Studies in Second Language Acquisition (2023), **45**, 812–837
doi:10.1017/S0272263123000062

RESEARCH ARTICLE

The effects of interleaved and blocked corpus-based practice on L2 pragmatic development

Ying Zhang

University of Utah, Salt Lake City, UT, USA
E-mail: ying.z.zhang@utah.edu

(Received 11 July 2022; Revised 30 December 2022; Accepted 04 February 2023)

Abstract

A handful of second/foreign language (L2) studies have examined the effects of practice schedules and reported the advantage of interleaved practice (i.e., practice multiple skills simultaneously) over blocked practice (i.e., practice one skill first and then proceed to the next one). However, no studies in the realm of L2 pragmatics have explored this theme. This study investigated the influence of interleaved corpus-based practice and blocked corpus-based practice on L2 pragmatic development. Sixty-three L2 learners of English from a university in China received instruction on two pragmatic features: suggestions and requests. After the instruction, they were randomly assigned to an interleaved-practice group ($n = 31$) or a blocked-practice group ($n = 32$). Results from multimedia discourse completion tasks on the immediate and delayed posttests showed facilitative and long-term effects of interleaved practice on pragmatic accuracy. Moreover, the results revealed positive and durable influence of blocked practice on fluency. Implications are discussed.

Introduction

Second/foreign language (L2) practice has recently received renewed attention in the field of L2 acquisition (Suzuki et al., 2022). As a vital element in L2 development, *practice* refers to consistent and deliberate learning activities with the objective of developing L2 learners' knowledge and skills (Dekeyser, 2007). According to the skill acquisition theory (Dekeyser, 2015), systematic, deliberate, and extensive practice can convert *declarative knowledge* (i.e., knowledge of that) into *procedural knowledge* (i.e., knowledge of how) and *automatized knowledge* (i.e., accurate and fast application of the learned knowledge). The proceduralization and automatization can enhance accuracy and fluency of L2 learners' performance.

A crucial question regarding L2 practice is how practice should be arranged to optimize L2 acquisition. One subarea that is largely underexplored is L2 practice schedules (e.g., blocked practice versus interleaved practice). Blocked practice refers

to focusing on practicing one target feature before proceeding to another target feature on multiple days (e.g., Day 1: a-a-a, Day 2: b-b-b, and Day 3: c-c-c) or within one day (e.g., a-a-a-b-b-b-c-c-c over the course of one lesson), whereas interleaved practice refers to practicing multiple target features simultaneously on multiple days (e.g., Day 1: a-b-c, Day 2: a-b-c, and Day 3: a-b-c) or within one day (e.g., a-b-c-a-b-c-a-b-c during one instructional session). An increasing number of non-L2 studies in fields such as cognitive psychology (Kang, 2016), mathematics (Rohrer et al., 2014), and chemistry (Eglington & Kang, 2017) have reported the advantage of interleaved practice over blocked practice. However, this line of research is still scarce in the realm of L2 acquisition. Only a handful of L2 studies on grammar (Nakata & Suzuki, 2019; Pan et al., 2019; Suzuki & Sunada, 2020), speaking (Suzuki, 2021), and vocabulary (Schneider et al., 2002) have explored this topic.

Crucially, no studies in the domain of L2 pragmatics have investigated the effects of practice schedules on *pragmatic competence* (i.e., the ability to adopt proper linguistic forms to perform communicative acts based on determinants such as contextual factors). This topic is worth exploring because practice schedules (e.g., blocked versus interleaved practice) are crucial for effective L2 pragmatics instruction and curriculum development. Moreover, optimal practice schedules are conducive to the proceduralization and automatization of L2 pragmatic knowledge, which enables L2 learners to use pragmatic features accurately and fluently (Dekeyser, 2007) and to conduct smooth intercultural communication (Zhang, 2021). Thus, more empirical studies are called for to critically assess and explore the effectiveness of different practice schedules in L2 pragmatic competence.

The current study seeks to address this gap by integrating two strands of research (i.e., L2 pragmatics and L2 practice) and investigating which practice schedule optimizes L2 pragmatic development. This study intends to contribute to the fields of instructed second language acquisition and instructional pragmatics by examining whether the superiority of interleaved practice over blocked practice found in non-L2 fields (e.g., cognitive psychology, chemistry, mathematics) and L2 grammar (e.g., Suzuki & Sunanda, 2019; Suzuki et al., 2022) can extend to L2 pragmatics learning/teaching. The present study is the first to explore the effects of blocked practice and interleaved practice on L2 pragmatic development.

Literature review

Theoretical background

The advantage of interleaved practice over blocked practice can be explained by the discriminative contrast hypothesis (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013), stating that interleaved practice helps learners distinguish similar concepts. When engaged in interleaved practice, learners devote their attention to the distinctions between constantly switched categories, which allows them to compare exemplars from similar categories and pay attention to subtle distinctions between these categories (e.g., comparable to the similarities between blueberries and blackberries). In other words, "when between-category discriminability is low, interleaved practice can be more beneficial than blocked practice" (Suzuki et al., 2022, p. 673). This phenomenon is known as *interleaving effects*. It is crucial to note that to produce interleaving effects, the categories or concepts need to be similar. When two categories are divergent from each other and the level of between-category discriminability is high (e.g., blueberries versus potatoes), the effectiveness of interleaved practice may become limited. For example,

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

the advantage of interleaved practice over blocked practice was found in Eglington and Kang's (2017) study, which explored undergraduate students' acquisition of similar organic chemical compounds. However, interleaving effects were not found in Carpenter and Muller's (2013) study, which examined how L2 learners of French acquired sounds with high between-category discriminability (e.g., /o/ in *cadeau* versus /s/ in *brebis*) assessed by multiple-choice questions.

Although prior studies in various fields have shown the advantage of interleaved practice over blocked practice, it is worth pointing out that these two types of practice schedules can be efficacious in different aspects. The superiority of blocked practice over interleaved practice might be found in L2 fluency development. The effectiveness of blocked practice in fluency may be explained by the speech production model (Levelt, 1989), which consists of three main elements: (a) the conceptualizer (e.g., generating a preverbal idea), (b) the formulator (e.g., grammatical encoding, phonological encoding, the encoding of form-function-context mappings), and (c) the articulator (e.g., producing output in an L2). According to the speech production model, when L2 learners complete a task for the first time, their attention is directed to working memory for the purpose of conceptualization, for instance, creating a preverbal idea. As a result, limited attentional resources can be allocated for formulation and articulation. However, practicing the same task consecutively (i.e., blocked practice) can familiarize learners with the task. Thus, they may free up their attentional resources for the encoding of form-function-context mappings and articulation, which might enhance the fluency of learners' L2 pragmatic performance.

Effectiveness of blocked practice and interleaved practice

No studies in the existing literature have investigated the effects of interleaved practice and blocked practice on L2 pragmatic competence. Most of the L2 practice studies exploring the influence of interleaved practice and blocked practice have been focused on grammar (Nakata & Suzuki, 2019; Suzuki et al., 2022), suggesting the superiority of interleaved practice over blocked practice on grammatical accuracy. For example, Nakata and Suzuki (2019) examined the effects of interleaved practice and blocked practice on Japanese English as a foreign language (EFL) learners' acquisition of syntactic structures with low between-category discriminability (i.e., the simple past tense and the present perfect tense; first-conditional, second-conditional, and third-conditional constructions). The results of grammaticality judgment tests showed that the interleaved-practice group significantly outperformed the blocked-practice group on a one-week delayed posttest. Similar results were found in Suzuki et al. (2022). They investigated the influence of interleaved practice and blocked practice on Japanese EFL learners' acquisition of five similar relative-clause constructions. The results of a picture description test demonstrated that the interleaved-practice group produced more accurate relative-clause constructions than the blocked-practice group on both immediate and delayed posttests.

Only a handful of studies have examined the influence of interleaved practice and blocked practice on L2 oral fluency. Importantly, the small number of studies have revealed inconclusive findings with respect to which practice schedule is more effective in enhancing fluency. For instance, in Suzuki's (2021) study, the results of oral narrative tasks showed that blocked practice led to greater fluency development (e.g., higher articulation rate) compared with interleaved practice. However, in Suzuki et al.'s (2022)

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

research, the results of a picture description test demonstrated slight superiority of interleaved practice over blocked practice in Japanese EFL learners' fluency in producing five similar relative-clause constructions on the immediate posttest. Due to the mixed findings, more studies are needed to further investigate the effects of interleaved practice and blocked practice on L2 oral fluency.

