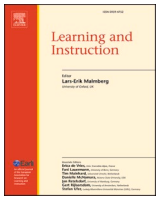
Learning and Instruction 95 (2025) 102045

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc



Interleaved practice enhances grammar skill learning for similar and dissimilar tenses in Romance languages☆,☆☆,☆☆☆

Steven C. Pan [a,*], Sergio Rodríguez Flores [b,c], Michelle E. Kaku [d], Wing Hei Esmee Lai [d]

[a] Department of Psychology, National University of Singapore, 9 Arts Link, 117572, Singapore
[b] Centre for Language Studies, National University of Singapore, Singapore
[c] Department of Spanish Language and General Linguistics, Universidad de Murcia, Spain
[d] Department of Psychology, University of California, Los Angeles, CA, 502 Portola Plaza, 90095-1563, USA

A B S T R A C T

Background: Interleaved practice (or *interleaving*), the strategy of alternating between categories or concepts during study or practice, can enhance second language grammar skills. It remains to be determined, however, whether that enhancement specifically involves identifying tenses, conjugating verbs, or both, and whether close similarity between tenses is essential.
Aims: This study investigated the language skills that interleaving can enhance and the extent to which that enhancement is limited to highly similar tenses.
Sample: Participants were college students (Experiment 1, 92 participants; Experiment 2, 109 participants) and adult learners (Experiment 3, 104 participants; Experiment 4, 88 participants).
Methods: In each experiment, participants completed two weekly learning sessions and a one-week delayed criterial test. In the blocked group, participants learned one tense per session. In the interleaved group, participants alternated between two tenses during each session. The criterial test assessed: verb conjugation skills (all experiments), tense identification ability for specific usage scenarios (Experiments 1–3), and the capacity to identify the language of a sentence written in a specific tense (Experiment 4).
Results: Interleaving improved verb conjugation skills in all experiments, tense identification ability in Experiments 1 and 3, and language identification skills in Experiment 4. Benefits of interleaving were observed across tenses varying in usage, meaning, and suffixes.
Conclusions: Interleaving enhances multiple language skills, including verb conjugation, tense identification, and language identification. Those benefits are not limited to highly similar tenses. Accordingly, these results challenge assumptions about interleaving and underscore its potential as an effective approach for improving language learning.

1. Introduction

Alongside vocabulary, reading, and writing, among the most important elements of learning a second language (L2) is mastering its grammar—that is, the rule-based system that governs the structure of the language. Doing so usually involves learning multiple grammatical tenses, which are categories that express temporal references (i.e., when an action occurred or is going to occur), aspect (how a verbal action extends over time), and other nuances such as politeness or commands. In many language courses, each grammatical tense is learned in isolation. For example, in an Italian course, one lesson might be devoted to the present tense, whereas a later lesson may be devoted to the future tense (e.g., Protej & Coggle, 2004). Although this one-topic-at-a-time approach (which learning scientists call *blocking*) may seem logical, recent research suggests that an alternative strategy known as *interleaved practice* (or *interleaving*)—which in L2 learning entails mixing tenses together—can yield more effective and longer-lasting learning.

☆ Thanks to Charis Kwok, Faith Siauw, Elisya Tung, and Kah Liang Tan for assistance with running Experiment 3, Josiah Hoo and Nathaneal Teo for assistance with data and analysis verification, to Hieu Vu Minh for programming assistance, and Jarrett Lovelett for helpful suggestions. This research was supported by a Faculty of Arts and Social Sciences (FASS) grant from the National University of Singapore and a University of California, Los Angeles, Chancellor's Research Fund grant, both awarded to S. C. Pan.☆☆ Portions of this research were presented at the 9th Annual Centre for Language Studies International Conference (CLaSIC) in December 2022, Singapore, and at the European Association for Research in Learning and Instruction Special Interest Group 6 & 7 Conference in August 2024, Tübingen, Germany.☆☆☆ Data, materials, and analysis scripts for this study can be accessed via the Open Science Framework at: https://osf.io/436ya/.

* Corresponding author. Department of Psychology, Faculty of Arts and Social Sciences, College of Humanities and Sciences, National University of Singapore, 9 Arts Link, Singapore, 117572.
E-mail address: scp@nus.edu.sg (S.C. Pan).

https://doi.org/10.1016/j.learninstruc.2024.102045
Received 23 June 2024; Received in revised form 17 September 2024; Accepted 24 October 2024
Available online 7 November 2024

1.1. The interleaving effect for inductive learning

Interleaving involves alternating between different categories, concepts, or skills during learning (for reviews, see Carpenter, 2014; Carvalho & Goldstone, 2019; Firth et al., 2021; Richter et al., 2022; Rohrer, 2012). For example, studying topics A, B, and C in the sequence $A_1B_1C_1B_2A_2C_2A_3C_3B_3$ (with subscripts representing different examples) exemplifies interleaving. It can produce the *interleaving effect*, in which subsequent test performance improves compared to learning through blocking.

Prior research has demonstrated the interleaving effect in various domains. Such studies have often focused on inductive learning (i.e., acquiring conceptual knowledge from exemplars) of visual categories, such as landscape artists' painting styles (Kornell & Bjork, 2008), radiograms (e.g., Hatala et al., 2003), bird families (e.g., Wahlheim et al., 2011), and chemical structures (Eglington & Kang, 2017). More recently, interleaving effects have also been observed for problem-solving skills in such domains as mathematics (e.g., Mielicki & Wiley, 2022; Rohrer et al., 2014) and physics (e.g., Samani & Pan, 2021). In L2 learning, interleaving has yielded mixed results for vocabulary (e.g., Finkbeiner & Nicol, 2003; Schneider et al., 2002) and pronunciation skills (e.g., Carpenter & Mueller, 2013), but studies involving grammar skills have produced more encouraging results (e.g., Nakata & Suzuki, 2019b).

Two prominent theories explain the interleaving effect. The discriminative contrast hypothesis (Kang & Pashler, 2012; see also Carvalho & Goldstone, 2014) posits that interleaving enables or causes learners to compare differences between categories, thereby enhancing the capacity to differentiate between them. Spacing effect-based accounts (e.g., Carpenter, 2014) suggest that the interleaving effect is derived from the well-established spacing effect (Ebbinghaus, 1885), where increased time between exposures to stimulus materials enhances learning. Crucially, interleaving inherently introduces spacing—that is, when it is used, there is always some passage of time between successive exposures to a given category or concept. In contrast, with blocking, each category or concept is learned in contiguous fashion and such spacing is absent. Hence, the interleaving effect might be due at least partly to the spacing effect. Studies of visual category learning in which the amount of temporal spacing between exemplars was manipulated (e.g., Birnbaum et al., 2013; Ge et al., 2021; Kang & Pashler, 2012), however, have generally concluded in favor of the discriminative contrast hypothesis (i.e., adding more spacing has not improved learning). Nevertheless, the extent to which that conclusion applies beyond visual materials remains unclear, and theoretical development on the interleaving effect is still ongoing (Carpenter & Pan, 2024; Foster et al., 2019).

A meta-analysis of interleaving and inductive learning by Brunmair and Richter (2019) highlighted *between-category similarity*—that is, shared perceptual or other characteristics—as a critical boundary condition of the interleaving effect. They found the effect to be largest for studies featuring artists' painting styles (Hedges' $g = .67$), which typically have high between-category similarity, and not significant in the case of verbal materials such as expository texts ($g = .21$) or words ($g = -.39$), which often vary in similarity and other characteristics. Given the paucity of studies involving interleaving with authentic educational materials and inconsistent findings, however, the extent of the interleaving effect across different materials and the role of between-category similarity remains to be determined.

1.2. Interleaving and second language grammar learning

Mastering the grammatical tenses of a language commonly requires learning at least two related skills through study and practice (Rodriguez, 2004) and may involve both declarative and procedural knowledge (Ullman, 2004). These skills are *tense identification*, the ability to recognize different tenses in specific usage scenarios (which involves studying rules and viewing examples), and *verb conjugation*, the ability to modify verbs to reflect the appropriate tense (which, in Romance languages, entails learning how to append the appropriate suffix to a root verb). Fig. 1 illustrates how these skills interact in Romance languages such as Spanish or French. Typically, L2 learners first acquire the rules that specify the usage of a given tense, then master the procedures and suffixes needed for verb conjugation in that tense.

There are theoretical reasons to expect that interleaving could benefit L2 grammar learning. For instance, interleaving-induced discriminative contrast may be useful for learning to tell apart confusable or highly similar grammatical tenses. Moreover, the temporal spacing introduced by interleaving might improve L2 learners' retention of rules, procedures, and suffixes. Interleaving may also impose a more appropriate level of difficulty than blocking at more advanced stages of L2 learning, leading to better learning (Suzuki & Sunada, 2019).

Although a growing body of research explores the potential of interleaving in various L2 learning scenarios (e.g., Carpenter & Mueller, 2013; Nakata & Suzuki, 2019a; Schneider et al., 2002; Suzuki, 2021; Suzuki & Sunada, 2020), very few studies have investigated interleaving and L2 grammar learning to date. In one example, Pan et al. (2019b, Experiments 3–4; see also Pan, Lovelett, et al., 2019) had English-speaking undergraduate students without prior Spanish experience learn the *Pretérito Perfecto Simple* and *Pretérito Imperfecto* tenses across two weekly sessions. The tenses were chosen for their similar usage and high potential for confusion (Castañeda, 2011), which according to the discriminative contrast hypothesis should make them highly suitable for interleaving. The blocked group learned and practiced one tense per session, whereas the interleaved group alternated between tenses during each session as they studied and practiced. On a one-week delayed verb conjugation test, interleaving advantages were observed (Cohen's $d = .53$ in Experiment 3; $d = .79$ in Experiment 4). Discriminative contrast processes were pinpointed as a major driver of those results.

In another example, Nakata and Suzuki (2019b) had Japanese-speaking students with prior English language experience study five different grammatical structures in English—simple past, present perfect, and first, second, and third conditional tense—in a blocked, interleaved, or blocked-then-interleaved order within a single session. As in Pan et al. (2019b), the tenses being learned are prone to confusion with one another. An interleaving effect was not observed on an immediate grammaticality judgment test but emerged when the test was administered one week later ($d = .64$). The authors attributed these results to discriminative contrast and spacing effect-based processes.

