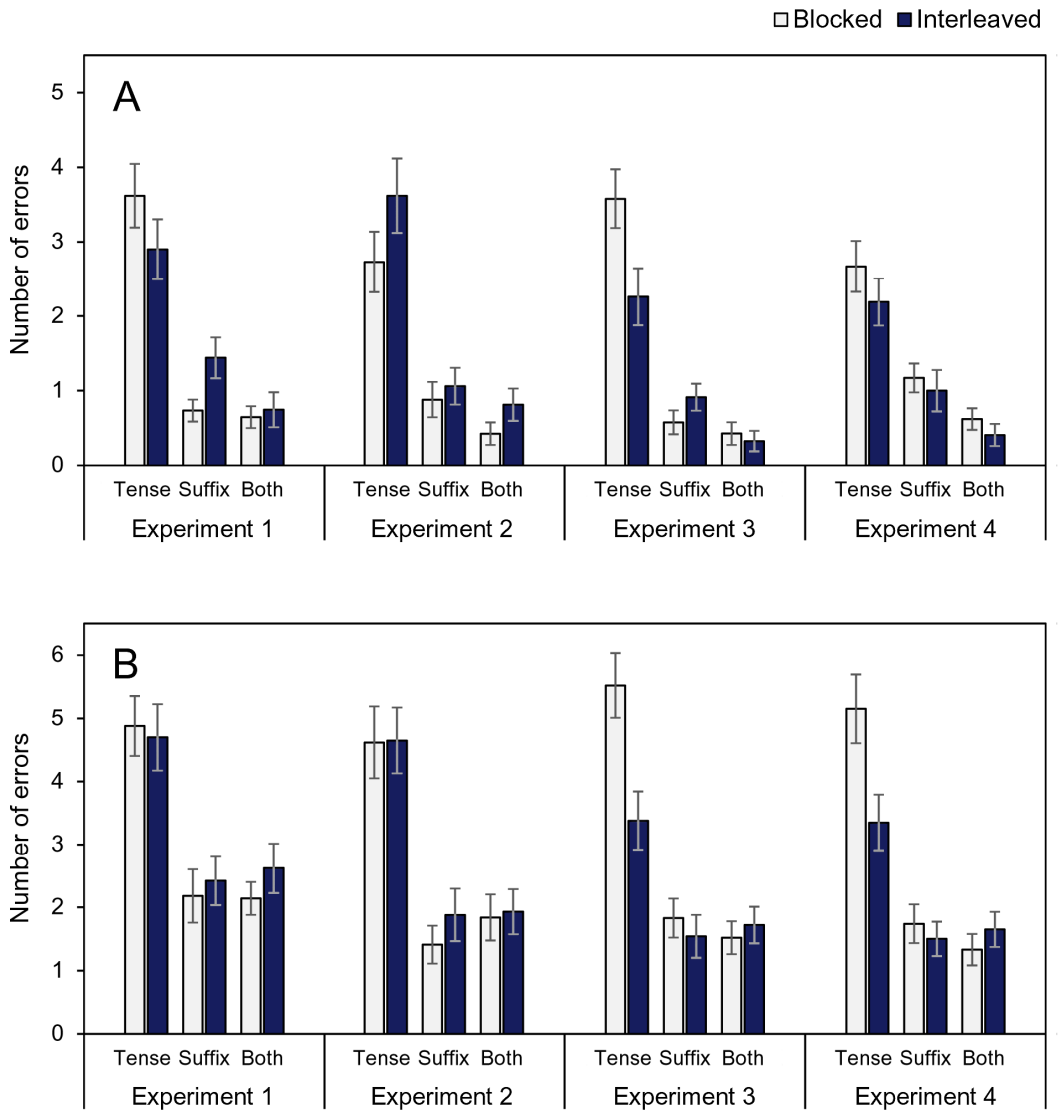

Figure 3. Mean number of verb conjugation errors on the delayed test of Experiments 1-4 (count data averaged over participants). *Tense* errors = verb conjugations that corresponded to the given pronoun but were in the incorrect tense, *suffix* errors = verb conjugations that were in the correct tense but had the incorrect suffix for that given pronoun, and *both* errors = verb conjugations involving incorrect tense and incorrect correspondence to the given pronoun. Panel A: short answer. Panel B: multiple-choice. Error bars = *SEM*.