Note that the referenced studies have primarily focused on grammatical properties with simple and decontextualized language usage (e.g., using the simple past tense on a grammaticality judgment test). The influence of interleaved practice and blocked practice on contextualized language usage (e.g., making requests or suggestions) is unclear. This topic is worth exploring because effective practice schedules facilitate the proceduralization and automatization of L2 pragmatic knowledge, which can promote accuracy and fluency of learners' L2 pragmatic behaviors (Dekeyser, 2007). The accurate and fluent pragmatic performance produced through effective practice can further help EFL learners use English as a lingua franca to conduct intercultural communication smoothly, avoid communication misunderstandings, and establish rapport among interlocutors from various cultural backgrounds (Zhang, 2021, 2022).

Studies on practice in the field of instructional pragmatics

A myriad of (quasi-)experimental studies and meta-analyses (e.g., Plonsky & Zhuang, 2019; Taguchi, 2015) have reported the positive influence of instruction (e.g., consciousness-raising activities and metapragmatic explanations) on promoting L2 learners' pragmatic competence. Specifically, consciousness-raising activities (e.g., form-comparison tasks) draw L2 learners' attention to the target pragmatic feature and enhance its salience. The awareness that students establish from consciousness-raising activities might improve their subsequent intake of the target pragmatic feature (Schmidt, 1993). With respect to metapragmatic explanations, they "could ensure learners' pragmatic awareness at the level of understanding" (Li, 2019, p. 121) and acquaint students with L2 pragmatic knowledge (e.g., pragmalinguistic forms used for making suggestions). Learners who are equipped with pragmatic knowledge pertaining to the target pragmatic feature might be more likely to notice the feature during real-life communication in the L2 (Zhang, 2021). In addition to awareness-raising activities and metapragmatic explanations, both Plonsky and Zhuang (2019) and Taguchi (2015) reported that pragmatics instruction entailing practice is more efficacious in developing learners' L2 pragmatic competence than instruction without follow-up practice.

In the line of research on L2 pragmatics practice, a small number of studies have explored the effects of the intensity/amount of practice on L2 pragmatic competence (e.g., Li, 2012, 2013). For instance, in Li's (2012) study, a regular training group, an intensive training group (i.e., practicing the target speech act twice as much as the regular training group), and a control group (i.e., without practice) received metapragmatic instruction on making requests in Chinese. The results of oral discourse completion tasks (DCTs) showed that the intensive training group produced significantly more appropriate requests than the control group and the regular training group. In a subsequent study, Li (2013) examined the influence of the amount of practice on L2 pragmatic competence. The results of oral DCTs revealed that practicing requests four times was sufficient for enhancing the accuracy of the L2 learners' production of requests in Chinese.

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

To summarize, studies that examine the effectiveness of practice in L2 pragmatic development are sparse (Li, 2019). The small number of studies on L2 pragmatics practice have revealed the positive effects of repeated practice on L2 pragmatic competence (Li, 2012, 2013). A more intriguing question is how repeated practice should be arranged to optimize L2 pragmatics learning/teaching (e.g., blocked practice versus interleaved practice). Crucially, no studies have investigated the influence of practice schedules on L2 pragmatic competence. This strand of research is significant because practice schedules are instrumental in improving the quality of L2 pragmatics instruction. Furthermore, optimal practice schedules contribute to the proceduralization and automatization of L2 pragmatic knowledge.

Target pragmatic features: Suggestions and requests

The speech act of suggestions refers to "a language function whereby one of the interlocutors may express his/her belief that a course of action could be profitably taken up by either or both of them" (Marmaridou, 1990, p. 565). The speech act of requests is defined as the behavior of the speaker who expresses his/her "expectation of the hearer with regards to prospective action" (Blum-Kulka et al., 1989, p. 11). One of the reasons why these two pragmatic features were selected as the foci of the current study is that they are ubiquitous but face-threatening (Brown & Levinson, 1987). As far as hearers are concerned, making a request or suggestion may impose on the negative face of the addressee who wants to be undisturbed and the positive face of the addressee who strives for agreement (Brown & Levinson, 1987; Taguchi, 2022). Realizing these two speech acts can also be face-threatening for speakers, who might hesitate to make a suggestion or request "for fear of exposing a need or risking the hearer's loss of face" (Blum-Kulka et al., 1989, p. 11). The face-threatening characteristics of these two speech acts may pose a great challenge for L2 learners to appropriately perform these two pragmatic features in the target language. Thus, it is important to equip L2 learners with pragmatic knowledge and engage them in activities to practice making requests and suggestions.

Another reason for selecting these two pragmatic features is based on the discriminative contrast hypothesis (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013). The low between-category discriminability of making requests and suggestions can be reflected in two respects: (a) *pragmalinguistic similarities* and (b) *sociopragmatic similarities* (Thomas, 1983). Pragmalinguistic similarities refer to adopting similar linguistic forms to perform communicative acts, whereas sociopragmatic similarities refer to examining similar contextual factors (e.g., power, social distance) when making pragmatic choices. Specifically, with respect to pragmalinguistic similarities, the linguistic forms used for providing suggestions in Martínez-Flor and Alcón Soler's (2007) study and the linguistic devices used for making requests in Taguchi et al.'s (2015) study show that these two speech acts employ similar pragmalinguistic forms, entailing *head acts* (i.e., the core parts that are used for realizing the target act independently of other components), internal modifiers, and external modifiers. Table 1 presents examples of making requests and suggestions with similar head acts, internal modifiers, and external modifiers.

In addition to pragmalinguistic similarities, these two target pragmatic features share sociopragmatic similarities. When speakers make requests or suggestions, similar contextual variables, such as the power differences between interlocutors (P), their social distance (D), and the degrees of imposition (R), need to be considered (Brown &

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Table 1. Pragmalinguistic similarities between making requests and making suggestions

Pragmatic strategy	Pragmalinguistic form	Example of making suggestions	Example of making requests
Head acts	Using the main suggestion-making or request-making linguistic forms to convey the target illocutionary force.	"**I suggest** that we go to Yosemite this weekend."	"**Can you** lend me this book?"
Internal modifiers	Adopting syntactic downgraders (e.g., the conditional tense) and/or words (e.g., *maybe*, *possibly*) to reduce the degree of imposition of a request or suggestion.	"**Could** we **maybe** go to Yosemite this weekend?"	"**Could** you **possibly** write a letter of recommendation for me?"
External modifiers	Adopting auxiliary expressions to mitigate the degree of imposition before or after the head act of a request or suggestion.	"**I was wondering**, maybe we could have a biweekly meeting."	"**Would it be okay if** I get an extension on this project?"

Levinson, 1987). For instance, when one makes a high-imposition request or suggestion to the addressee who has more power and longer social distance (e.g., a student asks a professor for a letter of recommendation, an employee makes a suggestion to his/her employer), both speech acts necessitate examining the influence of contextual factors on their pragmatic choices. More politeness is required in PDR-high situations to save the addressee's face by adopting indirect strategies (e.g., using internal or external modifiers).[1]

The present study

Based on the discriminative contrast hypothesis (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013), this study employed a pretest-posttest-delayed-posttest research design and investigated the influence of interleaved practice and blocked practice on Chinese EFL learners' acquisition of two pragmatic features with low between-category discriminability—namely, the speech acts of requests and suggestions. Moreover, to further explore the efficacy of interleaved practice and blocked practice with respect to different aspects of L2 pragmatic development, this study examined both accuracy and fluency of L2 learners' pragmatic behaviors. The present study is the first to investigate the effects of practice schedules (i.e., interleaved practice versus blocked practice) on accuracy and fluency of L2 learners' pragmatic performance.

Another novel contribution of this study is germane to practice tasks. From a methodological standpoint, practice tasks in previous studies exploring the effects of practice schedules are primarily form-focused (e.g., practicing first-conditional,

[1]In PDR-low situations, interlocutors have equal power (P), the social distance between the interlocutors is short (D), and the degree of imposition (R) is small, whereas in PDR-high situations, the listener has more power, the social distance between the interlocutors is longer, and the degree of imposition is larger.

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

second-conditional, and third-conditional constructions through grammaticality judgment tasks). Thus, practice treatments in the extant literature lack authenticity and contextualization. To address this methodological shortcoming, the current study incorporated corpus-based practice based on natural, authentic, contextualized, and real-life situations to help L2 learners practice making suggestions and requests.

According to the findings of previous L2 grammar studies (Nakata & Suzuki, 2019; Suzuki et al., 2022), it is predicted that the advantage of interleaved practice over blocked practice will be found in promoting learners' L2 pragmatic competence in making requests and suggestions. The proceduralization of pragmatic knowledge is assessed by two parameters: (a) pragmatic accuracy and (b) fluency. The present research is guided by the following two research questions (RQs):

1) Is interleaved corpus-based practice more effective than blocked corpus-based practice in improving Chinese EFL learners' **pragmatic accuracy** and **fluency** when they make requests and suggestions?
2) Is the effectiveness of interleaved corpus-based practice or blocked corpus-based practice in Chinese EFL learners' pragmatic accuracy and fluency **long-lasting** over time?

