

The effects of interleaving versus blocking on foreign language pronunciation learning

Shana K. Carpenter · Frank E. Mueller

Published online: 16 January 2013
© Psychonomic Society, Inc. 2013

Abstract Many studies have shown that students learn better when they are given repeated exposures to different concepts in a way that is shuffled or *interleaved*, rather than *blocked* (e.g., Rohrer *Educational Psychology Review*, 24, 355–367, 2012). The present study explored the effects of interleaving versus blocking on learning French pronunciations. Native English speakers learned several French words that conformed to specific pronunciation rules (e.g., the long “o” sound formed by the letter combination “eau,” as in *bateau*), and these rules were presented either in blocked fashion (*bateau, carreau, fardeau . . . mouton, genou, verrou . . . tandis, verglas, admis*) or in interleaved fashion (*bateau, mouton, tandis, carreau, genou, verglas . . .*). Blocking versus interleaving was manipulated within subjects (Experiments 1–3) or between subjects (Experiment 4), and participants’ pronunciation proficiency was later tested through multiple-choice tests (Experiments 1, 2, and 4) or a recall test (Experiment 3). In all experiments, blocking benefited the learning of pronunciations more than did interleaving, and this was true whether participants learned only 4 words per rule (Experiments 1–3) or 15 words per rule (Experiment 4). Theoretical implications of these findings are discussed.

Keywords Interleaving · Pronunciation learning · Discriminative contrast

In academic settings, students are often given multiple exposures to different concepts. Under typical conditions,

a particular concept (e.g., conjugating foreign verbs in the present tense) is practiced over and over again before moving on to the next concept (e.g., conjugating foreign verbs in the past tense). However, recent evidence suggests that it may be more beneficial to present different concepts in an order that is shuffled and less predictable (e.g., practicing a present-tense conjugation followed by a past-tense conjugation, followed by more present-tense conjugations, etc.). For example, Rohrer and Taylor (2007) taught students to calculate the volume of 4 different types of solid figures. Students worked through practice problems in an order that was either *blocked* by type of figure (i.e., all problems pertaining to one type of figure were finished before the student moved on to the next type of figure) or *interleaved* such that the same problems appeared in an order that was shuffled and unpredictable. On a later test requiring students to calculate the volumes for similar types of figures, students scored higher if they had learned the information through interleaving, as compared with blocking.

Other studies have reported benefits of interleaving on the learning of geometric concepts (e.g., Taylor & Rohrer, 2010) and algebraic rules (e.g., Mayfield & Chase, 2002). A small but growing number of studies has also reported benefits of interleaving for other types of cognitive and motor tasks as well (for reviews, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Rohrer, 2012). For example, Kornell and Bjork (2008) taught participants to classify paintings of particular artists by presenting example paintings in an order that was either blocked by artist (e.g., participants saw several paintings by Georges Braque, followed by several paintings by Judy Hawkins, and then Bruno Pessani, etc.) or interleaved such that paintings by one particular artist did not occur consecutively (e.g., Braque, Hawkins, Pessani, etc.). On a later test requiring participants to classify new paintings by the same artists, participants were better at matching the artist to

S. K. Carpenter (✉) · F. E. Mueller
Department of Psychology, Iowa State University,
W112 Lagomarcino Hall,
Ames, IA 50011, USA
e-mail: shacarp@iastate.edu

the painting when they had learned the artists through interleaving, as compared with blocking.

These findings were recently replicated by Kang and Pashler (2012) using a similar type of task, and benefits of interleaving have also been reported for a task in which participants learned to classify birds into their correct familial categories (e.g., finches, warblers, etc.; Wahlheim, Dunlosky, & Jacoby, 2011). Significant benefits of interleaving have been found even under conditions in which the degree of temporal spacing has been controlled (e.g., by inserting unrelated filler material between successive presentations of items from the same category, as in Kang & Pashler, 2012; see also Mitchell, Nash, & Hall, 2008; Taylor & Rohrer, 2010), suggesting that the benefits of interleaving do not appear to be due simply to the effects of temporal spacing.

One explanation for the benefits of interleaving is that it promotes discriminative contrast between concepts that are easily confused (see, e.g., Rohrer, 2012). For example, in Rohrer and Taylor's (2007) study, interleaving the types of solid figures affords students the opportunity to compare the solution for the current problem (e.g., calculating the volume of a wedge) with the solution for the previously presented problem (e.g., calculating the volume of a spheroid), allowing an opportunity to compare key differences between the solutions. Calculating the volume for the same type of figure in back-to-back blocked fashion provides less of an opportunity to notice these differences and may render it harder to distinguish between the different solutions later on. Indeed, Taylor and Rohrer (2010) observed a benefit of interleaving over blocking for solving mathematics problems that related to prisms and found that students who learned the problems through blocking, rather than interleaving, made more errors in discrimination—that is, mistakenly applied the solution for one type of problem to a different type of problem—on the final test.

In further support of the *discriminative contrast* hypothesis, interleaving has been shown to produce similar benefits for category induction learning when compared with a situation in which participants are allowed to simultaneously view multiple examples from different categories (e.g., Kang & Pashler, 2012). The benefits of interleaving have also been shown to be greater under simultaneous, rather than sequential, viewing conditions, which presumably encourage the processing of each category's distinguishing features (e.g., Wahlheim et al., 2011).

