sempre:
Society for Education, Music and Psychology Research

Article

Psychology of Music
2021, Vol. 49(4) 1027–1046

DOI: 10.1177/0305735620922595
journals.sagepub.com/home/pom
SAGE

Learning melodic musical intervals: To block or to interleave?

Sarah Shi Hui Wong, Si Chen and Stephen Wee Hun Lim

Abstract

Musical interval identification is a valuable skill for holistic and sophisticated musicianship. Yet, discriminating and identifying intervals is often challenging, especially for musical novices. Drawing on cognitive psychological principles, we built two experiments that investigated the utility of interleaving in enhancing novices' aural identification of melodic ascending intervals. Specifically, we designed a novel programmed intervention during which novices learnt six interval types in an interleaved schedule (different interval types learnt interspersed) and six interval types in a blocked schedule (each interval type drilled several times before proceeding to the next) within a single session. When implemented in combination with familiar reference songs and singing as supplementary learning aids, interleaving and blocking yielded comparable performance on a test requiring participants to classify novel instances of the studied interval types (Experiment 1). However, in the absence of reference songs and singing, a robust interleaving effect emerged—interleaving produced superior musical interval identification than blocking (Experiment 2). Yet, most participants were unaware of the benefits of interleaving, and misjudged blocking to be more effective. These findings highlight the potential influence of context under which interleaving is a beneficial technique for learning melodic musical intervals.

Keywords

aural skills, melodic intervals, category induction, interleaving, blocking

Musical intervals constitute basic relations in pitch, which is one of two primary dimensions of music that have been the focus of extensive psychological research and interest, the other dimension being rhythm (Krumhansl, 2000). Intervals can be defined as the fixed distance or frequency ratio between two pitches (Burns, 1999; Krumhansl, 2000), which can be presented either simultaneously (harmonic intervals) or successively (melodic intervals). For instance, a

Department of Psychology, Faculty of Arts & Social Sciences, National University of Singapore, Singapore

Corresponding authors:
Sarah Shi Hui Wong, Department of Psychology, Faculty of Arts & Social Sciences, National University of Singapore, Block AS4, 9 Arts Link, Singapore 117570.
Email: psywshs@nus.edu.sg

Stephen Wee Hun Lim, Department of Psychology, Faculty of Arts & Social Sciences, National University of Singapore, Block AS4, 9 Arts Link, Singapore 117570.
Email: psylimwh@nus.edu.sg

transposed melody of "Twinkle Twinkle Little Star" is likely to remain recognizable even if all of its notes have been shifted up or down, since the intervals—the pattern of changes between notes—have been preserved (McDermott, Lehr, & Oxenham, 2008). Accordingly, melodic intervals serve as the basis for melody, and sequences of such intervals have even been described as the "fingerprint" of music (Thompson, 2013).

Developing an aural awareness and understanding of intervals has been viewed as fundamental to good musicianship (Buttram, 1969), and has been found to be important for more advanced aural skills. For instance, interval identification is closely related to the skill of sight-singing (Ottman, 1956), and has more recently been shown to predict the detection of pitch errors in musical performance (Stambaugh & Nichols, 2020). Unsurprisingly, music schools often spend enormous amounts of time on interval identification as a common ear-training activity (Rogers, 2004), given that listening skills have been considered "prerequisite to all other musical pursuits" (Madsen & Geringer, 2000–2001, p. 103).

The task of aural interval identification is a challenging one, however (Killam, Lorton, & Schubert, 1975). To identify an interval, listeners must not only calibrate the frequency ratio between two heard pitches, but also discriminate this ratio among other ratios (i.e., other intervals), and provide a label for the heard interval (Little, Cheng, & Wright, 2019). Besides this approach, other methods of aurally identifying intervals have been conceptualized. For example, it may be possible for listeners who are fluent with functional scale degrees to meaningfully encode the pitches heard within a tonal context in relation to the appropriate scale degrees, then identify the interval based on one's higher order knowledge of the pitches' tonal functions (Karpinski, 2000; Wright, 2016). However, attaining competence in this aural skill is considerably difficult— even music conservatory students have been observed to struggle with this complex task (Ponsatí, Miranda, Amador, & Godall, 2016, 2020), while novices use relatively less efficient strategies and consequently perform more poorly at interval perception and identification (Burns & Ward, 1978; Siegel & Siegel, 1977; Zatorre & Halpern, 1979). For intervals up to an octave, for instance, musically untrained listeners demonstrate poorer differentiation of interval sizes than trained listeners, whose magnitude estimates of each interval's size reflect a more rapid increase with greater interval size (Russo & Thompson, 2005). Nonetheless, there is some evidence that novices can demonstrate increased sensitivity to unfamiliar intervals (Leung & Dean, 2018) and even make finer-grained distinctions in identifying specific intervals (Little et al., 2019) after receiving training (for a review of non-musicians as "experienced listeners," see Bigand & Poulin-Charronnat, 2006).