Suzuki and Sunada (2019) and Suzuki et al. (2022) had Japanese-speaking students (at least some having prior English language experience) learn and practice relative clause constructions in English (e.g., subjective vs. objective relative clauses) via an oral picture description task in which different construction types were mixed together (interleaving) or grouped separately (blocking), then take picture-based comprehension or production posttests. On immediate posttests, there were speed and accuracy advantages for interleaving over blocking (*d*s ranging from .19 to .58), whereas the delayed posttest results were inconsistent: Suzuki et al. found an interleaving advantage for accuracy ($d = .37$), but not for speed, whereas Suzuki and Sunada found no advantage for either. Methodological differences such as the number of constructions learned and the amount of practice have been suggested as possible sources of these discrepant results (Suzuki et al., 2022).

As just described, the few studies on interleaving and L2 grammar learning to date—which have addressed different tasks, learning materials, and different kinds of participants—encompass highly promising results (e.g., Nakata & Suzuki, 2019b; Pan, Lovelett, et al., 2019, Pan, Lovelett, et al., 2019) and mixed findings (e.g., Suzuki & Sunada, 2019; Suzuki et al., 2022). More research is needed to address outstanding and unresolved issues. For instance, most of the interleaving effects for L2 grammar learning have yet to be successfully replicated, and replication is critical in education research (Plucker & Makel, 2021). Moreover,

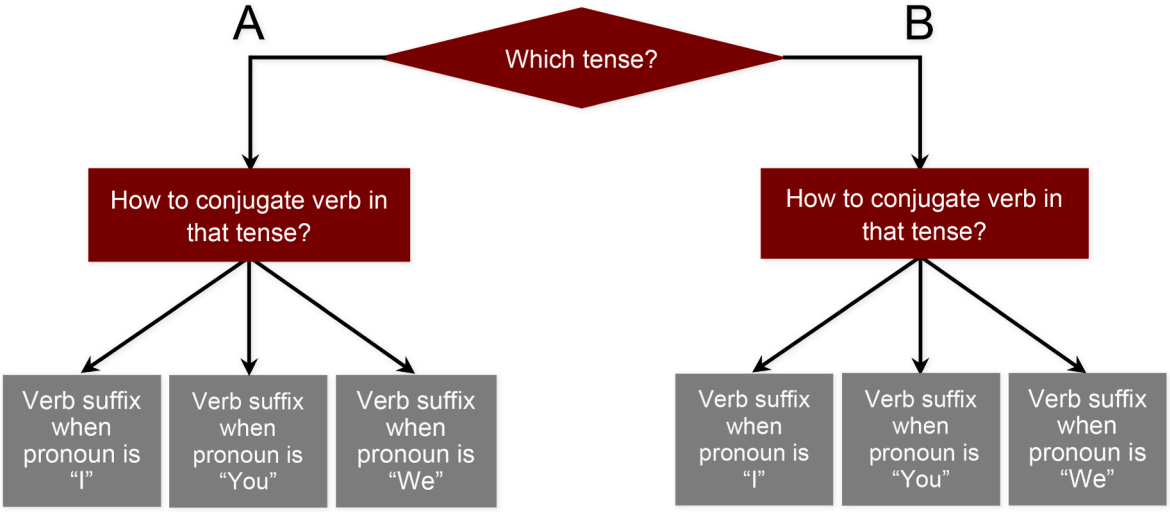


Fig. 1. Tense Identification and Verb Conjugation Process in Romance Languages
Note. Mastering grammatical tenses involves learning at least two skills. First, *tense identification* (or language identification, as in Experiment 4). Second, *verb conjugation* (i.e., using the correct suffix to conjugate the verb based on the tense and subject pronoun). A and B refer to different tenses.

such effects have yet to be extended to other languages beyond Spanish and English, to yet other grammar learning scenarios, and other learning contexts. In addition, the grammar skills that interleaving can enhance have yet to be fully characterized and especially in the case of verb conjugation skills. For example, on the criterial test of Pan et al. (2019b), participants' responses were limited to typing verbs; as such, the extent to which the results reflected proficiency in tense identification, verb conjugation ability, or both remained unknown. Tense identification more closely aligns with the category discrimination tasks where the strongest interleaving effects have been demonstrated (Brunmair & Richter, 2019), whereas verb conjugation does not map as directly onto such tasks, raising the possibility that interleaving is more beneficial for the former than the latter.

It is also unclear whether the benefits of interleaving for grammar learning are specific to highly similar grammatical tenses or other confusable materials. Although the discriminative contrast hypothesis and inductive learning research imply that close similarity is necessary (and meta-analytic findings indicate that the interleaving effect depends heavily on the types of materials being learned), those assumptions have yet to be evaluated in the context of L2 grammar learning. In an extreme case, no interleaving effects might be observed outside of the Spanish and English tenses that have been investigated thus far (for analogous discussions, see Fiedler, 2011; Sedlmeier et al., 1998).

1.3. The present study

The present study encompassed four experiments modeled after Pan et al. (2019b; Experiments 3–4). These experiments addressed (1) the reproducibility of interleaving benefits for L2 grammar learning, (2) the generalizability of these benefits across different languages and tenses, (3) the specific grammar skills that may or may not be enhanced by interleaving, and (4) the role of between-category similarity among to-be-learned grammatical tenses. As in the prior study, each experiment featured a three-week, three-session design with two learning sessions, the learning of two tenses (excepting Experiment 4, which featured one tense across two languages), and a one-week delayed criterial test. The present experiments differed, however, in the specific tenses and/or languages studied, the study setting and participants, and the design of the criterial test, which allowed for separate assessment of tense or language identification and verb conjugation skills in each experiment.

Across experiments, the effects of interleaving for learning seven different grammatical tenses and two Romance languages were investigated. These tenses, which are detailed in Table 1, varied in similarity of use and in the suffixes required for verb conjugation (we categorized similarity based on the defining rules for each tense, examining suffix morphology, and in consultation with L2 instructors). When selecting tenses, our goal was to maximize the diversity of the chosen tenses, constrained only by tense availability and the need to avoid combinations sharing identical suffixes, which could complicate data analysis.

2. Experiment 1

The first experiment was a near direct replication of Pan et al. (2019b, Experiments 3 and 4), in which participants learned to identify and conjugate Spanish verbs in the *Pretérito Perfecto Simple* and *Pretérito Imperfecto* tenses via interleaving or blocking. Both tenses are highly similar in usage (both refer to past actions or information and vary in terms of applicable situations) yet dissimilar in the suffixes used for verb conjugation (differing in multiple letters). Unlike those prior experiments, however, this experiment was conducted online and featured a delayed criterial test that assessed the two to-be-learned language

Table 1
Spanish and French grammatical tenses.

Expt.	Language	Grammatical tenses (Romance language)	Use (rules), meaning	Verb suffixes	Closest English equivalent
1	Spanish	*Pretérito Perfecto Simple, Pretérito Imperfecto*	High similarity	Low similarity	*Simple Past, Imperfect*
2	Spanish	*Presente de Indicativo, Pretérito Imperfecto*	Low similarity	Low similarity	*Simple Present, Imperfect*
3	French	*Conditionnel Présent, Futur Simple*	High similarity	High similarity	*Conditional Mood, Simple Future*
4	French, Spanish	*Présent de Subjonctif, Presente de Subjuntivo*	Moderate similarity	Low similarity[a]	*Present Subjunctive*

Note.
[a] Low similarity is in reference to the six verb suffixes that were learned in that experiment.

skills—tense identification and verb conjugation—separately. This approach allowed for distinguishing between the effects of interleaving on each skill and enabled conditional analyses of the types of errors occurring during tense identification and verb conjugation. Ultimately, this experiment addressed the question: What is the extent of the interleaving effect for the *Pretérito Perfecto Simple* and *Pretérito Imperfecto* tenses in terms of tense identification and verb conjugation?

2.1. Method

Data and materials for all experiments are archived at the Open Science Framework and accessible via this URL: https://osf.io/436ya/.

2.1.1. Participants

The target sample size for all experiments was determined via a priori power analysis in G*Power 3.1 (Faul et al., 2009), which indicated that 43 participants per group yields 80% power to detect a group difference of $d = .53$ (the interleaving effect in Pan et al., 2019b, Experiment 3) or larger on a two-tailed, independent-samples t-test at $\alpha = .05$. Essentially, our goal was to ensure sufficient power to detect an effect of interleaving or blocking at this effect size or greater, whether for tense identification or verb conjugation.

One hundred and four undergraduate participants were recruited from the subject pool at a large U.S. university in exchange for partial course credit. Only participants that had passed through a pre-screening questionnaire (which verified that they were fluent in English, plus had no prior training, were not a native speaker of, or had family members who were native speakers of Spanish, French, or another Romance language) were allowed to participate. The mean age was 20.4 years ($SD = 1.7$); the gender distribution of the sample was 83% female, 14% male, and 2% other. Sixty-four percent of the participants were of Asian/Pacific Islander ancestry, 22% were Caucasian/White, 2% were African-American/Black, 1% were Latinx/Hispanic, and 10% were of other ethnicities. All participants were fluent in English and had no prior Spanish language experience. Twelve participants did not complete all three sessions, leaving 92 participants (blocked group, $n = 46$; interleaved group, $n = 46$) in the final sample. All participants in this study gave informed consent and data collection for all experiments was conducted with ethics approval from the involved university.

2.1.2. Design

The independent variable was the type of learning schedule (Blocked vs. Interleaved). There were two dependent variables, tense identification and verb conjugation performance (on the criterial test). Each participant was randomly assigned to a blocked group or an interleaved group, then completed two weekly learning sessions followed by a one-week delayed criterial test. Within or across sessions, each tense was learned across three phases: In Phase 1, the defining rules for the tense were presented and practiced; in Phase 2, the suffixes that are to be used to conjugate verbs for different pronouns in the tense were learned and practiced; and Phase 3 consisted of a series of verb conjugation practice trials.

Although all participants viewed identical materials, presentation order differed between groups (see Fig. 2). The blocked group focused on one grammatical tense per session whereas the interleaved group learned both tenses in the first session and revisited them in the second session. The order of tense presentation was counterbalanced across participants.