Method

Participants

Sixty-three sophomores (35 females, 28 males) from EFL classes at a public university in China participated in this study. Their ages ranged from 19.5 to 22.6 years (M_{age} = 20.7). They were intermediate English learners based on their scores (M = 84.39, SD = 3.95) on the TOEFL (Test of English as a Foreign Language), which they took three months prior to this study.[2] Because the assessment tasks in this study involved speaking, the learners' scores in the speaking and listening segments of TOEFL were also reported. Their speaking and listening skills were at an intermediate level (speaking: M = 20.35, SD = 2.49; listening: M = 20.51, SD = .12).[3] A background survey showed that no students had lived outside of China and all of them intended to obtain their master's or doctoral degrees in the United States after graduating from college. Moreover, the students' responses in the survey and the interviews with their EFL instructors showed that their focused instruction did cover the metapragmatic explanations of the target speech acts.

All learners received explicit instruction germane to how to make pragmalinguistically and sociopragmatically appropriate requests and suggestions in English. After the pragmatics instruction, all students took the pretest (Details regarding the pretest are discussed in the section "Procedure"). Then, the 63 learners were randomly assigned to two groups to practice the two target pragmatic features. The interleaved-practice (IP) group consisted of 31 learners, whereas the blocked-practice (BP) group consisted of 32 learners.

[2]The total scores of TOEFL ranging from 42 to 91 are considered intermediate according to the guidelines provided by the official website of TOEFL: https://www.ets.org/toefl.html.

[3]Based on the guidelines provided by the official website of TOEFL, the scores ranging from 16 to 24 in the speaking section are considered intermediate, and the scores ranging from 9 to 21 in the listening section are considered intermediate.

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Procedure

The present study conducted a pretest, an immediate posttest, and a delayed posttest. The timeline of this study is illustrated in Figure 1.

During Week 1, all participants completed a background questionnaire. During Week 2, all students received two hours of pragmatics instruction on making requests and suggestions delivered by the researcher. The pedagogical design of the lecture was adapted from Li's (2013) research. The instruction started with an awareness-raising task where the participants were asked to circle the appropriate requests or suggestions in English (Details are discussed in the section "Materials for Instruction"). The instructor then provided the learners with explicit metapragmatic explanations with respect to how to make pragmalinguistically and sociopragmatically proper requests or suggestions in English (i.e., declarative pragmatic knowledge; details are provided in the section "Materials for Instruction"). After the instruction, the participants were expected to master the declarative knowledge, which was reflected by the pragmalinguistic and sociopragmatic accuracy of their requests and suggestions. An example of a pragmalinguistically and sociopragmatically accurate request in a PDR-high situation is "I was wondering if you could write a letter of recommendation for me." An example of a pragmalinguistically and sociopragmatically accurate suggestion in a PDR-high situation is "I would suggest that we have a biweekly meeting."

Then, all participants completed multimedia discourse completion tasks (MMDCTs) as the pretest (Details of MMDCTs are discussed in the section "Instrument"). The design of administering the pretest after the instruction was inspired by Li's (2012, 2013) studies and Li and Taguchi's (2014) research to "show the effects of practice only" (Li & Taguchi, 2014, p. 801) rather than the influence of metapragmatic instruction combined with practice. Having the pretest after the instruction can also avoid the potential influence of confounding variables, for instance, the variation in L2 learners' prior declarative knowledge of making requests or suggestions. Administering a pretest after the instruction controlled for this variable because the instruction acquainted all learners with the same declarative knowledge regarding the target speech acts. Another purpose of the pretest was to assess the students' initial knowledge regarding making requests and suggestions and to ensure that they successfully acquired the declarative knowledge (i.e., reflected by their high level of pragmatic accuracy) before moving on to the practice stage (Li, 2013). After the pretest, the learners were randomly assigned to an IP group or a BP group.

Inspired by the design of the practice schedules in Suzuki's (2021) study, at the practice stage of the current study, the IP group partook in intermixing practice (i.e., a-b-a-b-a-b …). For example, if the current practice task in the IP condition focuses on making a request, the subsequent practice involves making a suggestion. This pattern (i.e., request→ suggestion) was repeated 10 times in total. In contrast, the BP group first

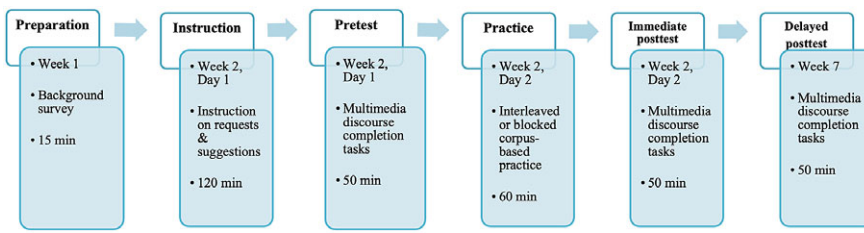

Figure 1. Timeline of the current study.

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

engaged in 10 practice tasks pertinent to making requests and then participated in 10 practice tasks involving making suggestions. In other words, if the current practice task in the BP condition pertains to making a request, the subsequent practice also describes a request-making scenario. It took approximately three minutes for the participants to complete each practice item. The entire practice session (i.e., 20 practice items) lasted one hour (More details are discussed in the section "Materials for Practice"). One thing worth mentioning is that the results from a series of pilot studies showed no order effects of the practice tasks. No significant difference was found between practicing making a request first and practicing making a suggestion first in either the BP ($p = .356$) or the IP condition ($p = .291$). The two groups' practice schedules in the current study are presented in Table 2.

After the practice session, both groups completed an immediate posttest. Moreover, to explore the retention of the benefits of the interleaved or blocked practice, both groups completed a delayed posttest five weeks later.

Materials for instruction

At the stage of the pragmatics instruction, two sets of handouts were provided for all students; one set of handouts was focused on making requests, and the other set was related to making suggestions. Specifically, regarding the set of handouts concerning requests, the participants took part in a consciousness-raising activity involving two pairs of dialogues pertinent to making requests (see Appendix A). Each pair of dialogues (e.g., Scenario 1A versus Scenario 1B) consisted of a pragmalinguistically and sociopragmatically appropriate request and a pragmalinguistically and sociopragmatically inappropriate request in English. The purpose of this handout was to raise the learners' awareness of the target speech act. The second handout provided a summary of the target pragmalinguistic forms and pragmatic strategies used for making requests (see Appendix B). The objective of the second handout was to acquaint the participants with explicit metapragmatic information with respect to making requests.

Similarly, the other set of handouts consisted of a consciousness-raising exercise related to making suggestions (see Appendix C) and a summary of the target pragmalinguistic forms and pragmatic strategies used for making pragmalinguistically and sociopragmatically appropriate suggestions in English (see Appendix D).

Materials for practice

Each target speech act entailed 10 practice tasks via computerized DCTs. Each DCT was designed based on the *Michigan Corpus of Academic Spoken English* (MICASE; Simpson et al., 2002). MICASE contains more than 190 hours of authentic real-life data from an array of speech events (e.g., conversations between friends in a PDR-low context, conversations between a professor and a student in a PDR-high situation) with 1.8 million words from 152 transcripts. This corpus encompasses various pragmatic features, including making requests and suggestions. The development of the materials for practice (i.e., corpus-based DCTs) consisted of three steps: (a) summarizing target pragmalinguistic forms used for making requests and suggestions based on previous studies (Martínez-Flor & Alcón Soler, 2007; Taguchi et al., 2015) (see Appendices B and D), (b) searching request-making and suggestion-making data in PDR-high and PDR-low situations from MICASE (Here is a request-making example in a PDR-high context extracted from Transcript ID: COL999MG053 in MICASE. A student, a first language

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Table 2. Practice schedules

Blocked practice	1. Request	2. Request	3. Request	4. Request	5. Request	6. Request	7. Request	8. Request	9. Request	10. Request
	11. Suggestion	12. Suggestion	13. Suggestion	14. Suggestion	15. Suggestion	16. Suggestion	17. Suggestion	18. Suggestion	19. Suggestion	20. Suggestion
Interleaved practice	1. Request	2. Suggestion	3. Request	4. Suggestion	5. Request	6. Suggestion	7. Request	8. Suggestion	9. Request	10. Suggestion
	11. Request	12. Suggestion	13. Request	14. Suggestion	15. Request	16. Suggestion	17. Request	18. Suggestion	19. Request	20. Suggestion

[L1] speaker of English, makes a request to a professor by asking, "I was wondering if you could comment on the differences of teaching at a small school versus a research university?"), and (c) developing corpus-based DCTs, including designing the prompts and background information based on the contexts of the examples from MICASE and creating exemplar responses by modifying the authentic data (e.g., replacing terminology with common words used in everyday life, removing repetitions and filler words).