How can novices' aural identification of musical intervals be facilitated? Drawing on cognitive psychology research (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kornell & Bjork, 2008; Sana, Yan, & Kim, 2017), we compared two strategies to enhance novices' learning of melodic intervals: *blocking* versus *interleaving*. Whereas blocking involves repeatedly studying exemplars from the same category (e.g., the same interval type) before going on to the next category (i.e., AAABBBCCC), interleaving involves studying exemplars from different categories (e.g., different interval types) in an interspersed manner (i.e., ABCABCABC).

The interleaving effect

Traditionally, blocking or repetition has often been recommended by music educators (e.g., Williamon, 2004; see Stambaugh, 2011 for a discussion), and is a popular practice technique that is widely adopted by musicians (e.g., Austin & Berg, 2006; Barry, 1992; Leon-Guerrero, 2008; Maynard, 2006; Rohwer & Polk, 2006). Such preferences for blocking may stem from a priori beliefs or intuitions, or a sense of subjective fluency that blocking induces, whereby repeated exposure to the same to-be-learnt category may increase one's experienced ease of processing that subsequently shapes perceptions of blocking's effectiveness (Yan, Bjork, & Bjork, 2016). However, numerous studies in cognitive and educational psychology have reported an *interleaving effect*, whereby interleaving produces superior learning than blocking across domains such as musical performance (Abushanab & Bishara, 2013; Carter & Grahn, 2016; Stambaugh, 2011), motor skills (Shea & Morgan, 1979), mathematics (Rohrer, Dedrick, & Burgess, 2014; Taylor & Rohrer, 2010), and foreign language acquisition (Nakata & Suzuki, 2019; Pan, Tajran, Lovelett, Osuna, & Rickard, 2019; cf. Carpenter & Mueller, 2013).

The advantage of interleaving over blocking has also been related to the "contextual interference effect" observed in verbal learning (Battig, 1966) and motor skill acquisition (Carter & Grahn, 2016; Magill & Hall, 1990; Shea & Morgan, 1979), whereby interference during practice counterintuitively facilitates subsequent skill learning. Under this view, interleaving introduces more contextual interference than blocking because learners must alternate between studying different variations or categories rather than the same ones repeatedly, thus encouraging greater elaborative processing that leads to improved retention (for a review, see Magill & Hall, 1990). For such reasons, interleaving has also been considered a "desirable difficulty" (Bjork, 1994), in that it may appear to cause immediate difficulties for learners, but in fact enhances long-term performance.

Of particular interest, the interleaving effect has been observed in *category induction*—learning the diagnostic features of a category through exposure to its exemplars (for a review, see Rohrer, 2012). A meta-analysis of the interleaving effect based on 59 studies (Brunmair & Richter, 2019) has revealed a moderate positive effect of interleaving over blocking in inductive learning, particularly for studies using visual materials such as paintings by various artists for discriminating between their painting styles (Kornell & Bjork, 2008), naturalistic pictures for distinguishing species of birds and butterflies (Birnbaum et al., 2013; Wahlheim, Dunlosky, & Jacoby, 2011), and diagrams to categorize organic chemical compounds (Eglington & Kang, 2017).