2.1.3. Materials

To reduce confusion, all L2 materials in this study were presented with accompanying English language translations and without diacritical marks. Learning materials were adapted from Pan et al. (2019b). Examples of the practice trials and questions used during the learning sessions and on the criterial test for all experiments are included in Appendix A.

2.1.3.1. Learning sessions. Phase 1 included 4 rules per tense, each presented using 3 example English sentences, and 8 additional English sentences for practice trials. Phase 2 involved 3 pronoun-verb combinations per tense (e.g., equivalent to "*I*", "*You*," and "*We*" in English), with each having one example sentence and one fill-in-the-blank practice question. The practice questions were in English except for the Spanish root verb. Phase 3 used 18 fill-in-the-blank questions per tense, formatted similarly to Phase 2's materials but with different Spanish root verbs (all ending in "*-ar*") and novel sentences.

2.1.3.2. Criterial test. The criterial test included 42 trials divided between the two to-be-learned tenses. Each trial included a sentence that was presented in Spanish fill-in-the-blank form (i.e., with a blank where a conjugated verb was meant to be inserted), along with a different Spanish "*-ar*" root verb and a translation of the sentence into English. All criterial test trials were dissimilar from the practice materials used during the learning sessions (i.e., different verbs, different sentences).

2.1.4. Procedure

Experiment 1 was conducted through the university's subject pool website. Participants signed up for all three sessions before starting the first session, with subsequent sessions becoming available at 7-day intervals (automatically scheduled from the completion of the first

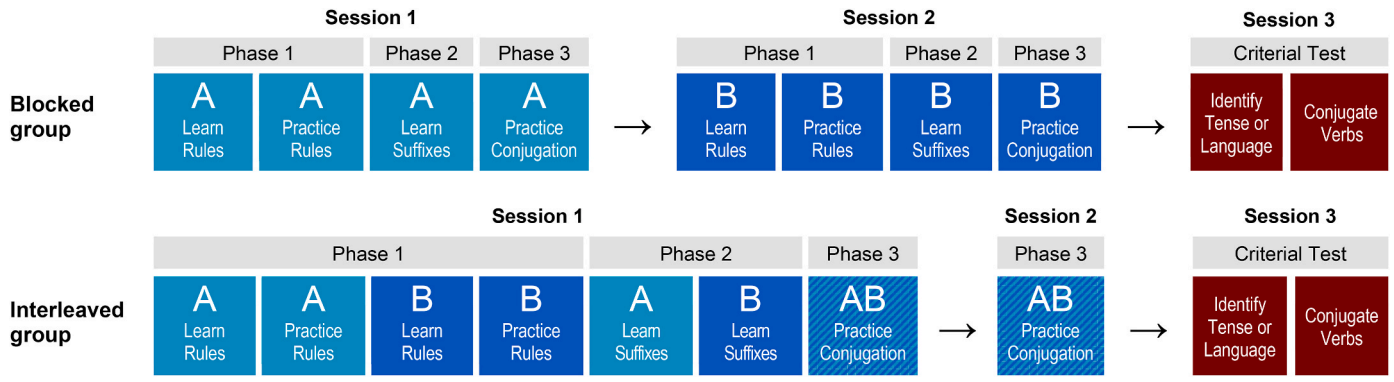


Fig. 2. Experimental Design

Note: In the Blocked group, participants focused on one grammatical tense per session, whereas the Interleaved group learned and practiced two tenses in session 1, then revisited them in session 2. All participants completed a criterial test in session 3, where they first identified the grammatical tense (Experiments 1–3) or language (Experiment 4), then conjugated the verb accordingly (Experiments 1–4). In the figure, A = first grammatical tense, B = second grammatical tense, and AB = alternating between tenses.

session). Reminder emails were sent the day before the second and third sessions and in cases of completion delays. All sessions were self-paced, computer-presented, and conducted via internet browsers.

2.1.4.1. Learning sessions. The general procedure for each phase, based on Pan et al. (2019b), was as follows. The specific rules and suffixes that were learned are presented in Appendix B.

2.1.4.1.1. Phase 1 (tense rules). Participants studied four defining rules one at a time, each accompanied by example sentences. They then viewed a rules summary slide before completing eight practice trials. These trials involved identifying whether presented sentences fit any learned rules. Feedback, including correct answers and relevant rules, was provided after each response. Participants repeated the practice trials once in a newly randomized order. By the end of Phase 1, participants were expected to have acquired a foundational understanding of the learned tense's defining characteristics.

2.1.4.1.2. Phase 2 (verb suffixes). Participants learned to conjugate verbs in the specified tense for three pronoun-verb scenarios: "*I*" (*"yo"*), "*You*" (*"tu"*), and "*We*" (*"nosotros"*), with root verbs ending in "*-ar*." Each scenario was studied individually, along with the appropriate verb suffix and example sentence. Participants completed a single fill-in-the-blank practice trial for each scenario, receiving the correctly conjugated verb as feedback. The order of scenarios was consistent for each participant. By the end of Phase 2, participants were expected to understand how to conjugate "*-ar*" verbs in the given tense for the specified pronouns.

2.1.4.1.3. Phase 3 (verb conjugation practice). The third phase began with a summary slide detailing defining rules and verb suffixes, followed by a series of 18 randomly-ordered fill-in-the-blank verb conjugation practice trials. After each trial, a feedback screen displayed the correct answer, the relevant pronoun, and grammatical tense.

2.1.4.1.4. Blocked and interleaved schedules. The blocked group focused exclusively on one tense per session without mixing between them, whereas the interleaved group alternated between the two tenses across sessions. Specifically, the blocked group completed Phases 1–3 for a given tense in the first session and repeated the sequence for the other tense in the second session (see top row of Fig. 2). In contrast, the interleaved group completed Phase 1 and Phase 2 for both tenses in succession (see bottom row of Fig. 2), followed by Phase 3 with 18 randomly-ordered practice trials evenly split between the two tenses, all in the first session. Their second session only included Phase 3, featuring 18 randomized practice trials evenly split between the two tenses. Both groups spent approximately 30 min in total across both sessions. After each session, they answered metacognitive questions on difficulty and learning, rated on a five-point Likert scale (results discussed in the General Discussion).

2.1.4.2. Criterial test. The third session consisted of the criterial test, the design of which differed from that in Pan et al. (2019b). Each test trial involved a two-step procedure. First, in a two-option multiple-choice question, participants identified which of the previously-learned tenses best represented the presented sentence. Second, they typed the correctly conjugated verb that fit the sentence. Trials were presented sequentially and in random order. The experiment concluded after all trials were completed.

2.2. Results

2.2.1. Learning sessions

Table 2 reports descriptive statistics from Phases 1–3 in the first and second learning sessions, where applicable, in the blocked and the interleaved groups from each experiment of this study. In Experiment 1, performance was relatively similar between groups during Phases 1 and 2, whereas during Phase 3, the blocked group's performance was substantially higher (by $>$ .30 proportion correct). Those patterns match that observed in prior work (e.g., Pan, Lovelett, et al., 2019, Pan, Lovelett, et al., 2019). Adherence to the three-session schedule was excellent, with 96% of participants in the final sample completing each session on the assigned day.

Table 2
Learning Session Mean Performance (SE).

Expt.	Session	Group	Phase 1: Tense rules		Phase 2: Verb suffixes	Phase 3: Verb conjugation practice
			First cycle	Second cycle		
1	1	Blocked	.78 (.030)	.85 (.027)	.89 (.034)	.88 (.026)
		Interleaved	.80 (.024)	.88 (.019)	.84 (.031)	.58 (.034)
	2	Blocked	.83 (.025)	.91 (.020)	.84 (.036)	.91 (.025)
		Interleaved	–	–	–	.49 (.033)
2	1	Blocked	.91 (.017)	.97 (.0077)	.90 (.023)	.91 (.015)
		Interleaved	.90 (.015)	.95 (.012)	.90 (.020)	.67 (.033)
	2	Blocked	.88 (.019)	.94 (.016)	.85 (.033)	.90 (.017)
		Interleaved	–	–	–	.55 (.027)
3	1	Blocked	.79 (.024)	.87 (.023)	.79 (.038)	.88 (.016)
		Interleaved	.78 (.023)	.87 (.021)	.79 (.031)	.55 (.028)
	2	Blocked	.72 (.026)	.83 (.021)	.83 (.035)	.89 (.014)
		Interleaved	–	–	–	.54 (.033)
4	1	Blocked	.74 (.024)	.83 (.025)	.79 (.047)	.86 (.025)
		Interleaved	.70 (.028)	.83 (.026)	.83 (.031)	.51 (.035)
	2	Blocked	.76 (.029)	.84 (.025)	.86 (.040)	.90 (.020)
		Interleaved	–	–	–	.48 (.038)

2.2.2. Criterial test

Violin and box plots depicting the criterial test results for the two assessed language skills are presented in the first-row panels of Fig. 3 and descriptive statistics for those results are presented in Table 3. Criterial test results were analyzed using *t*-tests wherein *p*-values were used to assess statistical significance. For non-significant results, equivalence testing was also performed to evaluate whether the effect sizes were small enough to be considered practically negligible. The reliability of the tense identification and verb conjugation scores (split-half reliability) was .89 and .95 for the interleaved group, and .97 and .94 for the blocked group, respectively.

2.2.2.1. Tense identification and verb conjugation performance. An independent-samples *t*-test comparing mean proportion correct scores between the interleaved and blocked groups for tense identification found no significant difference, $t(90) = 1.95$, $p = .055$, $d = .41$, whereas an analogous *t*-test for the case of verb conjugation revealed a significant interleaving benefit, $t(90) = 2.36$, $p = .020$, $d = .49$. These results correspond to inspection of the relevant panels of Fig. 3, in which a substantial interleaving advantage is evident for verb conjugation, whereas for tense identification, there is a smaller numerical interleaving advantage.