For each corpus-based DCT, the learners first read a prompt describing the context and interlocutors. After 20 seconds, the participants were asked to provide their oral responses (i.e., requests or suggestions), which were recorded through the computer's built-in microphone (see Figure 2). Then, an exemplar response with a metapragmatic explanation of the target speech act was provided as feedback for 20 seconds (see Figure 3). In total, the practice session included 20 tasks (i.e., 10 practice tasks × 2 target speech acts).

Instrument: Multimedia discourse completion tasks

In spite of being one of the most adopted tools to measure L2 learners' pragmatic performance, DCTs have many shortcomings (Nguyen, 2019). For instance, many scholars have pointed out that data collected from DCTs usually do not reflect real-life exchanges because DCTs often lack contextual cues (Golato, 2003; Sykes & González-Lloret, 2020). Despite the downsides, the advantages of DCTs are manifold. For example, they are effective in eliciting data directly germane to the target pragmatic feature while controlling for extraneous variables (e.g., social distance, degrees of imposition)(Golato, 2003).

To address the drawbacks of traditional DCTs, MMDCTs (i.e., encompassing text-formatted prompts as well as dialogues presented aurally and visually via pictures) were designed for the current study. Specifically, to provide contextual cues, each MMDCT

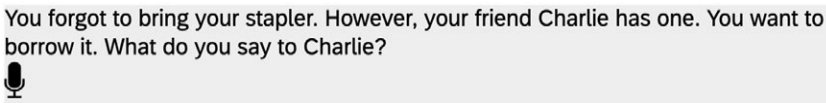

Figure 2. Sample practice task about making a request.

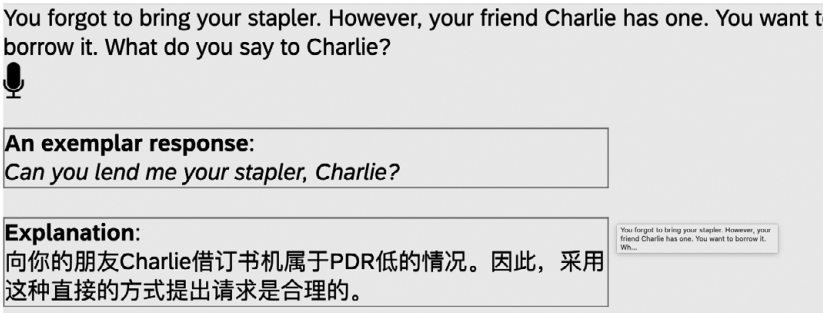

Figure 3. Sample feedback.
Note. The metapragmatic explanation translated into English is "Because borrowing a stapler from your friend Charlie is a PDR-low situation, adopting this direct strategy to make the request is appropriate."

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

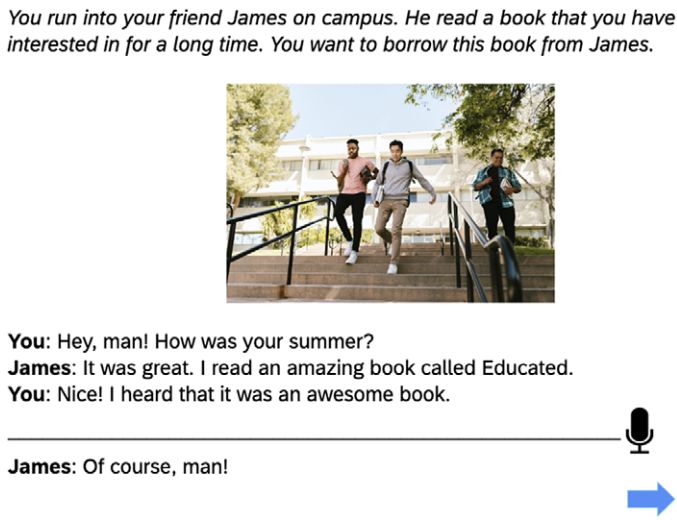

Figure 4. Sample MMDCT of making a request.

started with a prompt supplying contextual information (e.g., the relationship between conversationalists, background information). Additionally, a relevant image was incorporated below the prompt to provide more contextualization (Brown & Abeywickrama, 2018). The prompt, image, and incomplete dialogue disappeared after 30 seconds. The learners were then instructed to speak to the computers. Their answers regarding requests or suggestions were recorded through the computers' built-in microphones. After completing each MMDCT, the participants clicked on an arrow to proceed to the subsequent test item. Different from the practice stage, no feedback was provided at the testing stage. Each test lasted 50 minutes. Figure 4 and Figure 5 present a sample MMDCT involving making a request or suggestion, respectively.

Moreover, three equivalent versions of the tests were designed for the pretest, immediate posttest, and delayed posttest to avoid practice effects. The three tests were counterbalanced across the three testing stages between the IP group and the BP group. The internal reliability of each of the three tests as indexed by Cronbach's alpha coefficient was high for both the measure of pragmatic accuracy and the measure of fluency (i.e., pragmatic accuracy: pretest = .879, immediate posttest = .891, and delayed posttest = .893; fluency: pretest = .878, immediate posttest = .885, and delayed posttest = .896). Each test consisted of 32 items; 16 of them were target test items (i.e., eight items about requests and eight items about suggestions), and 16 test items were distractors (e.g., complimenting, congratulating). The sequence of the test items was randomized. Additionally, the target test items were piloted with 25 L1 speakers of American English and 11 proficient L2 speakers to make sure that the scenarios successfully elicited requests and suggestions. The target test items included in the three tests elicited requests and suggestions from 100% of the 36 L1 and L2 speakers.

With respect to the design of the target test items, first, each of them was designed based on the authentic data related to making requests or suggestions from MICASE (Simpson et al., 2002). Second, in accordance with the three contextual factors in Brown and Levinson's (1987) politeness theory (i.e., power, social distance, and the degree of imposition), the 16 target test items consisted of four PDR-high requests, four PDR-low

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Figure 5. Sample MMDCT of making a suggestion.

requests, four PDR-high suggestions, and four PDR-low suggestions. The distinctions between the PDR-high and PDR-low scenarios were confirmed through a pilot study. A questionnaire with a five-point Likert-type scale entailing the target test items was conducted with 22 L1 speakers of American English and 10 proficient L2 speakers. The questionnaire adapted Taguchi's (2022) survey and asked the 32 L1 and L2 speakers to rate how easy or challenging it was to perform the target pragmatic features (1 = *very easy*, 5 = *very challenging*). The results showed a significant difference between the PDR-high scenarios ($M = 4.10$, $SD = .29$) and the PDR-low scenarios ($M = 2.18$, $SD = .61$), $t = 16.336$, $p < .001$, $d = 2.888$ (large), suggesting that the target speech acts in the PDR-high scenarios were perceived as more challenging to perform compared with the target pragmatic features in the PDR-low situations. In addition, the scenarios in the target test items are different from the situations in the corpus-based DCTs at the practice stage.

Data analysis

The first RQ asked whether interleaved corpus-based practice is more effective than blocked corpus-based practice in improving Chinese EFL learners' **pragmatic accuracy** and **fluency** when they make requests and suggestions.

To assess **pragmatic accuracy** (i.e., declarative pragmatic knowledge), based on the speech act theory (Austin, 1962), politeness theory (Brown & Levinson, 1987), and the rating scales in former studies (e.g., Martínez-Flor & Alcón Soler, 2007), a four-level rubric (see Table 3) was adopted to measure the two groups' speech acts of making requests and suggestions, which share pragmalinguistic and sociopragmatic similarities as discussed in the section "Target Pragmatic Features: Suggestions and Requests." This rubric was also discussed with two EFL instructors who were L1 speakers of American English and have taught EFL for more than 10 years. With 16 target test items on the pretest, immediate posttest, and delayed posttest, respectively, the scores of each test ranged from 0 to 48. Three trained raters with PhD degrees in second language studies

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Table 3. Rubric for rating pragmatic accuracy of the target speech acts

Score	Descriptor	Example
3	–The communicative function is realized. –The target speech act is both **sociopragmatically and pragmalinguistically appropriate**.	A PDR-high scenario: *Ping gives suggestions to her professor about business etiquette in Chinese culture. She says:* "I **would probably suggest** that you use both hands to exchange business cards."
2	–The communicative function is somewhat realized. –The target speech act is **sociopragmatically appropriate but pragmalinguistically inappropriate**.	"I **would require** you to use both hands to exchange business cards."
	– Or the target speech act is **pragmalinguistically appropriate but sociopragmatically inappropriate**.	"I **suggest** that you receive business cards with both hands."
1	–The communicative function is not realized. –The target speech act is both **sociopragmatically and pragmalinguistically inappropriate**.	"I **require** you to exchange business cards with both hands."
0	The participant produces an irrelevant response.	"I hope you will have a great trip."

scored the data independently. High interrater reliability was achieved among the three raters for all three tests: pretest = .863, immediate posttest = .892, and delayed posttest = .920. The average scores among the raters were used for the data analyses.