Another explanation for the benefits of interleaving, not necessarily mutually exclusive from the discriminative contrast hypothesis, is the notion that interleaving promotes more effective retrieval practice than does blocking (see, e.g., Dunlosky et al., 2013). Presentation of an item representing a particular concept can remind participants of earlier items representing that same concept. When presentations are blocked, the retrieval of previous items representing a given concept is likely to be impoverished

because these representations are not very old and may still be active in working memory (see also, e.g., Rohrer, 2012). Memory representations for interleaved items, on the other hand, are more likely to have been deactivated at the time the next one is presented, increasing the chances that retrieval of a previous interleaved item will be more elaborate or “complete” than retrieval of a previous blocked item. Given that retrieval of prior information benefits learning (e.g., Roediger & Butler, 2011; Roediger & Karpicke, 2006a, b), particularly under conditions in which the retrieval process is more elaborate (e.g., Carpenter, 2009, 2011; Carpenter & DeLosh, 2006; Pyc & Rawson, 2009), interleaving may be generally more beneficial than blocking because it encourages elaborative retrieval of previous items.

According to these hypotheses, the benefits of interleaving may be diminished in tasks that are less likely to require discriminative contrast or that involve stimuli that are unlikely to be retrieved during interleaved presentations. The present study explored these possibilities by using a task that has not been fully explored in any of the known research on interleaving. The task required native English speakers to learn the pronunciations for unfamiliar French words. With this task, participants learned orthographic-to-phonological mapping rules (e.g., the long “o” sound formed by the “eau” letter combination in the French word *bateau*) through repeated exposures to words that represented those rules. Participants were presented with multiple examples of French words sharing the same rule, either in blocked fashion (e.g., *bateau, fardeau, rameau . . . tandis, brebis, vernis . . . darder, combler, valser*, etc.) or in interleaved fashion (e.g., *bateau, tandis, darder, fardeau, brebis, combler, rameau, vernis, valser*, etc.), and were later tested for their pronunciation proficiency of these words.

There are reasons to expect that benefits of interleaving might not occur in this task. First, the pronunciation learning task may be less likely to require discriminative contrast, as compared with some of the tasks used in previous studies. These studies have typically presented similar-looking stimuli (e.g., paintings depicting natural landscapes; Kang & Pashler, 2012; Kornell & Bjork, 2008) and required participants to extract key differences between the stimuli that were diagnostic of category membership (e.g., unique brush strokes used by different artists). The birds used in the study by Wahlheim et al. (2011) shared similar physical features, making it difficult for a novice to identify the unique features (e.g., the shape of the beak) that defined a particular bird family. If the task involves the identification of distinguishing features from similar-looking examples, a method that highlights the differences between categories (i.e., interleaving) may benefit learning more than a method that highlights the similarities within a category (i.e., blocking).

However, when the goal is to notice commonalities among different-looking stimuli, blocking may be

advantageous because it helps learners notice the features that are shared by members of the same category. For example, Goldstone (1996, Experiment 2) presented participants with line segments that contained various degrees of distortion from their respective prototypes. The categories represented by the line segments were alternated either frequently (i.e., similar to interleaving, such that line segments belonging to the same category rarely occurred back-to-back) or infrequently (i.e., similar to blocking, such that line segments belonging to the same category often occurred back-to-back). A later test that required categorizing new line segments revealed better performance under conditions in which the categories had been learned through infrequent (i.e., blocked), rather than frequent (i.e., interleaved), alternations. Learning a set of different pronunciation rules would seem to rely more on noticing the common features among stimuli than on noticing key differences between them, so interleaving may not benefit learning in the current task because of its potential to rely more on the analysis of shared features than on discriminative contrast.

Second, to the extent that retrieval underlies the benefits of interleaving (e.g., Dunlosky et al., 2013), those benefits may be reduced or eliminated in a task for which retrieval of prior examples is very difficult or unlikely. Unlike the visual tasks that have been used in past research on interleaving (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim et al., 2011), pronunciation learning requires knowledge of auditory-to-visual mapping. Auditory memory traces are brief and may last only a few seconds (e.g., Baddeley, Thomson, & Buchanan, 1975), making it difficult to remember the phonological characteristics of an earlier item. Common features among interleaved stimuli may therefore be difficult to notice because memory for the phonology of the earlier item may have already been forgotten by the time the next one is presented. The phonological trace of an earlier blocked item, on the other hand, is more likely to remain in working memory at the time the next one is presented, making it easier to notice common features and extract a rule or pattern among stimuli.

These possibilities were explored in four experiments varying the way in which blocked versus interleaved presentations were administered (within or between subjects), the way in which pronunciation proficiency was assessed (multiple-choice or recall tests), the amount of exposure to the stimuli, and whether or not participants were aware ahead of time that the words conformed to specific pronunciation rules. Experiment 1 explored the effectiveness of blocking versus interleaving on learning French pronunciations under conditions in which participants were not informed that the words conformed to any pronunciation rules. They simply saw the words presented (half through blocking and half through interleaving) and then completed a multiple-choice test requiring them to choose the correct

pronunciation for each word out of three options that were provided.

Experiment 1

Method

Participants

Nineteen native English speakers who reported having no prior familiarity with the French words participated in order to fulfill partial requirements for introductory psychology courses at Iowa State University.