To better characterize the non-significant difference in tense identification, we conducted an equivalence test using the two one-sided test (TOST) approach with an equivalence margin of $\pm.5$ standard deviations. The 90% confidence interval for the mean difference between the interleaved and blocked groups (.014–.18) did not fall within the margin, indicating that we cannot conclude practical equivalence. Thus, while the difference in tense identification performance between the

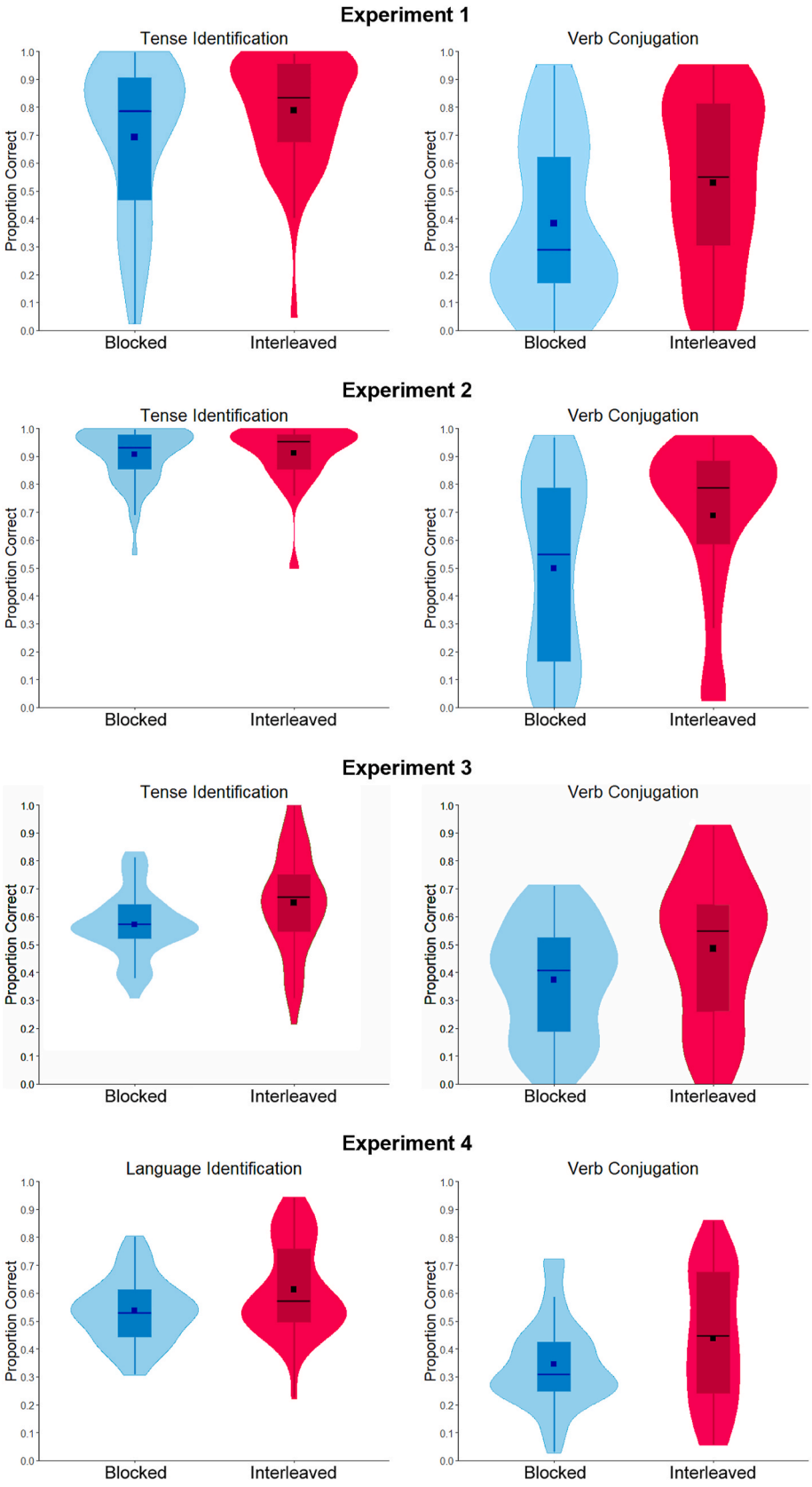


Fig. 3. Criterial test results.

Table 3
Criterial test mean performance (SE).

Experiment	Group	Tense identification (Experiments 1–3) or language identification (Experiment 4)	Verb conjugation
1	Blocked	.69 (.040)	.38 (.043)
	Interleaved	.79 (.029)	.53 (.044)
2	Blocked	.91 (.013)	.50 (.045)
	Interleaved	.91 (.014)	.69 (.037)
3	Blocked	.57 (.016)	.37 (.028)
	Interleaved	.65 (.025)	.49 (.036)
4	Blocked	.54 (.017)	.35 (.024)
	Interleaved	.61 (.024)	.44 (.035)

blocked and interleaved groups was not statistically significant (p = .055), the effect size was too large to be considered negligible.

2.2.2.2. Conditional error analyses. Drawing on the data generated by the two-step criterial test procedure, we investigated two error types that participants may have committed on the test.

2.2.2.2.1. Correct tense, incorrect suffix. In cases where participants correctly identified the tense, errors in verb conjugation could still occur, indicating incomplete learning of the rules or procedures for verb conjugation in that tense. An analysis of criterial test trials where the tense was correctly identified (accounting for >80% of all test items) showed that the blocked group had a significantly higher error rate (M = .57) than the interleaved group (M = .41), $t(90) = 2.23$, p = .028, d = .46. Therefore, in cases where the two groups demonstrated similar ability to identify grammatical tense, the interleaved group showed better verb conjugation skills as well.

2.2.2.2.2. Incorrect tense, correct suffix. In cases where the incorrect tense was identified, participants might conjugate the verb appropriately for the tense they mistakenly believe the sentence represents. Such patterns would suggest accurate knowledge of verb conjugation procedures and suffixes. On criterial test trials where the tense was incorrect (fewer than 20% of all items), the interleaved group had a numerically higher, though not statistically significant, rate of correct conjugation (M = .54) compared to the blocked group (M = .41), $t(86) = 1.58$, p = .12, d = .34. A TOST equivalence test with a ±.5 standard deviation margin found that the 90% CI for the mean difference (−.27 to .007) did not fall within the equivalence margin, indicating that the difference was not negligible.

2.3. Discussion

In Experiment 1, benefits of interleaving for Spanish L2 grammar learning were observed and particularly for the case of verb conjugation skills. Moreover, those benefits were observed on a modified criterial test versus that used in prior research and in an online setting. Conditional analyses further indicated that correct verb conjugation was more likely following interleaving than blocking when the tense was correctly identified. Thus, Experiment 1 demonstrated that the benefits of interleaving for Spanish L2 grammar learning are not specific to the most similar and arguably most confusable elements of the *Pretérito Perfecto Simple* and *Pretérito Imperfecto* tenses (i.e., tense identification).

3. Experiment 2

Experiment 2 investigated the effects of interleaving for learning two Spanish tenses that are not easily confused, the *Presente de Indicativo* and *Pretérito Imperfecto* tenses. These tenses are highly dissimilar both in usage (the former refers to actions occurring in the present, whereas the latter refers to certain past actions) and in the suffixes used for verb conjugation (differing in multiple letters; for details, see Appendix B). As in Experiment 1, participants completed a one-week delayed criterial test that assessed tense identification and verb conjugation skills separately. This experiment addressed the question: What is the extent of the interleaving effect for tense identification and verb conjugation when the tenses being learned are highly dissimilar?

3.1. Method

The design and analysis plan for Experiment 2 were preregistered at https://aspredicted.org/cp6s-56mw.pdf.

3.1.1. Participants

Undergraduate student participants were recruited from the same subject pool, using the same pre-screening questionnaire, and compensated in the same manner as in Experiment 1. The mean participant age was 20.4 years (SD = 2.2); the gender distribution of the sample was 83% female, 15% male, and 1% other. Sixty-two percent of the participants were of Asian/Pacific Islander ancestry, 20% were Caucasian/White, 4% were African-American/Black, and 13% were of other ethnicities. One-hundred and twenty-three participants signed up for Experiment 2; 14 participants did not complete all three sessions, leaving 109 participants (blocked group, n = 53; interleaved group, n = 56) in the final sample.

3.1.2. Design, materials, and Procedure

The design was identical to that of Experiment 1, with the sole change being the use of the *Presente de Indicativo* tense in place of the *Pretérito Perfecto Simple* tense (with the same number and type of items for Phases 1, 2, and 3). *Presente de Indicativo* has not previously been investigated in studies of L2 learning and interleaving. Accordingly, participants learned the equivalent of a present tense and a past tense. All procedures, including reminders to participants in case of delayed session completion, were identical to that used in the preceding experiment.

3.2. Results

3.2.1. Learning sessions

Performance during the learning sessions (see Table 2) was relatively similar between groups during Phases 1 and 2, whereas during Phase 3, the blocked group's performance was substantially higher (by > .20 proportion correct). Adherence to the three-session schedule was excellent, with 98% of participants in the final sample completing each session on the assigned day.

3.2.2. Criterial test

Violin and box plots in the second-row panels of Fig. 3 depict the criterial test results for the two language skills assessed and descriptive statistics for those results are presented in Table 3. The preregistration specified pairwise comparisons and conditionalized analyses, both of which were conducted. Additionally, equivalence testing analogous to that performed for the prior experiment was used to evaluate non-significant results; such tests were not preregistered. The reliability of the tense identification and verb conjugation scores was .85 and .97 for the interleaved group, and .82 and .98 for the blocked group, respectively.

3.2.2.1. Tense identification and verb conjugation performance. An independent-samples t-test comparing mean proportion correct scores between the interleaved and blocked groups for tense identification found no significant difference, $t(107) = .23$, p = .81, d = .045, whereas an analogous t-test for verb conjugation revealed a significant interleaving benefit, $t(107) = 3.27$, p = .0014, d = .63. These results correspond to inspection of the relevant panels of Fig. 3, which show no indications of an interleaving advantage for tense identification and evidence of an interleaving advantage for verb conjugation. A TOST equivalence test with a ±.5 standard deviation margin performed on the

tense identification results found that the 90% CI for the mean difference (−.028 to .037) fell within the equivalence margin, suggesting that the difference between the interleaved and blocked groups for the tense identification was negligible.

3.2.2.2. Conditional error analyses. We performed conditional error analyses analogous to those conducted for the first experiment.