Comparatively few studies have investigated the interleaving effect in learning auditory material or categories, with some exceptions. For instance, an interleaving advantage has been observed for learning categories of psychopathological disorders from auditorily presented case studies (Zulkiply, McLean, Burt, & Bath, 2012). More recently, interleaving has also been found to outperform blocking in musical category learning. In a study by Wong, Low, Kang, and Lim (2020), novices learnt various classical music composers' styles by listening to their respective music pieces in a blocked (i.e., listening to works by one composer at a time before going on to the next) or interleaved (i.e., alternating between listening to different composers' works) manner. In a final test, learners were more accurate at classifying novel music pieces by composers whose works had earlier been presented interleaved, as opposed to blocked. Yet, most learners were unaware of this benefit and instead misjudged blocking to be more effective.

There are at least two mechanisms that can potentially explain the interleaving effect in category induction. First, interleaving involves spacing exemplars from the same category apart, as opposed to presenting them back-to-back during blocking. Such spacing in interleaving facilitates more effortful retrieval and mental processing during each reoccurrence of a category's exemplars (Birnbaum et al., 2013; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006) while sustaining learners' attention and reducing mind-wandering (Metcalfe & Xu, 2016), thereby promoting learning. However, spacing alone is inadequate to account for the benefits of interleaving (Kang & Pashler, 2012). Rather, a second mechanism relates to the discriminative contrast that interleaving promotes—when exemplars from different categories are interleaved, learners' attention is drawn to the key differences across categories and they are thus better able to discriminate among these categories, relative to blocking that directs learners' attention toward detecting similarities among exemplars within the same category (Birnbaum et al., 2013; Carvalho & Goldstone, 2017; Goldstone & Steyvers, 2001; Kang & Pashler, 2012). In line with this view, the interleaving effect has been found to be stronger for learning material with high inter-category similarity, which thus invokes greater difficulty in discriminating among the different categories (Brunmair & Richter, 2019). Accordingly, given that musical interval learning involves discriminating among various categories or types of intervals, interleaving may be a relatively more effective strategy than blocking. Curiously, though, interleaving has received relatively little attention in extant interval learning research and classroom practices.

Extant research and practices in musical interval learning

Blocking. In Rogers' (2004) overview of pedagogical approaches in music theory, one practical recommendation for interval practice involves blocking intervals by particular families or classes, such that intervals with similar characteristics (e.g., stable vs. unstable intervals) are grouped during study. Indeed, in music classrooms, blocking is evident in the common practice of drilling easier or more familiar intervals before proceeding to more difficult or unusual ones that are reserved for last (Jeffries, 1967; Rogers, 2004). Notably, this procedure of blocking intervals in increasing order of difficulty has been adopted in taped self-instructional melodic interval drills that have been developed for use with music students at various universities (Spohn, 1963; Tarratus & Spohn, 1967).

At the same time, blocking is the predominant schedule used in most research studies on musical interval learning, in contrast to cognitive psychology research on the interleaving effect. For instance, in Wasserman's (1974) 8-week programmed melodic interval learning procedure that involved self-learning via tapes, high-school students in a vocal program were drilled on a specific interval type until they could correctly identify it, before proceeding to learn the next interval type. Even while some studies have incorporated elements of interleaving, blocking still tends to take precedence. For instance, in Jeffries' (1967) programmed melodic interval training, each learning session comprised introducing students to two new intervals and repeatedly drilling each of these intervals in separate blocks, although all learnt intervals were subsequently presented in random order for review. Likewise, Smith, Kemler Nelson, Grohskopf, and Appleton (1994) adopted blocking as the primary schedule when teaching novices to identify three melodic interval types by associating them with the opening notes of three familiar folk tunes—before learning the next interval type, each interval type was drilled multiple times on different registers across the keyboard in various forms (e.g., listening to the interval alone, listening to the corresponding folk tune, and imagining hearing the remaining melody after listening to the interval). Only after each interval type had been drilled were novices presented with a random assortment of nine instances of the three interval types.