To measure **fluency** (i.e., procedural pragmatic knowledge) of the learners' target speech acts, the current study focused on *speed fluency* assessed by speech rate. Even though speech rate is usually regarded as a global measure of fluency (Huensch & Tracy-Ventura, 2017), "it provides adequate indication of the overall speed at which a person speaks" (Bui et al., 2019, p. 5). Moreover, this temporal measure can extend the traditional analyses of target pragmatic features with a sole focus on accuracy and provide a complementary approach to dissecting L2 pragmatic production by assessing fluency (an indicator of processing), which is usually neglected in L2 pragmatics research. Specifically, speech rate in this study was measured by the average number of uttered English syllables per second when the participants made a request or suggestion, excluding false starts, repetitions, and repairs (Derwing et al., 2009). Three raters manually calculated the speech rate independently. High interrater reliability was achieved among the raters for all three tests (pretest = .882, immediate posttest = .891, and delayed posttest = .895).

SPSS 26 was used to conduct the statistical analyses in the current study. Before running the statistical analyses, the researcher checked the data to make sure that the assumptions for *t*-tests were met. First, the assumption of a continuous dependent variable has been met. In this study, the EFL learners received scores on a measurement scale, and there was an equal distance between each point on this scale. For instance, the difference in pragmatic accuracy between scoring 45 and 46 points on a 48-point test is the same as the difference between scoring 30 and 31 points, that is, one point. Second, according to the results of Shapiro–Wilk tests, the two groups' scores on the three tests were normally distributed (see Table 4). Third, the results of Levene's tests showed that the assumption of homogeneity of variances has been met (see Table 5).

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Table 4. Results regarding normality

Group	Test	*p* for pragmatic accuracy	*p* for fluency
Blocked-practice group	Pretest	.407	.818
	Immediate posttest	.831	.581
	Delayed posttest	.927	.414
Interleaved-practice group	Pretest	.594	.061
	Immediate posttest	.757	.782
	Delayed posttest	.101	.348

Table 5. Results regarding homogeneity

Test	*p* for pragmatic accuracy	*p* for fluency
Pretest	.556	.485
Immediate posttest	.249	.149
Delayed posttest	.353	.457

To answer RQ 1 investigating whether interleaved corpus-based practice is more effective than blocked corpus-based practice in improving Chinese EFL learners' **pragmatic accuracy** and **fluency** when they make requests and suggestions, an independent samples *t*-test was carried out to explore the between-groups differences on the immediate posttest. Furthermore, the effect size of the between-groups contrast was assessed (small = .40, medium = .70, large = 1.00; Plonsky & Oswald, 2014). To address RQ 2 regarding the **long-lasting** effectiveness of interleaved corpus-based practice or blocked corpus-based practice in Chinese EFL learners' pragmatic accuracy and fluency over time, a paired-samples *t*-test was conducted to examine the within-groups differences in the pretest data and the delayed posttest data. Furthermore, the effect sizes of the within-groups distinctions were evaluated (small = .60, medium = 1.00, large = 1.40; Plonsky & Oswald, 2014). In addition, an independent samples *t*-test was conducted to investigate the between-groups differences on the delayed posttest. It is worth mentioning that running multiple statistical analyses on the same data may enhance the possibility of committing a Type I error (Brown, 1990). One of the approaches to attenuating the adverse influence is using a Bonferroni correction, which generates a more conservative alpha level and decreases the likelihood of making a Type I error in such cases (Loewen & Plonsky, 2016). Therefore, in the current study, an alpha level of .05 was reduced to .017 due to conducting the *t*-tests three times on the data set ($\alpha = .05/3$). This new alpha level was used to examine the significance of the *p* values in this study.[4]

Prior to answering the two RQs, the researcher examined the pretest data to explore whether there were preexisting differences between the IP group and the BP group on

[4]To verify whether the *t*-tests and a mixed analysis of variance (ANOVA) would reach the same statistical conclusions, the researcher used SPSS 26 to conduct a mixed ANOVA with test (i.e., the pretest, immediate posttest, and delayed posttest) as a within-groups variable and group (i.e., the IP group and the BP group) as a between-groups variable. These two statistical tests arrived at the same conclusions. The results of the *t*-tests and the results of a mixed ANOVA (including post hoc analyses) consistently showed that the pragmatic accuracy of the IP group's target speech acts significantly outperformed the pragmatic accuracy of the BP group's target speech acts on the immediate and delayed posttests and the BP group's speech rate was significantly higher than the IP group's speech rate on the immediate and delayed posttests.

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

the pretest in terms of pragmatic accuracy and fluency. The results from an independent samples *t*-test demonstrated no noticeable difference in pragmatic accuracy of the target speech acts on the pretest between the IP group ($M = 39.98$, $SD = 3.71$) and the BP group ($M = 39.01$, $SD = 4.03$), $t(61) = .991$, $p = .326$, $d = .250$. Moreover, no significant difference in fluency of the target pragmatic features was found on the pretest between the IP group ($M = 2.69$, $SD = .32$) and the BP group ($M = 2.63$, $SD = .35$), $t(61) = .786$, $p = .435$, $d = .198$.

Results

This section reports the results in two aspects: (a) pragmatic accuracy and fluency on the immediate posttest and (b) pragmatic accuracy and fluency on the delayed posttest.

Pragmatic accuracy and fluency on the immediate posttest

First, with respect to the immediate effects of interleaved corpus-based practice and blocked corpus-based practice on **pragmatic accuracy**, Table 6 demonstrates the means, standard deviations, and 95% confidence intervals (CIs) of the two groups' scores on the pretest and the immediate posttest. The results of an independent samples *t*-test showed that the pragmatic accuracy of the IP group's target speech acts ($M = 46.32$, $SD = .85$) significantly outperformed the pragmatic accuracy of the BP group's target speech acts ($M = 39.80$, $SD = 2.99$) on the immediate posttest, $t(61) = 11.666$, $p < .001$, $d = 2.940$ (large), indicating the superiority of interleaved practice over blocked practice in pragmatic accuracy.

Second, in terms of the immediate effects of interleaved corpus-based practice and blocked corpus-based practice on **fluency**, Table 7 displays the means, standard deviations, and 95% CIs of the speech rate of each group's target speech acts on the pretest and the immediate posttest. Different from the advantage of interleaved practice

Table 6. Descriptive statistics of each group's scores of pragmatic accuracy on the pretest and the immediate posttest

				95% Confidence interval	
Group	Test	*M*	*SD*	Lower	Upper
Interleaved-practice group	Pretest	39.98	3.71	38.62	41.34
	Immediate posttest	46.32	.85	46.01	46.63
Blocked-practice group	Pretest	39.01	4.03	37.56	40.47
	Immediate posttest	39.80	2.99	38.72	40.88

Table 7. Each group's speech rate on the pretest and the immediate posttest

				95% Confidence interval	
Group	Test	*M*	*SD*	Lower	Upper
Interleaved-practice group	Pretest	2.69	.32	2.58	2.81
	Immediate posttest	2.75	.31	2.64	2.86
Blocked-practice group	Pretest	2.63	.35	2.50	2.75
	Immediate posttest	4.03	.45	3.87	4.19

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

over blocked practice found in pragmatic accuracy, the results from an independent samples *t*-test revealed that the BP group's speech rate ($M = 4.03$, $SD = .45$) was significantly higher than the IP group's speech rate ($M = 2.75$, $SD = .31$) on the immediate posttest, $t(61) = 13.223$, $p < .001$, $d = 3.332$ (large), suggesting the superiority of blocked practice over interleaved practice in fluency.