Materials and design

Materials consisted of 64 French words, 8 of which represented one of 8 distinct pronunciation rules (see Appendix 1). During encoding, a randomly selected half of these words (4 words per rule) were presented to each participant according to a blocked versus interleaved schedule. The method of presentation was based closely on that of Kornell and Bjork (2008, Experiment 1; see also Wahlheim et al., 2011). For 4 of the 8 pronunciation rules, the 4 example words were presented in blocked fashion, and for the other 4 rules, the words were interleaved. The items for each rule were randomly assigned for each individual participant to a predetermined sequence of blocked (B) or interleaved (I) groups of items that were presented in the order BIIBBIIB. For example, a participant would first see and hear a blocked group of 4 words that represented a single rule (e.g., *bateau, fardeau, tonneau, rameau*), followed by an interleaved group of 4 words that each represented a different rule (e.g., *baver, chardon, osseux, admis*). This was followed by another interleaved group that demonstrated the same 4 rules as the previous interleaved group (e.g., *tamiser, chacal, neveux, tandis*). Two more blocked groups were then presented consecutively (e.g., *cervelle, malade, emplette, cuvette; bouton, genou, mouton, voulu*), followed by 2 more interleaved groups (e.g., *darder, cochon, brumeux, compris; raconter, charnel, vaniteux, verglas*) and, finally, by 1 more blocked group (e.g., *pavot, carnet, sommet, brevet*). The pronunciation rules that were assigned to blocked versus interleaved conditions, and the order in which the 4 items of a given rule were presented was always randomly determined for each individual participant. Each of the 32 words was presented only once.

A multiple-choice test was created by producing 3 different audio recordings of each word. The recordings were made by a native speaker of French who created 1 correct pronunciation for each word and 2 wrong pronunciations. One type of wrong pronunciation was created by pronouncing the rule-relevant portion of the word correctly (e.g., the

long “o” at the end of *bateau*) and a different part of the word incorrectly (e.g., the “a” in *bateau*). We refer to this as the *rule-correct* pronunciation, since the portion of the word relevant to the rule was pronounced correctly. For the other type of wrong pronunciation, the rule-relevant portion of the word was pronounced incorrectly (e.g., a long “a” at the end of *bateau*), whereas the rest of the word was pronounced correctly. We refer to this as the *rule-incorrect* pronunciation. The participant’s job was to listen to each of the 3 recordings for each word and choose which one they believed was correct. Although only 1 response was completely correct, choosing the rule-correct option might indicate that participants had acquired some knowledge of the rule.

Procedure

Participants began by reading instructions on the computer screen, informing them that they would be learning the pronunciations for a number of French words and would later be tested over their knowledge of these pronunciations. The instructions did not inform participants that the words conformed to specific pronunciation rules but simply encouraged participants to do their best to learn the pronunciations for each word that they were about to see and hear. Before beginning the experiment, participants put on headphones and listened to a recording of an example French word (not included among the experimental stimuli) to verify that they could hear the audio stimuli.

Participants then saw and heard the 32 words (4 words conforming to each of 8 pronunciation rules) in either blocked or interleaved fashion, as described previously. Participants were not informed of these presentation methods or the order in which they were delivered. They simply saw and heard a continuous presentation of 32 items, 1 at a time. For each presentation, the French word was presented alone in the middle of the computer screen, simultaneously with an audio recording of a native French speaker pronouncing the word. Each word was presented on the screen for 4 s, with a 1-s blank screen occurring between each word.

Immediately after seeing and hearing the 32 words, participants were given instructions for the multiple-choice test. These instructions informed participants that they would be tested over their knowledge of the pronunciations of the words that they had just seen and heard, in addition to new words that they had not seen or heard but that have similar pronunciations. The order of items on the test was grouped into 8 blocks, each consisting of 1 item that represented 1 of the 8 pronunciation rules. Since participants were exposed to the correct answer as 1 of the options on the test, it is possible that learning could occur during the test (see Kornell & Bjork, 2008). Grouping items into these blocks ensured that any 2 words representing the same pronunciation rule would not occur too close together and made it possible for us to assess

any learning that might be occurring on the test by examining performance across test blocks.

On the multiple-choice test, participants first encountered the 32 items that they had seen and heard previously, followed by the 32 new items that they had not seen or heard. Each item on the test was presented 1 at a time in the center of the computer screen, simultaneous with the 3 audio recordings of the word (correct, rule-correct, and rule-incorrect). The order of the 3 recordings was randomly determined on each individual trial for each participant. Participants listened to all 3 recordings and then were asked to decide which 1 they thought was correct. They made their decision by pressing the 1, 2, or 3 number key to indicate whether they believed that the correct pronunciation was the first, second, or third option, respectively.

After completing this test for all 64 words, participants answered a question assessing their metacognitive awareness of the effects of blocking versus interleaving. Instructions on the computer screen explained the nature of the blocked versus interleaved presentation and asked participants to indicate which method they believed had helped them learn the pronunciations better. Participants entered a response by pressing 1 of 2 keys to indicate either blocked or interleaved presentation. Participants were then asked whether their native language was English and whether they were familiar with any of the French words prior to participating in the experiment. Data were replaced for any participant whose native language was not English or who reported having familiarity with any of the French words.

Results

Multiple-choice accuracy

Table 1 (top section) displays the rate at which participants chose each of the 3 options (correct, rule-correct, and rule-incorrect) on the multiple-choice test. Participants chose the correct option more often after learning words through blocking, as compared with interleaving. A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) repeated measures analysis of variance (ANOVA) on correct responses revealed a significant advantage of blocking over interleaving, $F(1, 18) = 13.99$, $p = .001$, $MSE = .01$, and a significant advantage for old words over new words, $F(1, 18) = 5.67$, $p = .028$, $MSE = .01$, but no interaction, $F = 1.61$.

The blocking advantage for correct responses was fairly stable across final test blocks, and overall performance did not vary across test blocks. A 2 (presentation method: blocking vs. interleaving) \times 8 (test block) repeated measures ANOVA revealed a significant advantage of blocking ($M = .51$, $SD = .12$) over interleaving ($M = .43$, $SD = .12$), $F(1, 18) = 13.95$, $p = .002$, $MSE = .04$, but no main effect of test block, $F = 1.05$, and no interaction, $F = 0.94$.