3.2.2.2.1. Correct tense, incorrect suffix. In cases where the tense was correctly identified (representing over 90% of all test items), the blocked group had a higher rate of conjugating the verb incorrectly ($M = .47$) than the interleaved group ($M = .27$), $t(107) = 3.40$, $p = .00095$, $d = .65$. That result suggests that the blocked group did not learn verb conjugation skills as well as the interleaved group.

3.2.2.3. Incorrect tense, correct suffix. In cases where the tense was not correctly identified (representing <10% of all test items), the interleaved group had a higher rate of conjugating the verb correctly for the tense that they believed the sentence represented ($M = .59$) than the blocked group ($M = .37$), $t(83) = 2.42$, $p = .018$, $d = .53$. Hence, in cases where they mistakenly identified the tense, the interleaved group was less likely to compound that error during verb conjugation.

3.3. Discussion

Experiment 2 revealed that previously observed benefits of interleaving for two highly similar Spanish tenses are not an isolated case. Even for a highly dissimilar combination of Spanish tenses, interleaving was still beneficial for verb conjugation skills. Conditional error analyses reinforced that conclusion (in fact, if tense identification accuracy is disregarded, then the advantage of interleaving over blocking would have been even higher). Both groups, however, easily distinguished between the two tenses, resulting in no interleaving advantage for tense identification (although a ceiling effect may have been involved). Overall, these results heighten the possibility that discriminative contrast is not the sole driver of the benefits of interleaving for Spanish L2 grammar learning, at least for verb conjugation skills.

4. Experiment 3

Experiment 3 investigated potential benefits of interleaving for L2 grammar learning in another Romance language, French. It involved the *Conditionnel Présent* and *Futur Simple* tenses, which even proficient French speakers can confuse due to similarities in usage and suffixes (both represent future actions; the former refers to possible, wishful, or supposed actions whereas the latter expresses intentions as well as suppositions) (Chevalier-Karfis, 2023; Sallee & Hebert, 2011). As such, the tenses used in this experiment exhibited a substantial amount of between-category similarity (and were, in that respect, the polar opposite of those used in Experiment 2). This experiment addressed the question: What is the extent of the interleaving effect for tense identification and verb conjugation in French when the tenses are highly similar in how they are used and the suffixes involved?

4.1. Method

The design and analysis plan for Experiment 3 were preregistered at https://aspredicted.org/vkx4-7wpn.pdf.

4.1.1. Participants

In a departure from the prior experiments, 110 adult participants were recruited from Prolific Academic, a crowdsourcing platform for online studies. Eligibility criteria included being located in an English-speaking country (Australia, Canada, New Zealand, the U.K., or the U.S.), having a 95% or higher approval rate on prior Prolific studies, fluency in English, and no prior experience (i.e., having taken a class, native fluency) with any Romance language. Participants were informed that they would learn French for a payment of at least GBP 8.70 or USD 10.50, with a GBP 1.65 or USD 2.00 bonus for on-time completion. Six participants did not complete all sessions, leaving 104 participants (blocked group, $n = 53$; interleaved group, $n = 51$) in the final sample. The mean age was 33.3 years ($SD = 11.6$); the gender distribution of the sample was 73% female and 27% male. Seventy-four percent of the participants were of Caucasian/White ancestry, 9% were Asian, 5% were Black, 4% had a multi-ethnic background, 7% were of other ethnic groups, and 1% declined to provide ethnicity information. Most participants had an undergraduate degree or higher (66%), some undergraduate coursework (18%), or a high school diploma (15%). Most participants (64%) were from the U.K., followed by Canada (30%), the U.S. (2%), and other countries.

4.1.2. Design

The design closely mirrored that of previous experiments. To accommodate time zone differences and Prolific Academic's automatic payment limits, each participant had up to 36 h to complete each session upon it becoming available. Consequently, most participants completed the second and third sessions within a seven or eight-day interval from the immediately preceding session.

4.1.3. Materials

All materials were similar to those used in previous experiments, with the same number and type of items for Phases 1, 2, and 3 for each tense, but involving the *Conditionnel Présent* and *Futur Simple* tenses in French. The verb suffixes learned involved root verbs ending in "*-er*" and pronouns equivalent to "He"/"She," "We," or "You" (plural) in French, as the suffixes did not overlap within or across tenses.

4.1.4. Procedure

Procedures were adapted for the Prolific Academic platform. Participants signed up for the first session knowing they would receive links to the second and third sessions at weekly intervals and be emailed when the next session was available. Only participants who completed a session received access to the next one. At the end of each session, participants were redirected to the Prolific website, where their participation was recorded. Payment was provided after the third session. The procedure for each phase of the learning sessions and the criterial test was identical to the preceding experiments.

4.2. Results

4.2.1. Learning sessions

Performance during the learning sessions (see Table 2) was comparable between groups in Phases 1 and 2, whereas in Phase 3, performance in the blocked group was higher (by approximately .33 proportion correct). Adherence to the three-session schedule was good, with 79% of participants in the final sample completing each session on the assigned day.

4.2.2. Criterial test

Violin and box plots of the criterial test results are presented in the third-row panels of Fig. 3 and descriptive statistics for those results are presented in Table 3. The reliability of the tense identification and verb conjugation scores was .91 and .96 for the interleaved group, and .75 and .89 for the blocked group, respectively.

4.2.2.1. Tense identification and verb conjugation performance. Independent-samples *t*-tests comparing mean proportion correct scores between the interleaved and blocked groups revealed significant interleaving advantages for tense identification, $t(102) = 2.64$, $p = .0096$, $d = .52$, and for verb conjugation, $t(102) = 2.45$, $p = .016$, $d = .48$. These results correspond to inspection of the relevant panels of Fig. 3, which show

indications of an interleaving advantage for tense identification and for verb conjugation.

4.2.2.2. Conditional error analyses. Analyses are as follows.

4.2.2.2.1. Correct tense, incorrect suffix. In cases where the tense was correctly identified (representing over 50% of all test items), the blocked group had a higher rate of incorrect verb conjugation ($M = .52$) compared to the interleaved group ($M = .35$), $t(102) = 2.50$, $p = .014$, $d = .49$. That result suggests that the blocked group did not acquire verb conjugation skills as effectively as the interleaved group.

4.2.2.2.2. Incorrect tense, correct suffix. In cases where the tense was not correctly identified (over 30% of all test items), the interleaved group had a higher rate of correctly conjugating the verb for the tense they believed the sentence represented ($M = .61$) compared to the blocked group ($M = .44$), $t(101) = 2.78$, $p = .0065$, $d = .55$. Thus, even when they mistakenly identified the tense, the interleaved group was less likely to compound that error with incorrect verb conjugation.

4.3. Discussion

Experiment 3 found benefits of interleaving for learning two highly similar tenses in French. The observed benefits most closely resembled the patterns observed in Experiment 1—that is, interleaving improved tense identification and verb conjugation skills. Conditional error analyses provided further evidence that interleaving enhanced verb conjugation performance over blocking.

5. Experiment 4

Experiment 4 examined the effects of interleaving on learning the present subjunctive tense in two Romance languages, French (*Présent de Subjonctif*) and Spanish (*Presente de Subjuntivo*), simultaneously. Owing to the two languages' shared Latin origins, these tenses exhibit moderate similarity in usage and suffixes (although the subsets of suffixes used in this experiment shared few letters and hence their similarity can be categorized as low). While learning multiple languages concurrently is somewhat uncommon, it is not unheard of (e.g., Fukui & Yashima, 2021; Henry, 2010) and such learning forms a cornerstone of the E.U.'s plurilingual approach (Council of Europe, 2006). Identifying the language that a sentence represents is a separate skill in its own right (for related discussions, see Erard, 2012; Lightbown & Spada, 2013), including in the context of intercomprehension (where differentiating between related languages facilitates effective communication; for details see Doyé, 2005). This experiment addressed the question: What is the extent of the interleaving effect in the case of learning a tense and associated verb conjugation procedures in two languages simultaneously?

5.1. Method

The design and analysis plan for Experiment 4 were preregistered at https://aspredicted.org/7t4j-5h23.pdf.

5.1.1. Participants

One-hundred and five adult participants were recruited using Prolific Academic in the same manner as the preceding experiment, including under the same eligibility criteria and in exchange for the same amount of monetary compensation. Participants were informed that they would be learning about "Romance languages". Data were excluded from 15 participants did not finish all three sessions and 2 participants that did not follow study instructions, leaving 88 participants (blocked group, $n = 44$; interleaved group, $n = 44$) in the final sample. The mean participant age was 37.5 years ($SD = 13.3$); the gender distribution of the sample was 71% female, 28% male, and 1% non-binary. Sixty-seven percent of the participants were of Caucasian/White ancestry, 14% were Asian, 6% were Black, 4% had a multi-ethnic background, 5% were of other ethnic groups, and 4% declined to provide ethnicity information. Most participants' highest level of education was an undergraduate degree or higher (66%), whereas the remainder had completed some undergraduate coursework (18%) or finished high school or secondary school (15%). Most of the participants (67%) were located in the U.K., followed by Canada (30%) and other countries (3%).

5.1.2. Design, materials, and Procedure

The design, materials, and procedure were largely identical to the prior experiment, except as follows. The materials involved the *Présent de Subjonctif* and *Presente de Subjuntivo* tenses in French and Spanish. In Phase 1, there were only three rules learned per tense (reflecting the manner in which those tenses are commonly defined and keeping the number of rules equal between both tenses), and there were six practice trials instead of eight; and in Phase 2, verb suffixes corresponding to the equivalent of the pronouns "I", "You," and "You" (plural) in French and Spanish, and all for the case of root verbs ending in "*-er*", were learned (the verb suffixes involved did not overlap between languages). Reflecting the fewer rules for each tense, the number of criterial test items was also reduced to 36, and on each test trial, participants were asked to identify language rather than tense prior to conjugating verbs (all of the sentences were written in the present subjunctive tense in either language; for each sentence, participants were asked to identify whether it best reflects the present subjunctive tense in French or Spanish).

5.2. Results

5.2.1. Learning sessions

Performance during the learning sessions (see Table 2) was relatively comparable between groups in Phases 1 and 2, whereas in Phase 3, performance in the blocked group was higher (by $> .40$ proportion correct). Adherence to the three-session schedule was good, with 73% of participants in the final sample completing each session on the assigned day.