Pragmatic accuracy and fluency on the delayed posttest

First, with respect to the retention of the effectiveness of interleaved corpus-based practice or blocked corpus-based practice in **pragmatic accuracy**, Table 8 demonstrates the descriptive statistics of the two groups' scores on the three tests. Figure 6 presents the scores graphically. The results of a paired samples *t*-test showed that the IP group's pragmatic accuracy increased significantly from the pretest ($M = 39.98$, $SD = 3.71$) to the delayed posttest ($M = 47.12$, $SD = .64$), $t(30) = 11.078$, $p < .001$, $d = 1.990$ (large), indicating the long-term effectiveness of interleaved practice in pragmatic accuracy after five weeks. In contrast, no significant difference was found in the BP group's pragmatic accuracy between the pretest ($M = 39.01$, $SD = 4.03$) and the delayed posttest ($M = 40.20$, $SD = 3.38$), $t(31) = -1.324$, $p = .195$, $d = -.234$. Moreover, the results from an independent samples *t*-test showed that the IP group ($M = 47.12$, $SD = .64$) significantly outperformed the BP group ($M = 40.20$, $SD = 3.38$), $t(61) = 11.209$, $p < .001$, $d = 2.825$ (large) on the delayed posttest regarding the pragmatic accuracy of the target speech acts.

Second, with respect to the durable effects of interleaved corpus-based practice or blocked corpus-based practice on **fluency**, Table 9 displays the two groups' speech rate

Table 8. Descriptive statistics of each group's scores of pragmatic accuracy on the three tests

				95% Confidence interval	
Group	Test	*M*	*SD*	Lower	Upper
Interleaved-practice group	Pretest	39.98	3.71	38.62	41.34
	Immediate posttest	46.32	.85	46.01	46.63
	Delayed posttest	47.12	.64	46.89	47.36
Blocked-practice group	Pretest	39.01	4.03	37.56	40.47
	Immediate posttest	39.80	2.99	38.72	40.88
	Delayed posttest	40.20	3.38	38.99	41.42

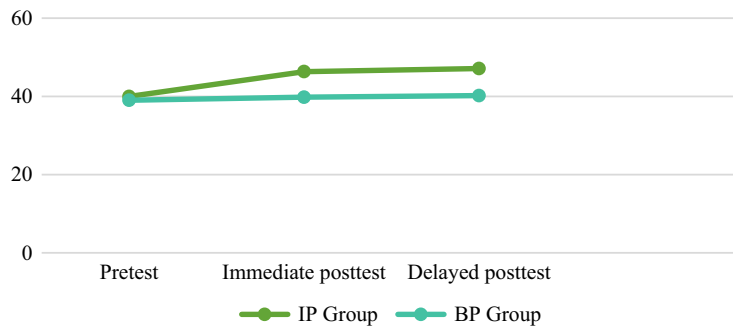

Figure 6. Two groups' scores of pragmatic accuracy on the three tests.

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Table 9. Each group's speech rate on the three tests

Group	Test	*M*	*SD*	95% Confidence interval	
				Lower	Upper
Interleaved-practice group	Pretest	2.69	.32	2.58	2.81
	Immediate posttest	2.75	.31	2.64	2.86
	Delayed posttest	2.79	.40	2.64	2.94
Blocked-practice group	Pretest	2.63	.35	2.50	2.75
	Immediate posttest	4.03	.45	3.87	4.19
	Delayed posttest	4.00	.50	3.82	4.18

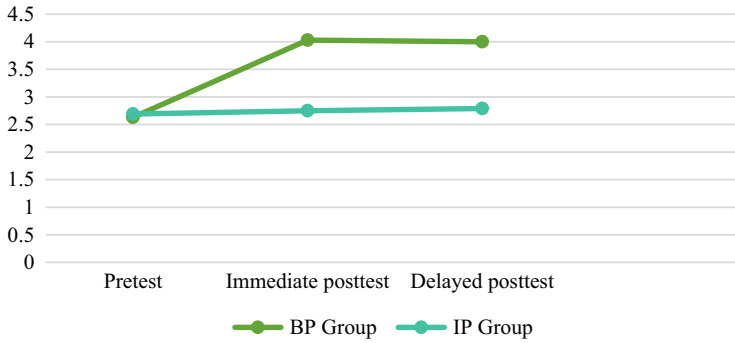

Figure 7. Each group's speech rate on the three tests.

on the three tests. Figure 7 illustrates the speech rate graphically. The results from paired samples *t*-tests revealed that the BP group's speech rate increased significantly from the pretest ($M = 2.63$, $SD = .35$) to the delayed posttest ($M = 4.00$, $SD = .50$), $t(31) = 12.199$, $p < .001$, $d = 2.156$ (large), indicating the long-term efficacy of blocked practice in fluency over time. However, there was no significant difference between the IP group's fluency on the pretest ($M = 2.69$, $SD = .32$) and their fluency on the delayed posttest ($M = 2.79$, $SD = .40$), $t(30) = -.966$, $p = .342$, $d = -.174$. Furthermore, the results from an independent samples *t*-test showed that the BP group's delayed posttest speech rate ($M = 4.00$, $SD = .50$) was significantly higher than the IP group's delayed posttest speech rate ($M = 2.79$, $SD = .40$), $t(61) = 10.594$, $p < .001$, $d = 2.670$ (large).

Discussion

The current study explored the effectiveness of interleaved corpus-based practice and blocked corpus-based practice in L2 pragmatic development. This section discusses the findings in two respects: (a) the immediate effects of interleaved practice and blocked practice on pragmatic accuracy and fluency and (b) the long-lasting influence of these two practice schedules on pragmatic accuracy and fluency.

Immediate effects of interleaved practice and blocked practice on pragmatic accuracy and fluency

The first RQ investigated whether interleaved practice is more effective in increasing EFL learners' pragmatic accuracy and fluency than blocked practice. First, regarding

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

pragmatic accuracy, the results of the immediate posttest revealed the superiority of interleaved practice over blocked practice in L2 pragmatic accuracy. This finding is congruent with the findings of previous L2 grammar studies (e.g., Suzuki et al., 2022). Focusing on Japanese EFL learners' application of syntactic structures with low between-category discriminability (i.e., five relative-clause constructions), the results of an immediate posttest in Suzuki et al. (2022) showed the advantage of interleaved practice over blocked practice in grammatical accuracy. Moreover, the finding of the current study corroborates the discriminative contrast hypothesis (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013), stating that interleaved practice is more effective in the acquisition of low between-category discriminability than blocked practice. When speech acts with low between-category discriminability (e.g., requests and suggestions) were compared, the interleaved practice promoted L2 pragmatic accuracy. The mixing exemplars regarding suggestions and requests might help the IP group differentiate the two similar categories (Kang & Pashler, 2012).

In addition to the discriminative contrast hypothesis, the transfer-appropriate processing theory (Morris et al., 1977) might account for the positive effects of interleaved practice on the IP group's pragmatic accuracy. This theory purports that learners' performance can be improved if the testing condition is congruent with the learning condition. The target test items on the immediate posttest were not presented consecutively (i.e., not in a sequence of request-request-request-suggestion-suggestion-suggestion). However, this testing condition was similar to how the practice tasks were presented for the IP group at the practice stage. These two similar conditions for the IP group—namely (a) the learning condition (i.e., practice tasks were presented in an intermixing manner) and (b) the testing condition of the immediate posttest (i.e., two sequential test items did not involve the same speech act)—may result in their significantly higher score than the BP group's score regarding pragmatic accuracy.

Contrary to the advantage of interleaved practice over blocked practice found in pragmatic accuracy, the results revealed the superiority of blocked practice over interleaved practice in **fluency** on the immediate posttest. The BP group's improved fluency indicates that the declarative pragmatic knowledge acquired during the pragmatics instruction was successfully converted into procedural pragmatic knowledge after the blocked practice (DeKeyser, 2007). This finding is consistent with the findings of previous studies on L2 oral fluency (Suzuki, 2021). Suzuki (2021) demonstrated that the repetition of the same tasks in a row helped the EFL learners' articulation rate increase on the immediate posttest, demonstrating the proceduralization of linguistic knowledge.

The facilitative effects of blocked practice on fluency development might be explained by the speech production model (Levelt, 1989). Drawing on this model, when L2 learners complete a task for the first time, their attention is directed to working memory for the purpose of conceptualization, for instance, creating preverbal ideas. Consequently, limited attentional resources can be allocated for formulation and articulation. The BP group practiced the same type of tasks involving the same pragmatic feature consecutively, which helped them become familiar with the task. This familiarization could alleviate the processing load required for conceptualization. Therefore, they might free up their attentional resources for the formulation of form-function-context mappings and articulation of their requests or suggestions, which may bring about more fluent speech on the immediate posttest.

Furthermore, from the perspective of *task repetitions*, during the practice treatment in the BP condition, it is likely that certain pragmalinguistic resources were activated in the first task and were subsequently reactivated in the following task involving the same

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

speech act. The constant reactivation of the same pragmalinguistic resources in the BP condition might contribute to the development of procedural pragmatic knowledge, which may improve the learners' fluent pragmatic performance. Thus, the positive effects of repeatedly practicing the same type of task involving the same pragmatic feature at the practice stage may be transferred to the immediate posttest. In contrast, the IP group could not use the same pragmalinguistic tools activated from the previous task because the preceding and current practice tasks did not focus on the same speech act.