Table 1 Proportions of options chosen on the multiple-choice test as a function of word type (old vs. new) and presentation method (blocking vs. interleaving)

	Old words			New words		
	Correct	Rule-correct	Rule-incorrect	Correct	Rule-correct	Rule-incorrect
Experiment 1						
Blocking	.56 (.12)	.22 (.10)	.22 (.08)	.46 (.18)	.29 (.15)	.24 (.12)
Interleaving	.44 (.15)	.34 (.13)	.22 (.12)	.42 (.14)	.31 (.11)	.27 (.12)
Experiment 2						
Blocking	.57 (.13)	.23 (.10)	.20 (.11)	.51 (.13)	.27 (.10)	.22 (.13)
Interleaving	.47 (.18)	.30 (.12)	.23 (.13)	.47 (.11)	.29 (.12)	.23 (.11)
Experiment 4						
Blocking	.59 (.08)	.17 (.04)	.23 (.07)	.54 (.10)	.23 (.10)	.23 (.10)
Interleaving	.52 (.10)	.23 (.08)	.25 (.09)	.53 (.10)	.26 (.04)	.21 (.11)

Note. Standard deviations are given in parentheses. For the *correct* option, the entire word was pronounced correctly; for the *rule-correct* option, the portion of the word containing the rule was pronounced correctly, and a different part of the word was pronounced incorrectly; for the *rule-incorrect* option, the portion of the word containing the rule was pronounced incorrectly, and the rest of the word was pronounced correctly

Multiple-choice errors

Analysis of error responses revealed that participants were more likely to choose the rule-correct option after learning words through interleaving, as compared with blocking. A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) repeated measures ANOVA on rule-correct responses revealed that participants chose this option more often after learning the words through interleaving, as compared with blocking, $F(1, 18) = 10.42$, $p = .005$, $MSE = .01$, and there was a significant interaction whereby this difference was greater for old words than for new words, $F(1, 18) = 5.14$, $p = .036$, $MSE = .01$. The same analysis applied to rule-incorrect responses indicated that these errors were committed slightly more often for new words than for old words, $F(1, 18) = 3.60$, $p = .07$, $MSE = .01$; however, no effects were observed for blocking versus interleaving, and there was no interaction, $F_s < 1$.

Metacognition

We examined participants' responses to the postexperiment question regarding which presentation method they believed was more effective. Responses to this question were lost for 4 participants due to a programming error. Of the remaining 15 participants, 12 (80 %) indicated that they believed blocking was more effective than interleaving.

Discussion

Experiment 1 revealed significant benefits of blocking over interleaving on the learning of French pronunciations. This result differs from those of previous studies that have

reported benefits of interleaving for other types of learning, such as classifying paintings (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008) or birds (e.g., Wahlheim et al., 2011). Whereas the rate of rule-incorrect errors did not differ as a function of blocking versus interleaving, the rate of rule-correct errors was significantly higher after words had been learned through interleaving, as compared with blocking. This seems to indicate that participants acquired some knowledge of the rule (e.g., the correct sound of *eau* in *cadeau*) through interleaving but did not learn the pronunciation of the entire word (e.g., the correct sounds of both *cad* and *eau*) as well through interleaving as they did through blocking.

In Experiment 1, participants were not informed at the beginning of the experiment that the words conformed to specific pronunciation rules. It is possible that interleaving is beneficial for pronunciation learning but that participants in Experiment 1 did not notice the pronunciation rules as well as they could have, because they were not aware at the outset of the experiment that these rules existed. When participants are aware that the pronunciation rules exist, they may be more likely to notice the shared features among items. If interleaving enhances their ability to do this, benefits of interleaving over blocking may be more likely to emerge under conditions in which participants are more likely to notice the common features among stimuli. Experiment 2 was designed to explore this possibility.

Experiment 2

Experiment 2 replicated the same design as that in Experiment 1, but this time participants were informed at the beginning of

the experiment that the to-be-learned words conformed to specific pronunciation rules. Participants were encouraged to try to discover those rules by paying close attention to the appearance and sound of each word. The presentation and test phases were then identical to those in Experiment 1.

Method

Participants

Twenty-five individuals were recruited from the same participant pool as before. All participants reported being native English speakers and having no prior familiarity with the French words used in the experiment.

Materials, design, and procedure

All materials and procedural details were identical to those in Experiment 1, with one exception. After reading the same basic instructions from Experiment 1, participants in Experiment 2 were informed that the French words they were about to learn represented specific pronunciation rules that they should try to discover. To illustrate this, participants were given an example of a French pronunciation rule that did not appear among the experimental stimuli (the “oi” sound). They saw and listened to three example words that illustrated this rule (*avoir*, *boisson*, and *voici*). When participants had verified that they understood the task, they completed the experiment in the same fashion as in Experiment 1.

Results

Multiple-choice accuracy

As is displayed in Table 1 (middle section), blocking again resulted in a higher proportion of correct responses than did interleaving on the multiple-choice test. A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) repeated measures ANOVA on correct responses again revealed a significant advantage of blocking over interleaving, $F(1, 24) = 5.01$, $p = .035$, $MSE = .02$, but no significant advantage for old words over new words, $F = 1.67$, and no interaction, $F = 1.26$.

The rate of correct responses across test blocks was again fairly stable and revealed a consistent advantage of blocking over interleaving. A 2 (presentation method: blocking vs. interleaving) \times 8 (test block) repeated measures ANOVA revealed a significant advantage of blocking ($M = .54$, $SD = .10$) over interleaving ($M = .47$, $SD = .12$), $F(1, 24) = 5.01$, $p = .035$, $MSE = .09$, but no main effect of test block and no interaction, $F_s < 1$.