Reference songs. Beyond blocking or interleaving the musical intervals to be learnt, it is worth further noting that using familiar folk tunes as reference songs during the aural training process benefited novices' interval identification in Smith et al.'s (1994) study. Specifically, novices who had learnt the intervals with the aid of reference songs outperformed those who had received standard instruction without the familiar melodies. Some novices even showed near expert-caliber discrimination performance on an interval identification test after undergoing training with the reference songs, in line with extant views on non-musicians as "experienced listeners" (Bigand & Poulin-Charronnat, 2006). Indeed, the advantage of musical training for melody recognition has been found to be less pronounced for familiar than unfamiliar melodies

(Dalla Bella, Peretz, & Aronoff, 2003), implying that non-musicians can potentially be as efficient as musicians in recognizing familiar tunes, which may thus serve as useful aids. That is, familiar musical tokens can be helpful in guiding novices to make connections between novel musical events such as intervals with their existing schemas in long-term memory (Smith et al., 1994), although some music educators have cautioned against over-reliance on such tunes (Karpinski, 2000; Rogers, 2004).

Singing. Some studies have combined the use of reference songs with singing in musical interval learning. For instance, in recent observational studies of the Aural Identification of Musical Harmonic Intervals proposal developed by Ponsatí et al. (2016, 2020), music conservatory students learnt musical intervals over 20 weeks in a four-stage process: (1) memorized singing of reference songs, (2) singing the first fragment of the songs, (3) singing the first interval of the songs aloud, and (4) internal singing of the first interval of the songs, while practicing aural discrimination and identification of the intervals throughout the four stages.

Singing has been regarded as a quintessential aspect of aural skill development (Klonoski, 1998) and as being crucial for audiation—the ability to hear music in one's mind even in the absence of physical sound (Gordon, 1985). In turn, audiation has been proposed to play an important role when hearing and distinguishing intervals through inner-ear recognition (Boberg, 1975), and when mentally manipulating intervals to experience their distances (Garner, 2009). Consonant with theories of cognitive processing, singing may also be helpful for interval learning because it engages multiple modalities when hearing and verbally producing melodic intervals, which may then facilitate multisensory representations of these intervals that benefit learning (see Shams & Seitz, 2008, for a review of the benefits of multisensory learning).

Taken together, these studies suggest that in the broader context of aural training when learning to identify musical intervals, a number of techniques have been shown as effective and can potentially be used in combination. Accordingly, blocking or interleaving does not necessarily have to be applied alone, but can be integrated with other effective elements such as reference songs and singing to form a holistic approach toward enhancing musical interval learning.

The present research

Developing novices' ability to accurately discriminate and identify various musical intervals serves as a step in guiding them toward sophisticated musicianship. To this end, drawing on learning principles that are grounded in cognitive psychology (see, e.g., Wong & Lim, 2017, for a discussion), such as the use of interleaving, may be particularly useful in informing educational practice. Whereas cognitive psychology studies have often endorsed the use of interleaving, this learning technique has arguably received less attention than blocking in extant research and practices in musical interval learning. To address this critical gap given its pedagogical implications and relevance, we empirically tested the efficacy of interleaving versus blocking in novices' inductive learning of musical intervals.

As identifying harmonic intervals has been theorized to involve isolating the relevant pitches and thereby translating them into a melodic type (e.g., Rogers, 2004; Zatorre & Halpern, 1979), we focused on the basic aural skill of identifying melodic intervals, particularly the 12 common interval types in Western chromatic music: Minor 2nd (m2), Major 2nd (M2), Minor 3rd (m3), Major 3rd (M3), Perfect 4th (P4), Tritone (TT), Perfect 5th (P5), Minor 6th (m6), Major 6th (M6), Minor 7th (m7), Major 7th (M7), and Perfect 8^{ve} (P8). Furthermore, as most studies on musical interval learning have involved extended training periods and/or focused only on some of the 12 interval types (e.g., Jeffries, 1967, 1970; Little et al., 2019), we sought to design a learning procedure that would enable the effective *and* efficient learning of all 12 interval types within a single session.

Experiment 1

To contextualize our study in the landscape of current musical interval learning practices, we developed a programmed learning intervention that integrated interleaving versus blocking with the use of reference songs and singing. Whereas Ponsatí et al.'s (2016, 2020) Aural Identification of Musical Harmonic Intervals procedure targeted professional-grade music students and their aural identification of harmonic intervals, we positioned our melodic interval learning intervention at a level that was accessible even to musical novices. We aimed to test the interleaving effect in combination with the use of reference songs and singing on learning interval types and categorizing novel instances of the studied interval types on a subsequent test.