5.2.2. Criterial test

Violin and box plots depicting the criterial test results are presented in the bottom row panels of Fig. 3 and descriptive statistics for those results are presented in Table 3. The reliability of the tense identification and verb conjugation scores was .79 and .90 for the interleaved group, and .55 and .70 for the blocked group, respectively (Note: the reduced reliability for language identification in the blocked group may reflect more random guessing in that group; for related discussion see Zimmerman & Williams, 2003).

5.2.2.1. Language identification and verb conjugation performance. Independent-samples *t*-tests comparing mean proportion correct scores between the interleaved and blocked groups revealed significant interleaving advantages for tense identification, $t(86) = 2.55$, $p = .013$, $d = .54$, and for verb conjugation, $t(86) = 2.18$, $p = .032$, $d = .46$. These results correspond to inspection of the relevant panels of Fig. 3, which show indications of an interleaving advantage for tense identification and for verb conjugation.

5.2.2.2. Conditional error analyses. Analyses are as follows.

5.2.2.2.1. Correct language, incorrect suffix. In cases where the language was correctly identified (representing over 50% of all test items), the blocked group was more prone to conjugating the verb incorrectly ($M = .57$ vs. $M = .42$), $t(86) = 2.57$, $p = .012$, $d = .55$. As in prior experiments, this result suggests better learning of verb conjugation skills in the interleaved group.

5.2.2.2.2. Incorrect language, correct suffix. In cases where the language was not correctly identified (over 40% of all test items), the interleaved group was more likely to correctly conjugate the verb for the tense they believed the sentence represented than the blocked group (*M*

= .53 vs. M = .38), $t(86) = 2.40$, $p = .019$, $d = .51$. Therefore, even in cases where they mistakenly identified the language, the interleaved group was less likely to compound that error with incorrect verb conjugation.

5.3. Discussion

Experiment 4 revealed that the benefits of interleaving for L2 grammar learning further extend to learning two Romance languages concurrently. In a scenario where the same tense is being learned in French and Spanish—in which learners might benefit from engaging in comparison or discriminative contrast between languages—a substantial interleaving benefit for a third skill, language identification, was observed. Benefits of interleaving for verb conjugation also occurred.

6. Supplementary analyses

6.1. Retention interval analyses

In each blocked group, one tense was introduced two weeks prior to the criterial test, whereas the other tense was introduced one week prior (versus the interleaved group, where both tenses were introduced two weeks prior). Following Pan et al. (2019b), we examined whether that retention interval difference affected the criterial test results. One-sample t-tests comparing performance in the blocked group for tenses learned in the first versus second sessions showed no significant difference in verb conjugation performance (p-values $\geq$.15). Additionally, TOST equivalence tests with a $\pm$.5 standard deviation margin found that the 90% confidence intervals for those mean differences fell within the equivalence bounds, indicating negligible criterial test performance differences between sessions in each experiment.

For tense or language identification, the blocked group showed no significant difference between the tenses learned in the first versus second session across all experiments except Experiment 2 (p-values $\geq$.43), reflecting patterns similar to verb conjugation. TOST equivalence tests, analogous to those for verb conjugation, also indicated negligible differences in tense or language identification between the two sessions for the blocked group. In Experiment 2, however, a significant retention interval difference favored the tense learned in the second session ($M_{1st\ session}$ = .86 vs. $M_{2nd\ session}$ = .96), $t(52) = 4.71$, $p < .0001$), suggesting greater forgetting of the first learned tense. Uniquely in that experiment, however, tense identification scores were near ceiling, and moreover, there was a negligible difference in such scores among the interleaved and blocked groups.

6.2. Internal meta-analyses

To further characterize the interleaving effects across experiments, we conducted two internal meta-analyses separately for tense/language identification and verb conjugation skills, using the *metafor* package in R (Viechtbauer, 2010) and effect size and sample size information from each experiment (Note: these analyses were not preregistered and hence could be regarded as exploratory). The results are shown in Fig. 4. As shown in the figure, interleaving enhanced tense/language identification ability in all but Experiment 2. The estimated benefit of interleaving for tense/language identification was d = .36, 95% CI [.12, .60]; for verb conjugation skills, it was d = .52, 95% CI [.32, .73]. Overall, the effect sizes of these improvements can be considered meaningful from an educational standpoint (Kraft, 2020).

In the figure it is apparent that the individual effect sizes for tense/language identification (except Experiment 2) and verb conjugation skills are quite similar, occupying a relatively narrow band across experiments. That consistency implies reliable interleaving benefits despite varying tense/language combinations. Overall, the internal meta-analysis results are compatible with a potentially general or broad benefit of interleaving for learning to identify tense or language, at least when there is some degree of similarity between tenses, as well as a general benefit of interleaving for verb conjugation skills that apparently does not depend on similarity between tenses.

7. General Discussion

The present study yielded insights into the reproducibility, generalizability, associated language skills, and specificity of the benefits of interleaving for L2 grammar learning. In response to the issues raised at the outset of this manuscript, we found that (1) the advantages of interleaving for learning highly similar (in use) Spanish tenses were largely reproducible; (2) such advantages successfully generalized to other Spanish tenses, French tenses, tenses that share and do not share highly similar verb suffixes, and even learning a single tense in Spanish and French; (3) tense identification, verb conjugation, and language identification skills benefited from interleaving; and (4) low-to high-similarity tenses (in use or in the suffixes needed for verb conjugation) were better learned through interleaving than blocking, at least with

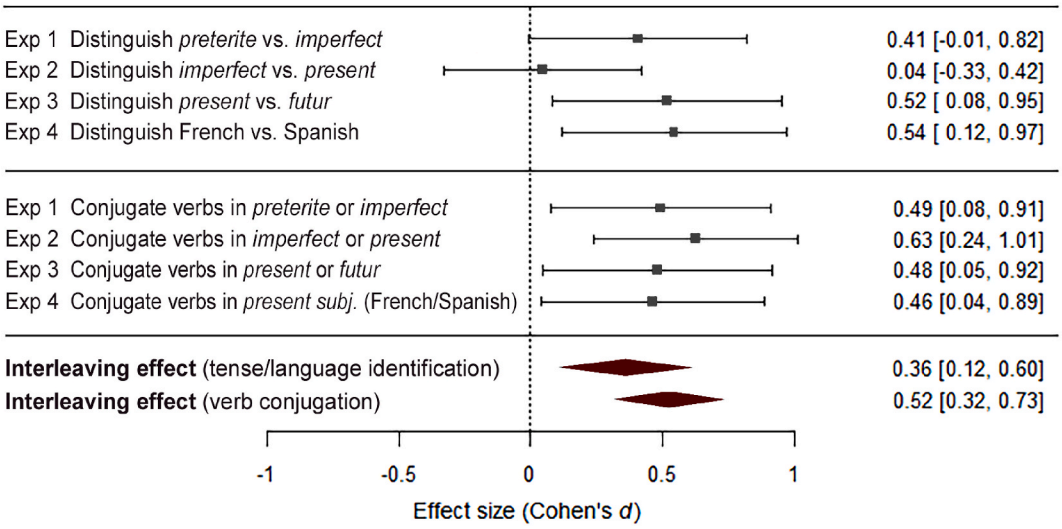


Fig. 4. Internal Meta-Analyses of Experiments 1-4

Note. Upper portion: benefits of interleaving versus blocking for tense (Experiments 1–3) or language identification (Experiment 4) skills. Lower portion: benefits of interleaving versus blocking for verb conjugation skills (Experiments 1–4). Abbreviated tense names shown (subj. = subjunctive).

respect to verb conjugation. Moreover, in a now-classic pattern that is a hallmark of many "desirably difficult" learning techniques (Bjork & Bjork, 2011; Pan & Bjork, 2022), interleaving consistently yielded lower performance than blocking during practice (Phase 3), but that pattern was almost always reversed on the criterial test. Further, a comparison of the patterns across experiments, including via internal meta-analyses, suggests a strong degree of consistency in the observed interleaving benefits. The implications of these results encompass theoretical accounts of the interleaving effect and how L2 instruction might be optimized.

The largely successful replication of Pan et al. (2019b; Experiments 3 and 4) on a modified criterial test provides insights into the interleaving benefits observed in that study. Specifically, it appears that such benefits did not stem solely from or largely depend on enhanced tense discrimination, as has previously been assumed. Instead, improvements in verb conjugation ability were almost certainly also involved. Moreover, when considering the patterns of results for different language skills across experiments, it appears that the interleaving effects in Pan et al. and in the present study involved more than one cognitive mechanism.

7.1. How interleaving impacts second language grammar learning

To reiterate, much of the literature on interleaving effects has focused on inductive category learning with visual materials, where evidence for a major role of discriminative contrast is compelling (e.g., Birnbaum et al., 2013; Ge et al., 2021; Kang & Pashler, 2012). In the present study, a case can also be made for the importance of discriminative contrast for learning tense and language identification skills. Alternating between tenses (or languages) likely enables comparison processes that enhance the ability to differentiate the materials being learned. At a task level, tense and language identification tasks focus on identifying the category membership of presented items rather than requiring complex procedures, similar to the classification tasks used in studies of interleaving and visual category learning. When the tenses or languages are at least somewhat challenging to tell apart (as analyses of similarity and the criterial test results of Experiments 1–4 confirm), then interleaving is beneficial for those tasks. In Experiment 2, however, when the tenses were highly dissimilar and easily distinguishable, there was no apparent benefit of interleaving for tense identification. Although the impact of a ceiling effect cannot be completely dismissed, that result aligns with the importance of discriminative contrast in interleaving effects for similar sets of materials.

With respect to verb conjugation skills, however, it appears that mechanisms beyond discriminative contrast are more important. An interleaving effect was observed for these skills in all experiments, irrespective of similarity in suffixes or related elements such as tense usage. That finding does not align with discriminative contrast being a critically important factor. Moreover, it suggests that stimulus similarity, a potent boundary condition in interleaving effects for other learning tasks (Brunmair & Richter, 2019), may have less impact on interleaving for verb conjugation skills.