Multiple-choice errors

Analysis of error responses revealed the same trend as that in Experiment 1, in that participants were more likely to choose the rule-correct option after learning words through interleaving, as compared with blocking. A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) repeated measures ANOVA on rule-correct responses revealed that these errors were committed more often after words had been learned through interleaving rather than through blocking, $F(1, 24) = 3.25$, $p = .08$, $MSE = .01$, but no effect for type of word and no interaction, $F_s < 1$. The same analysis applied to rule-incorrect responses revealed no significant effects or interactions, $F_s < 1$.

Metacognition

As in Experiment 1, the majority of participants indicated that they believed blocking was more effective than interleaving. Of the 25 participants, 23 (92 %) indicated that blocking helped them learn the pronunciations better than did interleaving.

Discussion

Experiment 2 replicated the same blocking advantage as that found in Experiment 1, this time under conditions in which participants were informed ahead of time about the existence of pronunciation rules. Thus, even when participants were aware that the words conformed to specific rules and were encouraged to discover those rules, the pronunciations were learned better through blocking than through interleaving.

Experiment 3 was designed to provide further data on the effects of blocking versus interleaving on pronunciation learning. This time, pronunciation proficiency was assessed through a test of recall, instead of multiple choice. This test required participants to produce a spoken pronunciation of the word rather than choose the correct pronunciation from the options provided.

Experiment 3

Experiment 3 replicated the same procedure as that in Experiment 2, but this time using a final test of recall, instead of multiple choice. After receiving the same instructions and seeing and hearing the words presented in the same fashion as in Experiment 2, participants in Experiment 3 were presented with each of the 64 words 1 at a time, and they were asked to pronounce each word out loud. Each of their pronunciations was

recorded and scored for accuracy by native French speakers.

Method

Participants

Twenty-six individuals were recruited from the same participant pool as before. All participants reported being native English speakers and having no prior familiarity with the French words used in the experiment.

Materials, design, and procedure

All procedural details were identical to those in Experiment 2, except that participants in Experiment 3 were tested for their pronunciation proficiency by attempting to pronounce each French word out loud. For the pronunciation test, each word was presented one at a time, and participants were asked to pronounce the word as best they could by speaking it into a microphone. After doing so, participants pressed a button to advance to the next word. After completing the pronunciation test, participants answered the same questions as those in Experiment 2 and were then thanked and debriefed.

Results

Scoring

The scoring system was developed by a native speaker of French, who scored each pronunciation on a scale from 1 to 4. A score of 1 was assigned to a pronunciation that was not correct in any way, such that it was difficult to identify the word that the participant was attempting to pronounce. A score of 2 was assigned to a pronunciation that contained imperfections, but not to such a strong degree that the identity of the word was obscured. A score of 3 was assigned to a pronunciation that was basically correct but with a noticeable accent. A score of 4 was assigned to a pronunciation that contained no detectable flaw or accent, such that the word appeared to be spoken by a native speaker. Half ratings (e.g., 1.5) were assigned to pronunciations that did not fit precisely into the 1–4 categories.

Two native speakers of French used this system to score 30 % of the responses in blind fashion. Whether accuracy was computed as a mean rating (i.e., the average score from 1 to 4) or as a proportion correct (i.e., the percentage of responses that received a 3 or higher), the interrater correlations between the two scorers were significantly positive across the four conditions (blocking vs.

interleaving for old vs. new words) (r s ranged from .71 to .95, all p s < .05). The remaining responses were scored in blind fashion by a single rater.

Pronunciation accuracy

A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) repeated measures ANOVA on the proportion correct scores revealed a significant advantage of blocking ($M = .42$, $SD = .17$) over interleaving ($M = .33$, $SD = .16$), $F(1, 25) = 8.28$, $p = .008$, $MSE = .02$, and an advantage of old words ($M = .41$, $SD = .16$) over new words ($M = .35$, $SD = .14$), $F(1, 25) = 7.96$, $p = .009$, $MSE = .01$, but no interaction, $F = 1.76$. The same analysis applied to average ratings revealed a significant main effect of blocking ($M = 2.38$, $SD = .31$) over interleaving ($M = 2.19$, $SD = .31$), $F(1, 25) = 10.91$, $p = .003$, $MSE = .08$, and an advantage for old words ($M = 2.34$, $SD = .30$) over new words ($M = 2.23$, $SD = .29$), $F(1, 25) = 6.64$, $p = .016$, $MSE = .04$. An interaction also emerged, $F(1, 25) = 5.93$, $p = .022$, $MSE = .03$, indicating that the difference in average ratings between the blocked and interleaved conditions was larger for old words ($M = 2.47$, $SD = 0.35$, and $M = 2.20$, $SD = 0.33$, respectively) than for new words ($M = 2.29$, $SD = 0.34$, and $M = 2.18$, $SD = 0.34$, respectively).

Metacognition

Participants' responses to the metacognition question again revealed the belief that blocking was more effective than interleaving. This tendency was even stronger in Experiment 3 than in the previous experiments. This time, all 26 participants (100 %) indicated that they believed that blocking was more effective than interleaving.

Discussion

Blocking led to superior pronunciation learning over interleaving when proficiency was assessed through either multiple-choice tests (Experiments 1 and 2) or a recall test (Experiment 3). One possible explanation for the blocking advantage so far is that participants may not have received enough exposure to the material. Learning may be slower through interleaving than through blocking (e.g., Rohrer, 2012), and in the present experiments, participants encoded only 4 words per rule. Blocking may have led to faster discovery of the rule than did interleaving, which may create conditions under which a greater number of blocked items than interleaved items have a chance to be encoded after the rule has been discovered.