Method

Participants. Forty-eight students (30 were female) aged between 19 and 25 ($M = 21.24$, $SD = 1.74$) from the National University of Singapore took part in the study. Participants received either course credit or cash reimbursement. All participants were novices with no self-reported prior knowledge of musical intervals, and had never formally undergone musical interval training. Participants reported an average of 2.45 years ($SD = 1.80$) of formal musical training on an instrument (including voice). All experiments reported here were conducted with the appropriate ethics-review-board approval by the National University of Singapore, and participants granted their written informed consent.

Design. Experiment 1 employed a fully within-subjects design. The primary independent variable of interest was *Learning Strategy*: interleaved (each trial in a study block was of a different interval type) versus blocked (all trials in a study block were of the same interval type). We included *Test Block* as a secondary independent variable for control purposes (i.e., to determine whether the observed effects, if any, persisted across all four test blocks). The dependent variable was *Test Performance*, as measured by the proportion of novel intervals that participants correctly identified in a test.

Materials

Interval stimuli. All 12 musical interval types in Western music were included in the experiment: m2, M2, m3, M3, P4, TT, P5, m6, M6, m7, M7, and P8. Each interval type was formed from 12 base pitches of the Western chromatic scale: C, D♭, D, E♭, E, F, F♯, G, A♭, A, B♭, and B. Accordingly, 12 interval stimuli of each interval type were generated for a total of 144 interval stimuli. All interval stimuli were melodic ascending intervals. This procedure ensured that participants were exposed to all intervals formed by the 12 base pitches, thus maximizing the comprehensiveness of their learning and the generalizability of any observed effects. To facilitate participants' ease of singing during the learning process, we used the interval range of A3–A♭5. In addition, each interval stimulus was presented with its notes repeated as a four-note sequence (e.g., C-C-G-G). Compared to presenting intervals as two-note events (e.g., C-G), such four-note sequences have been found to improve novices' encoding of relative pitch distance (Lee, Janata, Frost, Martinez, & Granger, 2015).

Each of the 144 interval stimuli was entered into a music notation software (Sibelius 7), and 5-s piano sound recordings were generated via the software's playback function. In each audio recording, the notes of each interval stimulus were crotchets played at a tempo of 50 beats per min (BPM). We determined this tempo following a pilot test among novices who reported that 50 BPM was most suitable for their learning, compared to faster or slower tempi.

For each of the 12 interval types, six interval stimuli of different base pitches were randomly designated as the study stimuli to be presented during the study phase (with each participant randomly receiving three out of the six interval stimuli; i.e., 36 study items in total), while the remaining six interval stimuli were presented as test items during the test phase (with each participant randomly receiving four out of the six interval stimuli; i.e., 48 test items in total).

Reference songs. In all, 12 reference songs corresponding to each of the 12 interval types were selected. These songs were reported to be of high familiarity to novice participants in our pilot test. For example, the reference songs for the P4 and P5 interval types were "Amazing Grace" and "Twinkle Twinkle Little Star", respectively. The full list of reference songs is available in the supplemental materials online.

We generated 30-s audio clips of the 12 reference songs, which were presented to participants for initial exposure and familiarization. In addition, a brief excerpt from the opening of each of the 12 reference songs containing their respective interval types was selected, input into Sibelius 7 on 12 base pitches, and subsequently exported for a total of 144 piano audio clips that were presented to participants during the study phase of the experiment. Similar to the interval stimuli, the two notes forming the opening interval of all reference song excerpts fell within the range of A3–A♭5.

Goldsmiths Musical Sophistication Index. We administered the Goldsmiths Musical Sophistication Index (Gold-MSI; version 1.0) to assess participants' musical sophistication as a potential predictor of their inductive learning performance. A 38-item self-report inventory, the Gold-MSI measures musical sophistication on several dimensions of musical skills and behaviors in the general population (Müllensiefen, Gingras, Musil, & Stewart, 2014), and has been validated across several studies (Baker, Ventura, Calamia, Shanahan, & Elliott, 2018; Degrave & Dedonder, 2019; Lima, Correia, Müllensiefen, & Castro, 2020; Lin, Kopiez, Müllensiefen, & Wolf, 2019). The Gold-MSI includes five subscales: *active musical engagement* (9 items; $\alpha = .80$), *perceptual abilities* (9 items; $\alpha = .80$), *musical training* (7 items; $\alpha = .77$), *emotional responses to music* (6 items; $\alpha = .62$), and *singing abilities* (7 items; $\alpha = .83$). A composite score of General Musical Sophistication was also derived from participants' scores on 18 items ($\alpha = .89$) of the Gold-MSI (e.g., Müllensiefen et al., 2014). All Cronbach's alphas reported here were computed from our study's data. Each item was rated on a 7-point Likert-type scale, with negative items reverse-coded such that higher scores indicated higher levels of musical sophistication.