Crucially, verb conjugation involves more than just discrimination. It is a multistep process requiring recall of suffixes, conjugation steps, and executing those steps with the appropriate suffix correctly. A potential mechanism for the interleaving effect in such cases has been highlighted in studies of interleaving and math learning, where interleaving effects occur even with dissimilar materials (e.g., Rohrer et al., 2014; see also Foster et al., 2019), and in which multi-step problem-solving procedures have to be learned. In that literature, it has been argued that interleaving benefits problem-solving skills by strengthening *associations* between problem types and solution strategies (for related theorizing from research on interleaving and logic rule learning, see Schneider, 1991). By this account, when practice is blocked, a particular solution strategy or procedure is loaded into working memory once and reused until practice is completed, whereas with interleaving, such information needs to be reloaded into memory on each trial, with problem type-to-procedure associations being strengthened as a result. It is plausible that interleaving can strengthen memory associations for verb conjugation skills in an analogous manner.

Strengthening of memory associations might also be ascribed to the temporal spacing between successive study or practice trials for each grammatical tense. Such spacing occurred within each learning session in the interleaved groups (as an unavoidable consequence of interleaving) and also across the first and second sessions (it should be noted, however, that comparable interleaving effects were observed by Pan, Lovelett, et al., 2019 in a single-session design, suggesting that between-session spacing is not essential for interleaving to be beneficial for such learning). Other mechanisms associated with the spacing effect beyond strengthened memory associations, however—for instance, encoding variability or attentional factors—cannot be entirely ruled out (for discussion see Carpenter & Pan, 2024).

In our view, the most likely explanation for the interleaving effects across experiments ultimately involves two different mechanisms depending on the task type: (a) discriminative contrast being important in the case of tense/language identification skills and (b) strengthened associations between tense-pronoun combinations, suffixes, and procedures being crucial in the case of verb conjugation skills. Temporal spacing in the interleaving groups may have contributed or been responsible for those strengthened associations.

7.2. Study limitations

Although the present study is arguably the most comprehensive to date in addressing a considerable range of grammatical tenses and language skills, several limitations must be acknowledged. First, the learning materials used were a subset of those required for mastering Spanish and/or French in real-world contexts, focusing only on regular verbs and excluding irregular ones, which are more challenging. Second, our experimental design did not fully reflect how grammatical tenses are commonly learned in actual L2 courses. Such courses often involve not just reading and a minimal amount of writing, but also auditory activities (i.e., listening, speaking) and more extensive written exercises. In such courses, it is also common for grammatical tenses to be encountered across more than one session, even in a "blocked" schedule. Any benefits of interleaving relative to such repeated blocked exposure may not be as large as those that were observed in the present experiments.

Further, while the present study did not find a larger interleaving effect for more similar versus less similar sets of tenses, the combinations of tenses being learned were not directly manipulated within experiments. It should also be acknowledged that some of the interleaving effect sizes that were detected fell below the target value of $d = .53$ used for a priori power analysis. Although our study had adequate power to detect moderate effects, those smaller effect sizes warrant caution in interpreting the robustness of the interleaving effect for L2 grammar learning.

7.3. Future research directions

Future research could address the aforementioned study limitations and other questions. For example, new studies could focus on the effects of interleaving for learning three or more tenses among learners that are new to a particular language (cf. Nakata & Suzuki, 2019b), learning the most frequently intermixed tenses, learning combinations of tenses with identical verb suffixes, or learning irregular verbs. Moreover, examining how the similarity of tenses impacts interleaving effects, through direct manipulation of tense combinations, could provide additional insights. Future studies could consider even longer retention intervals, larger sample sizes, tighter limits on study or practice time (cf. Pan et al., 2019b), combining interleaving with other learning strategies (e.g., practice testing, e.g., Pan et al., 2024) and exploring yet-to-be-addressed cases of interleaving in actual L2 learning settings (Schweppe et al., 2024).

Table 4
Metacognitive judgments from learning sessions 1 and 2.

Expt.	Session	Judgments of difficulty						Judgments of learning				
		Group	Very easy	Easy	Moderate	Somewhat difficult	Very difficult	Excellent	Good	Average	Fair	Poor
1	1	Blocked	35%	39%	20%	4%	2%	11%	41%	37%	11%	0%
		Interleaved	2%	11%	37%	30%	20%	0%	22%	39%	20%	20%
	2	Blocked	33%	35%	26%	4%	2%	15%	41%	24%	20%	0%
		Interleaved	0%	13%	28%	37%	22%	0%	13%	28%	28%	30%
2	1	Blocked	40%	49%	9%	2%	0%	25%	47%	21%	8%	0%
		Interleaved	9%	36%	29%	13%	14%	9%	29%	39%	14%	9%
	2	Blocked	32%	47%	19%	2%	0%	19%	45%	25%	9%	2%
		Interleaved	4%	11%	48%	32%	5%	2%	20%	43%	18%	18%
3	1	Blocked	11%	55%	23%	11%	0%	17%	38%	34%	11%	0%
		Interleaved	0%	8%	37%	31%	24%	0%	22%	39%	24%	16%
	2	Blocked	19%	26%	49%	6%	0%	11%	34%	40%	13%	2%
		Interleaved	0%	16%	37%	27%	20%	2%	1%	43%	22%	18%
4	1	Blocked	39%	25%	32%	5%	0%	30%	42%	16%	12%	0%
		Interleaved	2%	9%	52%	18%	18%	2%	18%	34%	27%	18%
	2	Blocked	45%	32%	23%	0%	0%	45%	32%	16%	7%	0%
		Interleaved	2%	11%	41%	23%	23%	0%	16%	34%	20%	30%

Note. Data are presented in percentages of participants per group.

At a broader level, it is important to emphasize that whereas there is growing evidence of interleaving benefits for L2 grammar learning, there is much weaker evidence for other forms of L2 learning such as vocabulary acquisition (e.g., Nakata & Suzuki, 2019a; Schneider et al., 2002) and learning pronunciation-based skills (Carpenter & Mueller, 2013; Suzuki, 2021). The lack of any need to engage in discriminative contrast between different categories—and, conversely, greater benefits of blocked practice on a specific category (e.g., pronunciation rule)—has been suggested as a reason for blocking being more advantageous in such studies. It remains for further research to clarify whether L2 grammar learning represents an area where interleaving is uniquely beneficial or if there are underlying principles that can account for the presence or absence of interleaving effects for different types of L2 learning.

Another direction for future research is to explore the metacognitive aspects of interleaving for L2 learning (which the present study did not emphasize). As shown in Table 4, there was a trend towards higher difficulty ratings and lower judgments of learning among the interleaved groups in all experiments. Whether those trends translate to greater propensity for blocking (and associated issues, such as how to persuade learners of the efficacy of interleaving; e.g., Onan et al., 2022; see also Hartwig et al., 2022) remain to be investigated.

7.4. Pedagogical implications

The present study reveals that interleaving can be a highly beneficial, robust, and broadly applicable strategy for promoting durable and flexibly applicable L2 grammar learning. Whereas prior research has demonstrated the potential of interleaving for learning limited sets of highly similar tenses, it now appears that grammar learning in Romance languages can benefit from interleaved schedules more generally. The pedagogical value of interleaving for L2 grammar learning is heightened by the finding that it can enhance not only tense identification skills, but also verb conjugation and language identification skills. It is important to note, however, that the present results were derived from online experimental contexts, and that any practical recommendations should be regarded as tentative. Nevertheless, when considering the current popularity of blocked learning in L2 courses and other widely-taught academic domains (Rohrer et al., 2020), the potential of strategies such as interleaving to optimize and even revolutionize language learning (Suzuki et al., 2019), at least in the case of grammar skills, warrants strong consideration.

CRediT authorship contribution statement

Steven C. Pan: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Sergio Rodríguez Flores:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Michelle E. Kaku:** Resources, Project administration, Formal analysis, Data curation. **Wing Hei Esmee Lai:** Methodology, Investigation, Conceptualization.

Declarations of interest

None.

Appendix A

Learning Session and Criterial Test Example Materials

Expt.	Grammatical tenses	Learning sessions			Criterial test	
		Phase 1 Practice rules	Phase 2 Learn suffixes	Phase 3 Practice conjugation	Step one: Tense/language identification	Step two: Verb conjugation
1	*Pretérito Perfecto Simple vs. Pretérito Imperfecto*	Is the following sentence *pretérito*? "*I walked through the park.*" a. Yes b. No	Type the correctly conjugated version of *bailar* that fits in the following sentence: "*I ____ with my friend.*"	Conjugate *bailar* into: "*You __ with my friend last month*."	What is the grammatical tense of the following sentence? *"Yo ___ a él por tres años"*. (*I supported him for three years) a. Pretérito Perfecto Simple b. Pretérito Imperfecto*	Conjugate the verb *apoyar* (to support) into the following sentence: *"Yo ___ a él por tres años"*. (*I supported him for three years)* Answer: _____
2	*Presente de Indicativo vs. Pretérito Perfecto Simple*	Is the following sentence *presente*? "*You speak Spanish.*" a. Yes b. No	Type the correctly conjugated version of *hablar* that fits in the following sentence: "*You____ to your neighbour.*"	Conjugate *hablar* into: "*I __ with my cousin every day*."	What is the grammatical tense of the following sentence? *"Yo ___ con un futuro mejor"*. (*I dream of a better future) a. Pretérito Perfecto Simple b. Presente de indicativo*	Conjugate the verb *soñar* (to dream) into the following sentence: *"Yo ___ con un futuro mejor"*. (*I dream of a better future)* Answer: _____
3	*Conditionnel Présent vs. Futur Simple*	Is the following sentence *conditionnel*? "*We would like a tea.*" a. Yes b. No	Type the correctly conjugated version of *manger* that fits in the following sentence: "*He ____ a dessert.*"	Conjugate *manger* into: "*You __ a soup as a starter."*	What is the grammatical tense of the following sentence? *"Il ___ partir en vacances"*. (*He wants to go on vacation) a. Conditionnel Présent b. Futur Simple*	Conjugate the verb *souhaiter* (to want) into the following sentence: *"Il ___ partir en vacances"*. (*He wants to go on vacation)* Answer: _____
4	*Présent de Subjonctif (French) vs. Presente de Subjuntivo (Spanish)*	Is the following sentence *présent de subjonctif (French)*? "*I have to play with my son now.*" a. Yes b. No	Type the correctly conjugated version of *acceder* that fits in the following sentence: "*I do not think I ____ the website tomorrow.*"	Conjugate *voyager* into: "*While I usually __ by train, I prefer to travel by plane*."	The following sentence best represents the present subjective tense in French or Spanish? *"I have a dinner tonight and I have to ___ a good wine"*. a. French b. Spanish	What would be the correctly conjugated form of the root verb *apporter* (to bring) that best fits this sentence? *"I have a dinner tonight and I have to ___ a good wine"*. Answer: _____

Note. Diacritical marks were removed from the presented materials.