When the presentations are interleaved, perhaps 4 words per rule is not enough to promote discovery of the rule and

additional encoding of items after the rule has been discovered. If interleaving is beneficial for pronunciation learning, but only after the rule has been discovered, a greater amount of exposure to the material should be more likely to reveal benefits of interleaving over blocking. Experiment 4 was designed to address this possibility.

Experiment 4

In Experiment 4, participants received a greater amount of exposure to the materials than in the previous experiments. Whereas participants in Experiments 1–3 learned 4 items per rule, participants in Experiment 4 learned 15 items per rule. Also, unlike in the previous experiments, blocking versus interleaving was manipulated between subjects, and pronunciation proficiency was assessed after a 5-min retention interval.

Method

Participants

Thirty-eight individuals were recruited from the same participant pool as before. All participants reported being native English speakers and having no prior familiarity with the French words used in the experiment. Nineteen participants were randomly assigned to learn the pronunciations through blocking, and 19 through interleaving.

Materials, design, and procedure

Experiment 4 used 5 of the pronunciation rules from the previous experiments (the *eau*, *er*, *eux*, *t*, and *e* endings). Twelve new words were added to the 8 existing words that represented each of the 5 rules, resulting in 20 words per rule (see Appendix 2). Fifteen words per rule were randomly selected for each participant and were presented for participants to see and hear during the encoding phase. Later, during the test phase, pronunciation proficiency was assessed for the 15 old words representing each rule that had previously been seen and heard, in addition to the 5 new words representing each rule that had not previously been seen or heard.

During encoding, participants in the blocking group saw and heard 15 words in immediate succession that shared the same rule. After these 15 words, participants saw and heard the next group of 15 words, representing a different rule, and so on until participants had seen and heard all 15 items for each of the five rules. Participants in the interleaving group saw and heard a group of 5 items that each represented a different rule. They then saw and heard another group of 5 items that represented the same rules, in the same order as before, and so on until they had seen and heard all 15 items for each of the 5 rules. In both conditions, the order of

presentation of each rule and the items within each rule were randomly determined for each participant.

At the beginning of the experiment, participants were given the same instructions as those from Experiments 2 and 3, in which they were informed about the existence of pronunciation rules and encouraged to discover them. As in the previous experiments, each word was presented on the computer screen for 4 s, simultaneously with an audio recording of a native French speaker pronouncing the word. Unlike in the previous experiments, however, metacognitive accuracy was not assessed, because blocking versus interleaving was manipulated between subjects. Also, unlike in the previous experiments, the encoding phase was followed by a distractor task in which participants were asked to name as many of the 50 U.S. states as they could within a 5-min time period. Following this, participants completed the same type of multiple-choice test as in Experiments 1 and 2.

Results

Multiple-choice accuracy

Experiment 4 confirmed the same pattern of results as in Experiments 1 and 2 (see Table 1, bottom section). A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) mixed ANOVA on correct responses revealed a marginally significant advantage for blocking over interleaving, $F(1, 36) = 2.99, p = .09, MSE = .01$. No effect emerged for type of word, $F = 1.21$; however, a marginally significant method \times word type interaction emerged, $F(1, 36) = 3.95, p = .05, MSE = .005$, indicating that the advantages of blocking over interleaving were stronger for old words than for new words.

The rate of correct responses across test blocks was again fairly stable and revealed a significant overall advantage of blocking over interleaving. A 2 (presentation method: blocking vs. interleaving) \times 20 (test block) mixed ANOVA revealed a significant advantage of blocking ($M = .58, SD = .08$) over interleaving ($M = .52, SD = .08$), $F(1, 36) = 5.48, p = .025, MSE = .14$, but no main effect of test block, $F = 1.04$. A significant interaction emerged, $F(19, 684) = 1.75, p = .025, MSE = .042$, such that the blocking advantage was particularly strong during blocks 14 and 15 (in which the blocking group chose the correct response 25 % more often, on average, than did the interleaving group), as compared with the other blocks (in which the blocking group chose the correct response 5 % more often, on average, than did the interleaving group).

Multiple-choice errors

Analysis of error responses revealed the same trend as that in Experiments 1 and 2, in that participants were more likely

to choose the rule-correct option after learning words through interleaving, as compared with blocking. A 2 (presentation method: blocking vs. interleaving) \times 2 (type of word: old vs. new) mixed ANOVA on rule-correct responses revealed that these errors were committed more often after words had been learned through interleaving, as compared with blocking, $F(1, 36) = 6.13$, $p = .018$, $MSE = .01$. A significant effect of word type also emerged, indicating that more rule-correct errors were made for new words than for old words, $F(1, 36) = 9.00$, $p = .005$, $MSE = .003$. No interaction emerged between presentation method and word type, $F = 1.53$. The rate of rule-incorrect errors did not differ as a function of presentation method, $F < 1$, or word type, $F = 1.57$.

Discussion

Experiment 4 confirmed the same blocking advantage as that observed in Experiments 1–3, but this time under conditions in which participants received 15 exposures to the stimuli representing each rule. These findings suggest that the blocking advantage does not appear to have been driven by an insufficient amount of exposure to the stimuli. Experiment 4 also replicated the advantage of blocking over interleaving in a between-subjects design, under conditions in which participants were tested for pronunciation proficiency 5 min after seeing and hearing the words.