Procedure The experiment was programmed and presented to participants via E-Prime 2.0, and comprised three phases: initial exposure, study, and test.

Initial exposure phase. Upon their arrival at the laboratory either individually or in groups of no more than four, participants were first introduced to the basic definition of a musical interval, the 12 types of musical intervals, as well as the 12 reference songs that corresponded to each interval type. A printed information sheet that summarized this information was also handed out to participants to facilitate their learning (available in the supplemental materials online). Participants were asked to familiarize themselves with audio clips of the 12 reference songs, each lasting 30 s. As participants listened to each song via headphones, the song's title and respective interval type were presented at the bottom of the screen. Participants were taught to associate each interval type with the relevant opening notes of its corresponding reference song (e.g., Smith et al., 1994), with this information also included in the printed handout that they received. The presentation order of the reference songs was randomized for each participant. After each reference song had been played, participants were asked to rate how familiar they were with the song on a 7-point scale (1 = *not at all*, 7 = *very much*). The initial exposure phase lasted approximately 5 min.

Study phase. The interval stimuli were presented in nine study blocks ordered as BIBBIBBIB. Specifically, "B" denotes blocked learning whereby all trials within a study block were of the same interval type but formed on different base pitches (e.g., P5 formed on C, P5 formed on F), whereas "I" denotes interleaved learning whereby each trial within a study block was of a different interval type but formed on the same base pitch (e.g., M2 formed on C, m6 formed on C). Participants studied three trials in each "B" block and six trials in each "I" block, thus completing 18 study trials per learning strategy (i.e., 36 study trials in total). Within each of the nine study blocks, the presentation order of the interval stimuli was randomized for each participant. The study phase lasted approximately 15 min.

Six interval types were learnt interleaved, while the remaining six were learnt blocked. Participants were randomly assigned to receive one of two counterbalanced versions of the study stimuli, whereby the six interval types that had been presented in blocked format in one version (m2, M3, TT, P5, M6, and m7) were presented as interleaved in the other, and vice versa for the remaining six interval types (M2, m3, P4, m6, M7, and P8). A sample study sequence is listed in the supplemental materials online. We conceived the grouping of the interval types to reduce repetitions of interval numbers within each condition (e.g., to avoid Minor 3rd and Major 3rd being presented in the same condition), as well as to ensure comparable difficulty of intervals across conditions (e.g., Ponsatí et al., 2016). Thus, any observed effects could be attributed solely to the learning strategy adopted, rather than varied difficulty in identifying particular interval types.

Participants completed 36 study trials spanning across the nine study blocks and encompassing all 12 interval types (i.e., three study trials per interval type). Each of the 36 study trials proceeded in the following four stages: (1) listening to the reference song excerpt containing the to-be-learnt interval, with the song title and its corresponding interval type presented at the bottom of the screen, (2) singing the reference song excerpt aloud at one's own pace, (3) listening to an audio recording of the corresponding interval stimulus with the same base pitch as the reference song, with the name of the interval type presented at the bottom of the screen, and (4) singing the interval stimulus aloud at one's own pace. In this manner, participants learnt the auditory characteristics and names of all 12 interval types, alongside the use of reference songs and singing as supplementary aids.

Test phase. At the end of the study phase, participants engaged in a brief filler task, in which they counted down in 3 s from 547 for a total of 15 s (e.g., Kornell & Bjork, 2008). Then, participants underwent the test phase, which comprised four blocks of 12 test trials each (i.e., 48 test trials in total). Within a test block, each of the 12 trials were of a different interval type, with a randomized order of presentation for each participant. On each test trial, an audio

Table 1. Test performance in interleaved and blocked conditions by test block (Experiment 1).