Another way to interpret the IP group's less fluent target speech acts might be related to their improved pragmatic accuracy of the target pragmatic features on the immediate posttest. To formulate sociopragmatically and pragmalinguistically appropriate requests or suggestions, especially in the PDR-high situations, the IP group might need to spend more time analyzing the influence of the contextual factors (e.g., power) on their L2 pragmatic choices. It is likely that this cognitive process might lead to the IP group's slower and more cautious utterances. This finding aligns with the findings of previous L2 pragmatics studies (e.g., Taguchi, 2022). Taguchi (2022) found that L2 learners' speech rate was significantly lower when producing PDR-high pragmatic features than PDR-low pragmatic features. In fact, prior studies have further pointed out that less fluent utterances can be indicative of learners' pragmatic sensitivity and careful consideration of contextual factors (Taguchi, 2022), whereas immediate and fluent responses in PDR-high scenarios may sometimes be regarded as inconsiderate or rude (Tateyama, 2001).

This possible interpretation is confirmed by a further analysis of the IP group's suggestions and requests on the posttests. To demonstrate politeness and attenuate the face-threatening nature of making suggestions or requests in the PDR-high scenarios, the IP group adopted more supporting devices such as internal or external modifiers (e.g., "maybe," "I was wondering," "Would it be okay if") than the BP group. Moreover, almost one fifth of the learners from the IP group used internal and external modifiers in a combined manner to reduce the degrees of imposition (e.g., "I was wondering if we could possibly have a meeting on another day."). This cognitive process of analyzing the influence of contextual factors might result in the IP group's less fluent speech acts on the immediate posttest.

Long-term influence of interleaved practice and blocked practice on pragmatic accuracy and fluency

The second RQ explored whether the effectiveness of interleaved practice or blocked practice is durable in pragmatic accuracy and fluency. First, the long-term effectiveness of interleaved practice in **pragmatic accuracy** was found in the IP group after five weeks. This finding is in line with the findings of previous L2 grammar studies (e.g., Nakata & Suzuki, 2019; Suzuki et al., 2022). Nakata and Suzuki (2019) reported the interleaved-practice group's improved grammatical accuracy of using syntactic features with low between-category discriminability (e.g., first-conditional, second-conditional, and third-conditional constructions) on a one-week delayed posttest. Moreover, Suzuki et al. (2022) found the sustained effects of interleaved practice on Japanese EFL learners' acquisition of five similar relative-clause constructions on a one-week delayed posttest. It is noteworthy that the interval between the immediate and delayed post-posts in these two studies (i.e., Nakata & Suzuki, 2019; Suzuki et al., 2022) was only one week. The fact that the lasting influence of interleaved practice was found on a delayed posttest

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

conducted five weeks later provides strong evidence for the durable effectiveness of interleaved practice in accuracy.

The long-lasting efficacy of interleaved practice found in pragmatic accuracy on the delayed posttest may be explained by *spacing effects*—that is, *distributed practice* (i.e., practice tasks are distributed over time) produces better retention of improved performance than *massed practice* (i.e., practice tasks appear sequentially without intervals). Spacing effects have been found in previous studies, for instance, L2 vocabulary (Nakata, 2015) and grammar (Miles, 2014). In the present study, the blocked practice entailing tasks focusing on one speech act conducted in a successive manner was akin to massed practice, whereas the interleaved practice containing tasks intermixed with two pragmatic features was congruent with distributed practice. Similar to distributed practice, interleaved practice may bring about better recall because learners encode materials differently in each practice task, which can lead to a variety of retrieval cues for learners to capitalize on (Glenberg, 1979). Spacing effects might contribute to the long-term effectiveness of interleaved practice in the IP group's pragmatic accuracy after five weeks.

The long-lasting efficacy of blocked practice in **fluency** may be related to task repetition. The line of research on task repetition has reported the durable effects of repeated tasks (the counterpart of blocked practice) on L2 fluency (e.g., De Jong & Perfetti, 2011). For example, the results of a four-week delayed posttest in De Jong and Perfetti (2011) showed the sustained effectiveness of task repetition in the learners' fluency. The authors explained that task repetition led to proceduralization of linguistic knowledge, which resulted in increased L2 fluency.

Conclusions, future directions, and implications

This pretest-posttest-delayed-posttest study investigated the effects of interleaved practice and blocked practice on L2 pragmatic development. The results revealed the positive and long-term influence of interleaved practice on pragmatic accuracy (declarative pragmatic knowledge) of the EFL learners' suggestions and requests. This finding suggests that the effectiveness of interleaved practice in accuracy found in the field of L2 grammar (e.g., Suzuki et al., 2022) might extend to L2 pragmatics acquisition, especially the acquisition of similar pragmatic features. Moreover, the results showed the facilitative and durable effects of blocked practice on fluency (procedural pragmatic knowledge) of the L2 learners' target speech acts. This finding offers more evidence for the effectiveness of blocked practice in fluency shown in previous L2 speaking studies (Suzuki, 2021). Furthermore, these findings fill the gap in the domains of L2 pragmatics and L2 practice and contribute to the fields of instructed second language acquisition and instructional pragmatics by casting light on the efficacy of practice schedules in different aspects of L2 pragmatic development. The positive effects of interleaved practice on pragmatic accuracy and facilitative effects of blocked practice on fluency found in the current study suggest that the advantage of interleaved practice or blocked practice is contingent on the dimensions of L2 pragmatic development (i.e., pragmatic accuracy or fluency) and the types of pragmatic knowledge (i.e., declarative pragmatic knowledge or procedural pragmatic knowledge).

From a methodological standpoint, this study enhances the authenticity and contextualization of practice tasks by incorporating corpus-based practice and engaging learners in practice tasks based on natural, authentic, contextualized, and real-life scenarios. From a theoretical perspective, the current study lends support to the

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

discriminative contrast hypothesis (Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013), stating that when the between-category discriminability of target features is low, interleaved practice is more effective than blocked practice with respect to accuracy.

One limitation of the present study is that all participants were Chinese EFL learners. Future studies may recruit L2 learners from various L1 backgrounds. This can contribute to the generalizability of the findings of this study. Another promising direction is to investigate the interplay between individual differences (e.g., aptitude) and practice schedules in relation to L2 pragmatic development. Moreover, in response to the recent calls for expanding the scope of research exploring the influence of interleaved practice and blocked practice (Suzuki, 2021), future research may examine the influence of interleaved practice and blocked practice on the acquisition of less similar pragmatic features. With respect to measuring fluency of L2 learners' pragmatic production, future studies may include learners' regular L2 speech rate (e.g., through narration tasks or opinion tasks) as a baseline in addition to assessing their speech rate of producing target pragmatic feature(s).

The findings of this study provide useful pedagogical implications for L2 teachers and curriculum developers pertaining to designing lesson plans and developing curricula. This study suggests positive and long-term effects of interleaved practice on pragmatic accuracy. After providing metapragmatic explanations of two similar pragmatic features for students, L2 teachers and curriculum developers may ask students to engage in interleaved practice. Learners may practice one pragmatic feature first and then practice the other one. Moreover, interleaved practice may enable learners to accurately perform pragmatic features in real-life contexts, where pragmatic features are not blocked by category (Nakata & Suzuki, 2019). In other words, interleaved practice has more ecological validity than blocked practice in that interleaved practice resembles real-life exchanges. We usually employ various pragmatic features discursively instead of using the same pragmatic feature consecutively in real-world communication. In addition, the findings of the present study reveal facilitative and durable effects of blocked practice on fluency. To foster learners' fluency development, L2 instructors and curriculum developers may incorporate blocked practice and instruct students to complete a series of practice tasks focusing on one pragmatic feature and then proceed to tasks involving other pragmatic features.

Acknowledgments. I would like to express my sincere gratitude to the editors of *Studies in Second Language Acquisition* and the anonymous reviewers for their valuable suggestions and helpful comments on earlier versions of this article.

Competing interests. The author declares none.

References

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural pragmatics: Requests and apologies*. Ablex Publishing Corporation.

Brown, H. D., & Abeywickrama, P. (2018). *Language assessment: Principles and classroom practices*. Pearson Longman.

Brown, J. D. (1990). The use of multiple *t*-tests in language research. *TESOL Quarterly*, *24*, 770–773. https://doi.org/10.2307/3587135

Brown, P., & Levinson, S. D. (1987). *Politeness: Some universals in language usage* (2nd ed.). Cambridge University Press.