General discussion

Four experiments demonstrated consistent benefits of blocking over interleaving for the learning of French pronunciations. This finding occurred in both a within-subjects design (Experiments 1–3) and a between-subjects design (Experiment 4), under conditions in which pronunciation proficiency was assessed either through a multiple-choice test (Experiments 1, 2, and 4) or through a recall test (Experiment 3) and under conditions in which participants received an extended amount of exposure to the stimuli (Experiment 4).

Blocking was particularly advantageous for helping participants choose a fully correct pronunciation over one that contained correct pronunciation of just the rule. Through immediate repetition of words sharing the same rule, a given rule may be more obvious and easier to notice during blocked presentation than during interleaved presentation. Once the rule is noticed (which would presumably occur sooner during blocked than during interleaved presentations), for subsequent words that share that same rule, participants can more readily focus on the portions of the words that do not contain the rule. For example, after noticing that *eau* makes a long *o* sound, participants are already aware of

this and can devote more attention on subsequent trials to the “non-*eau*” portions of the words (e.g., the *cad* in *cadeau*, *bat* in *bateau*, or *fard* in *fardeau*). This two-part approach (i.e., first learning the sound of *eau* and then the sounds of the “non-*eau*” parts) may facilitate pronunciation learning more than a whole-word approach (i.e., trying to simultaneously learn the sound of both the repetitive *eau* and the other, “non-*eau*” parts during the presentation of each word).

This may explain the tendency for participants to remember the entire word (both rule and nonrule features) better through blocking than through interleaving. It may also explain why the advantage of blocking over interleaving tended to be stronger for old words than for new words. Although the pronunciation rules (e.g., *eau*) are present in both old and new words, old words are more likely than new words to contain the specific nonrule features (e.g., *cad*, *bat*, or *fard*) that were presumably encoded better through blocking than through interleaving.

The advantage of blocking over interleaving differs from the findings of past studies reporting benefits of interleaving on category induction learning. A possible explanation for these findings is that the effectiveness of interleaving may depend upon the processing requirements of the task. Past studies reporting benefits of interleaving used tasks that required discriminative contrast (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Wahlheim et al., 2011), whereas the present task may have been more likely to require the analysis of shared features. The finding that interleaving impairs analysis of shared features is consistent with the findings of previous studies that have required participants to notice commonalities from diverse stimuli, such as line segments (e.g., Goldstone, 1996) and geometric patterns (e.g., Kurtz & Hovland, 1956). The present study confirms and extends these findings to a pronunciation learning task in which participants had to notice patterns in orthographic-to-phonological mapping.

It is also possible that the advantage of blocking over interleaving may be driven partly by the nature of the task itself. Given the brief duration of auditory memory traces (e.g., Baddeley et al., 1975), interleaving may be likely not to benefit pronunciation learning, because the auditory characteristics of earlier items may be too difficult to access. Other studies using difficult or complex tasks have failed to observe benefits of interleaving. For example, de Croock and van Merriënboer (2007) taught participants to diagnose and troubleshoot specific types of malfunctions that could occur in a distiller system (e.g., pipe leakage, sensor malfunction, etc.). Participants worked through practice cases in an order that was either blocked by type of malfunction or interleaved such that the order of problems representing different types of malfunctions was shuffled. On a later test that involved diagnosing the same types of problems,

participants performed better if they had received blocked presentations during learning. It is possible that with enough practice, participants may be able to learn complex concepts as well or better through interleaving than through blocking. Although, in Goldstone's (1996) study, even after 600 exemplars of each category had been presented, participants still learned the line segments better when the categories changed infrequently (more similar to blocking) rather than frequently (more similar to interleaving).

Future studies could further explore the efficacy of the *retrieval hypothesis*, perhaps by comparing blocked presentation with different schedules of interleaved presentations that vary in the number of intervening items that occur between repetitions of the pronunciation rule. It is possible that shorter interleaved presentations (e.g., containing only one or two intervening items) afford more of an opportunity for participants to retrieve previous auditory information and may benefit pronunciation learning more than do blocked presentations. A more direct test of the discriminative contrast hypothesis could also be conducted by comparing blocked versus interleaved presentation schedules for learning pronunciation rules that are easily confused (e.g., vowel combinations that look and sound similar but are not pronounced identically), rather than a relatively diverse set of rules, as in the present study.

Across Experiments 1–3, 92 % of participants indicated that blocking helped them learn the pronunciations better than did interleaving. This is consistent with previous studies reporting that participants tend to choose blocking over interleaving as a more effective learning method, even if actual learning benefits more from interleaving (e.g., Kornell & Bjork, 2008). Such metacognitive judgments may be based on the degree to which a given method enhances the fluency or ease of encoding (e.g., Benjamin, Bjork, & Schwartz, 1998; Kornell, Rhodes, Castel, & Tauber, 2011; Oppenheimer, 2008). Seeing and hearing the pronunciation rules in blocked fashion may have rendered them easier to notice than seeing and hearing them in interleaved fashion. In previous studies, participants have reported that they learned material better when it was presented in a fashion that was designed to increase its ease or fluency of encoding, such as increasing the font size of text (e.g., Rhodes & Castel, 2008), increasing the coherency of text (e.g., Rawson & Dunlosky, 2002), or providing pictures with text (e.g., Carpenter & Olson, 2012; Serra & Dunlosky, 2010).

Sometimes, judgments based on ease or fluency can be accurate indicators of how well students have learned something (e.g., Rawson & Dunlosky, 2002), but fluency can sometimes give the erroneous impression that something has been well-learned because it was easy to encode (e.g., Carpenter & Olson, 2012). Future research should further explore the role of fluency in metacognitive awareness of

the effects of blocking versus interleaving and the conditions under which this heuristic may facilitate the accuracy of metacognitive judgments for a variety of different materials.