Test block	Interleaved		Blocked	
	M	*SD*	*M*	*SD*
Block 1	.26	.20	.26	.20
Block 2	.29	.22	.27	.19
Block 3	.27	.21	.31	.21
Block 4	.27	.21	.26	.18

SD: standard deviation.
N = 44. Participants' test scores were computed as the proportion of correctly identified novel musical intervals.

clip of a novel interval stimulus (i.e., not presented during the study phase; formed on a base pitch that was different from that of the study stimuli for that interval type) from one of the 12 interval types was played. Participants then responded to the question, "What is the name of the interval you just heard?" by typing a number from 1 to 12, whereby each number corresponded to the name of an interval type as presented onscreen. After undergoing all 48 test trials, participants completed the Gold-MSI, before being debriefed and thanked. The test phase lasted approximately 20 min.

Results

Participants' test performance was scored as the proportion of novel musical intervals that they correctly identified in the blocked versus interleaved condition. Four out of the 48 (8%) participants who performed at or below chance level (i.e., an accuracy level of 1 out of 12 trials answered correctly) across both conditions on the test were excluded from all analyses, leaving a final sample of 44.

Participants' test scores were submitted to a 2 (*Learning Strategy*: interleaved vs. blocked) × 4 (*Test Block*) repeated measures analysis of variance (ANOVA). There was no main effect of Learning Strategy, $F(1, 43) = 0.01$, $MSe = 0.03$, $p = .92$, $\eta_p^2 < .001$. Participants' test performance did not differ across the interleaved ($M = .27$, $SD = .15$) and blocked conditions ($M = .28$, $SD = .14$). The Learning Strategy × Test Block interaction was also nonsignificant, $F(3, 129) = 0.71$, $MSe = 0.03$, $p = .55$, $\eta_p^2 = .02$, as was the main effect of Test Block, $F(3, 129) = 0.80$, $MSe = 0.03$, $p = .50$, $\eta_p^2 = .02$. Table 1 shows the means and standard deviations of participants' test scores across the four test blocks.

Analyzing participants' test performance for each of the 12 interval types across the interleaving and blocking conditions (Figure 1), we found that the m6, M6, and M7 intervals ranked among the most difficult to identify on overall. These results corroborate those of previous studies that have also reported poorer identification performance for these intervals (e.g., Jeffries, 1967; Killam et al., 1975; Ponsatí et al., 2016, 2020; Samplaski, 2005).

Participants' self-reported familiarity with the reference songs did not significantly correlate with their overall test performance ($r = .12$, $p = .43$). However, participants' overall test performance significantly and positively correlated with their General Musical Sophistication ($r = .40$, $p = .008$), Perceptual Abilities ($r = .39$, $p = .01$), and Musical Training ($r = .48$, $p = .001$) scores on the Gold-MSI. No other correlations with the Gold-MSI subscales were significant. Means and standard deviations of participants' scores on the Gold-MSI are available in the supplemental materials online.

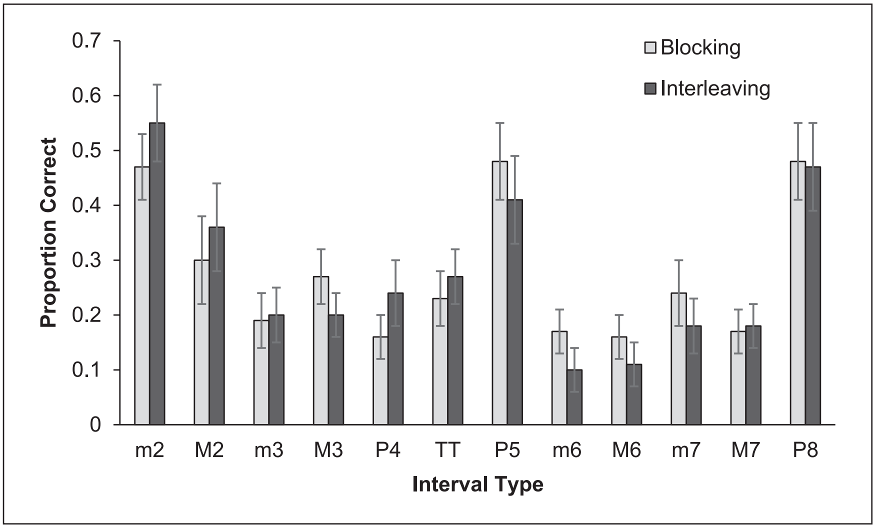

Figure 1. Learners' test performance for each interval type by learning condition in Experiment 1. Error bars represent standard errors.