Bui, G., Ahmadian., M. J., & Hunter, A.-M. (2019). Spacing effects on repeated L2 task performance. *System*, *81*, 1–13. https://doi.org/10.1016/j.system.2018.12.006

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*, 671–682. https://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*, 481–495. https://doi.org/10.3758/s13421-013-0371-0

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*, 533–568. https://doi.org/10.1111/j.1467-9922.2010.00620.x

DeKeyser, R. M. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). Routledge.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*, 533–557. https://doi.org/10.1017/S0272263109990015

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

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.

Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, *24*, 90–121. https://doi.org/10.1093/applin/24.1.90

Huensch, A., & Tracy-Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, *101*, 275–293. https://doi.org/10.1111/modl.12395

Kang, S. H. (2016). The benefits of interleaved practice for learning. In J. C. Horvath, J. M. Lodge, & J. Hat-tie (Eds.), *From the laboratory to the classroom: Translating science of learning for teachers* (pp. 79–93). Routledge.

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

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Li, S. (2012). The effects of input-based practice on pragmatic development of requests in L2 Chinese. *Language Learning*, *62*, 403–438. https://doi.org/10.1111/j.1467-9922.2011.00629.x

Li, S. (2013). Amount of practice and pragmatic development of request-making in L2 Chinese. In N. Taguchi & J. Sykes (Eds.), *Technology in interlanguage pragmatics research and teaching* (pp. 43–69). John Benjamins.

Li, S. (2019). Cognitive approaches in L2 pragmatics research. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 113–127). Routledge.

Li, S., & Taguchi, N. (2014). The effects of practice modality on the development of pragmatic performance in L2 Chinese. *The Modern Language Journal*, *98*, 794–812. https://doi.org/10.1111/modl.12123

Loewen, S., & Plonsky, L. (2016). *An A–Z of applied linguistics research methods*. Palgrave.

Marmaridou, S. (1990). Contrastive analysis at discourse level and the communicative teaching of languages. In J. Fisiak (Ed.), *Further insights into contrastive analysis* (pp. 561–571). Benjamins.

Martínez-Flor, A., & Alcón Soler, E. (2007). Developing pragmatic awareness of suggestions in the EFL classroom: A focus on instructional effects. *Canadian Journal of Applied Linguistics*, *10*, 47–76. https://journals.lib.unb.ca/index.php/CJAL/article/view/19734

Miles, S. W. (2014). Spaced vs. massed distribution instruction for L2 grammar learning. *System*, *42*, 412–428. https://doi.org/10.1016/j.system.2014.01.014

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, *37*, 677–711. doi:10.1017/S0272263114000825

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

Nguyen, T. T. M. (2019). Data collection methods in L2 pragmatics research: An overview. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 195–210). Routledge.

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

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Plonsky, L., & Zhuang, J. (2019). A meta-analysis of second language pragmatics instruction. In N. Taguchi (Ed.), *The Routledge handbook of second language acquisition and pragmatics* (pp. 287–307). Routledge.

Rohrer, D., Dedrik, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*, 1323–1330. https://doi.org/10.3758/s13423-014-0588-3

Schmidt, R. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 43–57). Oxford University Press.

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*, 419–440. https://doi.org/10.1006/jmla.2001.2813

Simpson, R. C., S. L. Briggs, J. Ovens, & J. M. Swales. (2002). *The Michigan corpus of academic spoken English*. The Regents of the University of Michigan. https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase

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

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*, 169–197.

Suzuki, Y., Yokosawa, S., & Aline, D. (2022). The role of working memory in blocked and interleaved grammar practice: Proceduralization of L2 syntax. *Language Teaching Research*, *26*, 671–695. https://doi.org/10.1177/1362168820913985

Sykes, J., & González-Lloret, M. (2020). Exploring the interface of interlanguage (L2) pragmatics and digital spaces. *CALICO Journal*, *37*, i–xv. https://doi.org/10.1558/cj.40433

Taguchi, N. (2015). Instructed pragmatics at a glance: Where instructional studies were, are, and should be going. *Language Teaching*, *48*, 1–50. https://doi.org/10.1017/S0261444814000263

Taguchi, N. (2022), Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, *56*, 308–335. https://doi.org/10.1002/tesq.3070

Taguchi, N., Naganuma, N., & Budding, C. (2015). Does instruction alter the naturalistic pattern of pragmatic development? A case of request speech act. *TESL-EJ*, *19*, Article ej75. https://www.teslej.org/wordpress/issues/volume19/ej75/ej75a3/

Tateyama, Y. (2001). Explicit and implicit teaching of pragmatic routines: Japanese sumimasen. In G. Kasper & K. Rose (Eds.), *Pragmatics in language teaching* (pp. 200–222).Cambridge University Press. https://doi.org/10.1017/CBO9781139524797.015

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, *4*, 91–111. https://doi.org/10.1093/applin/4.2.91

Zhang, Y. (2021). Combining computer-mediated communication with data-driven instruction: EFL learners' pragmatic development of compliment responses. *System*, *103*, Article 102624. https://doi.org/10.1016/j.system.2021.102624

Zhang, Y. (2022). The influence of practice contexts on L2 learners' compliment responses: Telecollaborative video-based practice versus face-to-face practice. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2152054

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16–27. https://doi.org/10.3758/s13421-012-0238-9

Appendix A

Handout: Awareness-raising activities regarding making requests

Please circle the more appropriate request(s) in English in each pair of the conversations (e.g., the request in Scenario 1A versus the request in Scenario 1B).

Scenario 1A:
Lisa will attend a conference this week. She asks her professor to reschedule the midterm exam. She says:

Hello Professor Jones. **Can you reschedule the midterm exam for next week?**

Scenario 1B:
Nate will apply for graduate schools this summer. He asks his professor to write him a letter of recommendation. He says:

Hello Professor Davis. **I was wondering if you could write a letter of recommendation for me.**

Scenario 2A:
James forgot to bring his textbook. He wants to borrow his friend Elle's textbook. He says:

I want to borrow your textbook for a second.

Scenario 2B:
George runs into a physics problem. He wants to ask his roommate John for help. George says:

Hi John. Can you help me with this physics problem?

Appendix B

Handout: A summary of the target pragmalinguistic forms used for making requests

Category	Target Pragmalinguistic Form	Example of Making Requests
Direct Strategy	1. Imperatives 2. Performatives 3. Obligation Statements 4. Questions	"**Lend** me this book, please." "**I want you to lend** me this book." "You **should** lend me this book." "**Can you** lend me this book?"
Indirect Strategy	5. Modal verbs (e.g., *could*, *would*) 6. Hedging (e.g., *possibly*, *maybe*) 7. External modifiers (e.g., *I was wondering*)	"**Could** you write a letter of recommendation for me?" "Could you **possibly** write a letter of recommendation for me?" "**I was wondering** if you could write a letter of recommendation for me."

https://doi.org/10.1017/S0272263123000062 Published online by Cambridge University Press

Appendix C

Handout: Awareness-raising activities regarding making suggestions

Please circle the more appropriate suggestion(s) in English in each pair of the conversations (e.g., the suggestion in Scenario 1A versus the suggestion in Scenario 1B).

Scenario 1A:
Dylan gives suggestions to his internship supervisor on having a biweekly meeting with all the interns.

You should have a biweekly meeting with us.

Scenario 1B:
Margaret gives suggestions to her professor, who will teach Computer Science II for the first time.

You could maybe go over the key concepts from Computer Science I before talking about new content.

Scenario 2A:
Ben is planning a three-day trip to Yosemite next week. His friend Carter gives him some tips on visiting national parks.

I require you to pack a large water bottle.

Scenario 2B:
Melissa will graduate from college this summer. Her older sister May gives her some suggestions on job applications. May says:

I suggest that you create a LinkedIn account.

Appendix D

Handout: A summary of the target pragmalinguistic forms used for making suggestions

Category	Target Pragmalinguistic Form	Example of Making Suggestions
Direct Strategy	1. Imperatives	"**Pack** a large water bottle."
	2. Performatives	"**I suggest that you** pack a large water bottle."
	3. Obligation Statements	"You **should** pack a large water bottle."
	4. Questions	"**Why don't you** pack a large water bottle?"
Indirect Strategy	5. Modal verbs (e.g., *could*, *would*)	"You **could** ask all the interns to take an online questionnaire."
	6. Hedging (e.g., *possibly*, *maybe*)	"**Maybe** you could ask all the interns to take an online questionnaire."
	7. External modifiers (e.g., *It would be helpful if you, I was wondering*)	"**I was wondering** if we could have a meeting on another day."

Cite this article: Zhang, Y. (2023). The effects of interleaved and blocked corpus-based practice on L2 pragmatic development. *Studies in Second Language Acquisition,* 45: 812–837. https://doi.org/10.1017/S0272263123000062