The present results represent the effects of blocking versus interleaving on pronunciation learning after a relatively short (i.e., up to 5 min) retention interval. It is possible that these effects may differ across a longer retention interval. The length of the retention interval has sometimes affected other memory phenomena, such as the testing effect (e.g., Coppens, Verkoeijen, & Rikers, 2011; Roediger & Karpicke, 2006a; Toppino & Cohen, 2009) and the spacing effect (e.g., Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Rohrer & Taylor, 2006), such that these manipulations sometimes produce stronger effects after longer, as opposed to shorter, retention intervals. Previous studies on interleaving have shown significant benefits of interleaving after relatively short retention intervals of several seconds (e.g., Kornell & Bjork, 2008; Wahlheim et al., 2011) or minutes (e.g., Kang & Pashler, 2012), suggesting that reliable effects of interleaving can occur even when memory is assessed relatively soon after learning. Using a comparable retention interval, the present study showed that blocking appears to be more beneficial than interleaving for learning pronunciations. A worthwhile question for future research is whether the effects of interleaving versus blocking might vary across time for learning different types of materials.

Another viable question for future research is how the learning of pronunciations may be affected by different ways of encoding the material. In the present study, participants saw and heard each word only once. Learning may be differentially affected by a number of activities in which students might engage while trying to learn pronunciations in everyday situations, such as hearing each word pronounced more than once or trying to pronounce the words aloud during learning.

Finally, it is worth exploring whether a mixture of blocking and interleaving is optimal for learning some types of materials. Rather than using a schedule that is exclusively blocked or interleaved, it may be more advantageous to start with a blocked schedule and then transition to interleaving. This possibility has been discussed (see, e.g., Dunlosky et al., 2013; Rohrer, 2012), but has not yet been fully explored in any of the known research on interleaving. Research on interleaving has yet to explore the efficacy of various interleaving schedules, or combinations of blocking and interleaving, that may yield the most optimal learning for a variety of different materials over different time frames.

Author note We thank Annie Morris for her assistance with scoring and Steve Burianek, Courtney Grotenhuis, Syamim Hasim, Lauren Miller, Kellie Mullaney, and Courtney Tapp for their assistance with data collection.

Appendix 1

Table 2 List of French words used in Experiments 1, 2, and 3

Pronunciation rules							
eau	er	e	ou	ch	eux	s	t
bateau	adosser	allumette	bouton	archipel	brumeux	admis	brevet
carreau	attraper	cervelle	genou	capuchon	neveux	brebis	carnet
corbeau	baver	corolle	goudron	chacal	osseux	coloris	pavot
fardeau	combler	cuvette	gourou	chardon	paresseux	compris	rabot
panneau	darder	emplette	mouton	charnel	sablonneux	lambris	sifflet
rameau	raconter	frappe	tabou	chiffon	somptueux	tandis	soldat
tableau	tamiser	malade	verrou	cochon	vaniteux	verglas	sommet
tonneau	valser	serpe	voulu	fichu	venimeux	vernis	tricot

Appendix 2

Table 3 List of French words used in Experiment 4

Pronunciation rules					
eau	er	e	eux	t	
bateau	adosser	allumette	brumeux	brevet	
carreau	attraper	cervelle	neveux	carnet	
corbeau	baver	corolle	osseux	pavot	
fardeau	combler	cuvette	paresseux	rabot	
panneau	darder	emplette	sablonneux	sifflet	
rameau	raconter	frappe	somptueux	soldat	
tableau	tamiser	malade	vaniteux	sommet	
tonneau	valser	serpe	venimeux	tricot	
bandeau	aduler	brosse	caverneux	capot	
drapeau	baser	datte	furieux	pivot	
museau	griffer	flotte	nombreux	sabot	
radeau	tomber	folle	radieux	vert	
cadeau	agrafer	monde	fameux	cornet	
manteau	corriger	pomme	lumineux	rivet	
poteau	isoler	rampe	odieux	sortant	
roseau	vanter	salle	vieux	volant	
cerveau	armer	sorte	frileux	ilot	
marteau	fonder	tasse	mieux	robinet	
pruneau	ramener	tresse	pluvieux	verset	
tombeau	bruler	trompe	yeux	volet	

References

- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *14*, 575–589.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mis-measure of memory: When retrieval fluency is misleading as a metacognitive index. *Journal of Experimental Psychology: General*, *127*, 55–68.
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1547–1552.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1563–1569.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268–276.
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*, 369–378.
- Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 92–101.
- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*, 1095–1102.
- Coppens, L. C., Verkoefen, P. P. J. L., & Rikers, M. J. P. (2011). Learning Adinkra symbols: The effect of testing. *Journal of Cognitive Psychology*, *23*, 351–357.
- de Croock, M. B. M., & van Merriënboer, J. J. G. (2007). Paradoxical effects of information presentation formats and contextual interference on transfer of a complex cognitive skill. *Computers in Human Behavior*, *23*, 1740–1761.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58.
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*, 608–628.
- Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103.
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction?”. *Psychological Science*, *19*, 585–592.
- Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*, 787–794.
- Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology*, *51*, 239–243.
- Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, *35*, 105–123.
- Mitchell, C. J., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 237–242.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*, 237–241.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, *60*, 437–447.
- Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 69–80.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27.
- Roediger, H. L., III, & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.
- Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.
- Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*, 355–367.
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498.
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*, 1209–1224.
- Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgments reflect the belief that diagrams improve learning from text. *Memory*, *18*, 698–711.
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837–848.
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*, 252–257.
- 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*, 750–763.