Appendix B

Spanish and French Grammatical Tense Rules and Verb Suffixes

Expt.	Language	Grammatical tense	Rules governing use	Verb suffixes
1	Spanish	*Pretérito Perfecto Simple*	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.	If the pronoun is *"I"* ("*yo*"), replace *"-ar"* with *"-e"* If the pronoun is *"you"* ("*tú*"), replace *"-ar"* with *"-aste"* If the pronoun is *"we"* ("*nosotros*"), replace *"-ar"* with *"-amos"*
1, 2	Spanish	*Pretérito Imperfecto*	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.	If the pronoun is *"I"* ("*yo*"), replace *"-ar"* with *"-aba"* If the pronoun is *"you"* ("*tú*"), replace *"-ar"* with *"-abas"* If the pronoun is *"we"* ("*nosotros*"), replace *"-ar"* with *"-abamos"*
2	Spanish	*Presente de Indicativo*	1. To describe something that is happening right now. 2. To describe something happening in the near future. 3. To express general truths and facts. 4. To describe a habitual actions and routines.	If the pronoun is *"I"* ("*yo*"), replace *"-ar"* with *"-o"* If the pronoun is *"you"* ("*tú*"), replace *"-ar"* with *"-as"* If the pronoun is *"we"* ("*nosotros*"), replace *"-ar"* with *"-amos"*
3	French	*Conditionnel Présent*	1. To express future intentions from a past point of view. 2. To imagine present or future situations that are impossible or unlikely in reality. 3. To express a wish/what you want to do. 4. To express uncertain information and rumors.	If the pronoun is *"he/she"* ("*il"/"elle*"), replace *"-er"* with *"-ait"* If the pronoun is *"you"* ("*tu*"), replace *"-er"* with *"-iez"* If the pronoun is *"we"* ("*nous*"), replace *"-er"* with *"-ions"*
	French	*Futur Simple*	1. To express future intentions from a present point of view. 2. To express conditional future situations that we believe are real or possible. 3. To give an order referencing a future action. 4. To express what will happen when/after another action had occurred.	If the pronoun is *"he/she"* ("*il"/"elle*"), replace *"-er"* with *"-a"* If the pronoun is *"you"* ("*tu*"), replace *"-er"* with *"-ez"* If the pronoun is *"we"* ("*nous*"), replace *"-er"* with *"-ons"*
4	French	*Présent de Subjonctif*	1. To express an opinion in negative form. 2. To express contrasting ideas. 3. To express an obligation.	If the pronoun is *"I"* ("*je*"), replace *"-er"* with *"-e"* If the pronoun is *"you"* ("*tu*"), replace *"-er"* with *"-es"* If the pronoun is *"you all"* ("*vous tous*"), replace *"-er"* with *"-iez"*
	Spanish	*Presente de Subjuntivo*	1. To express a doubt or a possibility. 2. To express a hope or a wish. 3. To express what will happen when or after another action had occurred.	If the pronoun is *"I"* ("*yo*"), replace *"-er"* with *"-a"* If the pronoun is *"you"* ("*tú*"), replace *"-er"* with *"-as"* If the pronoun is *"you all"* ("*vosotros*"), replace *"-er"* with *"-ais"*

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. https://doi.org/10.3758/s13421-012-0272-7

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 59–68).

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin, 145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

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). American Psychological Association. http://teachpsych.org/resources/documents/ebooks/asle2014.pdf#page=137.

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

Carpenter, S. K., & Pan, S. C. (2024). Spacing effects on learning and memory. In L. Mickes, & J. T. Wixted (Eds.), *Learning and Memory: A Comprehensive Reference* (3rd Edition). Elsevier. https://doi.org/10.1016/B978-0-443-15754-7.00020-1.

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. https://doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2019). When does interleaving practice improve learning? In J. Dunlosky, & K. A. Rawson (Eds.), *The cambridge handbook of cognition and education* (1st ed., pp. 411–436). Cambridge University Press. https://doi.org/10.1017/9781108235631.017.

Ebbinghaus, H. (1885). In H. A. Ruger, & C. E. Bussenius (Eds.), *Memory: A contribution to experimental psychology*. Dover Publications. Trans.

Eglington, L. G., & Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition, 6*(4), 475–485. https://doi.org/10.1016/j.jarmac.2017.07.005

Fiedler, K. (2011). Voodoo correlations are everywhere—not only in neuroscience. *Perspectives on Psychological Science, 6*(2), 163–171. https://doi.org/10.1177/1745691611400237

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied PsychoLinguistics, 24*(3), 369–383. https://doi.org/10.1017/S0142716403000195

Firth, J., Rivers, I., & Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *The Review of Education*, 642–684. https://doi.org/10.1002/rev3.3266

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, 47*(6), 1088–1101. https://doi.org/10.3758/s13421-019-00918-4

Ge, Y., li, F., Li, X., & Li, W. (2021). What is the mechanism underlying the interleaving effect in category induction: An eye-tracking and behavioral study. *Frontiers in Psychology, 12*, Article 770885. https://doi.org/10.3389/fpsyg.2021.770885

Hartwig, M. K., Rohrer, D., & Dedrick, R. F. (2022). Scheduling math practice: Students' underappreciation of spacing and interleaving. *Journal of Experimental Psychology: Applied, 28*(1), 100–113. https://doi.org/10.1037/xap0000391

Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education, 8*, 17–26. https://doi.org/10.1023/A:1022687404380

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. https://doi.org/10.1002/acp.1801

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

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Mielicki, M. K., & Wiley, J. (2022). Exploring the necessary conditions for observing interleaved practice benefits in math learning. *Learning and Instruction, 101583*. https://doi.org/10.1016/j.learninstruc.2022.101583

Nakata, T., & Suzuki, Y. (2019a). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition, 41*(2), 287–311. https://doi.org/10.1017/S0272263118000219

Nakata, T., & Suzuki, Y. (2019b). Mixing grammar exercises facilitates long-term retention: Effects of blocking, interleaving, and increasing practice. *The Modern Language Journal, 103*(3), 629–647. https://doi.org/10.1111/modl.12581

Onan, E., Wiradhany, W., Biwer, F., Janssen, E. M., & de Bruin, A. B. H. (2022). Growing out of the experience: How subjective experiences of effort and learning influence the use of interleaved practice. *Educational Psychology Review, 34*(4), 2451–2484. https://doi.org/10.1007/s10648-022-09692-3

Pan, S. C., & Bjork, R. A. (2022). Acquiring an accurate mental model of human learning: Toward an owner's manual. In A. Wagner, & M. Kahana (Eds.), *Oxford handbook of learning and memory: Foundations and applications*. Oxford University Press.

Pan, S. C., Lovelett, J., Phun, V., & Rickard, T. C. (2019). The synergistic benefits of systematic and random interleaving for second language grammar learning. *Journal of Applied Research in Memory and Cognition, 8*(4), 450–462. https://doi.org/10.1016/j.jarmac.2019.07.004

Pan, S. C., Selvarajan, G., & Murphy, C. S. (2024). Interleaved pretesting enhances category learning and classification skills. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1037/mac0000194

Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (2019b). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology, 111*(7), 1172–1188. https://doi.org/10.1037/edu0000336

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist, 56*(2), 90–100. https://doi.org/10.1080/00461520.2021.1895796

Protej, V. B., & Coggle, P. (2004). *Teach yourself beginner's Italian*. McGraw-Hill.

Richter, T., Nemeth, L., Berger, R., Borromeo Ferri, R., Hänze, M., & Lipowsky, F. (2022). Using interleaving to promote inductive learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 54*, 164–175. https://doi.org/10.1026/0049-8637/a000260

Rodriguez, J. (2004). *CliffsQuickReview Spanish II*. Houghton Mifflin Harcourt.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*(3), 355–367. https://doi.org/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. https://doi.org/10.3758/s13423-014-0588-3

Rohrer, D., Dedrick, R. F., & Hartwig, M. K. (2020). The scarcity of interleaved practice in mathematics textbooks. *Educational Psychology Review, 32*(3), 873–883. https://doi.org/10.1007/s10648-020-09516-2

Samani, J., & Pan, S. C. (2021). Interleaved practice enhances memory and problem-solving ability in undergraduate physics. *NPJ Science of Learning, 13*. https://doi.org/10.1038/s41539-021-00110-x

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46*(2), 419–440. https://doi.org/10.1006/jmla.2001.2813

Schweppe, J., Lenk-Blochowitz, A., Pucher, M., & Ketzer-Nöltge, A. (2024). Interleaved practice in foreign language grammar learning: A field study. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000917

Sedlmeier, P., Hertwig, R., & Gigerenzer, G. (1998). Are judgments of the positional frequencies of letters systematically biased due to availability? *Journal of Experimental Psychology Learning Memory and Cognition, 24*(3), 754–770.

Suzuki, Y. (2021). Optimizing fluency training for speaking skills transfer: Comparing the effects of blocked and interleaved task repetition. *Language Learning, 71*(2), 285–325. https://doi.org/10.1111/lang.12433

Suzuki, Y., Nakata, T., & Dekeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal, 103*(3), 713–720. https://doi.org/10.1111/modl.12585

Suzuki, Y., & Sunada, M. (2020). Dynamic interplay between practice type and practice schedule in a second language: The potential and limits of skill transfer and practice schedule. *Studies in Second Language Acquisition, 42*(1), 169–197. https://doi.org/10.1017/S0272263119000470

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1–2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*(5), 750–763. https://doi.org/10.3758/s13421-010-0063-y

Zimmerman, D. W., & Williams, R. H. (2003). A new look at the influence of guessing on the reliability of multiple-choice tests. *Applied Psychological Measurement, 27*(5), 357–371. https://doi.org/10.1177/0146621603254799