Discussion

In Experiment 1, we pitted the practice of blocking against interleaving that has been demonstrated to produce superior inductive learning in cognitive psychology research (e.g., Birnbaum et al., 2013; Kornell & Bjork, 2008). Surprisingly, inconsistent with the interleaving effect, we found that blocking yielded comparable melodic interval identification performance as interleaving in the context of a novel learning paradigm incorporating reference songs and singing as elements that commonly accompany instruction in real-world music educational settings.

Here, we consider a couple of explanations related to the inclusion of reference songs and singing. First, experimental studies demonstrating an advantage of interleaving in musical learning have often investigated this effect without integrating other learning aids or techniques (e.g., Stambaugh, 2011; Wong et al., 2020). In contrast, learners' attention in the present study may have been diverted to the reference songs and singing elements. Consequently, learners may have relied predominantly on the reference songs to build associations with their respective interval types, instead of capitalizing on the discriminative contrast that interleaving affords to appreciate the differences across interval types.

Second, the inclusion of reference songs and singing may have inherently introduced elements of blocking to the learning procedure even in the interleaved condition. Specifically, on each trial across both the interleaved and blocked conditions, learners were exposed to the to-be-learnt interval four times: (1) listening to the reference song containing the interval, (2) singing the reference song containing the interval, (3) listening to the specific interval, and (4) singing the specific interval. In the blocked condition, a sample learning schedule was thus as follows, where "A" denotes a particular interval type formed on various base pitches: $A_1A_1A_1A_1A_2A_2A_2A_2A_3A_3A_3A_3$. Correspondingly, in the interleaved condition, the learning schedule resembled the following, where "A," "B," and "C" denote different interval types formed on the same base pitch: $A_1A_1A_1A_1B_1B_1B_1B_1C_1C_1C_1C_1$ and so on. This could have been problematic because each trial was then essentially learnt "blocked", even if it had been presented in an overall interleaved schedule. In other words, the inclusion of reference songs and singing in Experiment 1 may have inadvertently interfered with the interleaving effect, subjecting the learning trials in the interleaved condition to basically a blocked schedule. In addressing these issues, we designed Experiment 2.

Experiment 2

In Experiment 2, we distilled the contributions of interleaving versus blocking to novices' musical interval learning. To this end, Experiment 2 was identical to Experiment 1 but without the use of reference songs and singing. That is, the learning trials' overall interleaved schedule was preserved through having participants learn the various melodic interval types solely by listening to examples of each type. In line with previous work on the interleaving effect, we predicted that interleaving alone would lead to superior performance in classifying novel instances of the studied interval types, compared to blocking. In addition, we investigated learners' metacognitive judgments of their learning, as extant research has found that learners are often unaware of the benefits of interleaving, and instead misperceive blocking to be more effective even if their actual performance indicated otherwise (e.g., Birnbaum et al., 2013; Kornell & Bjork, 2008).

Method

Participants. The participants were 49 students (29 were female) aged between 19 and 25 ($M = 20.9$, $SD = 1.94$) from the National University of Singapore. As in Experiment 1, all participants were musical novices with no self-reported prior knowledge of musical intervals, and had never formally undergone musical interval training. Participants reported an average of 1.82 years ($SD = 1.40$) of formal musical training on an instrument (including voice). Participants received either course credit or cash reimbursement for their participation.

Design. A 2 (*Learning Strategy*: interleaved vs. blocked) × 4 (*Test Block*) fully within-subjects design the same as that in Experiment 1 was used. The dependent variable was participants' *Test Performance*, as assessed by the proportion of novel musical intervals that they correctly identified during the test phase.

Materials. The same interval stimuli from Experiment 1 were used. For each of the 12 interval types, 6 interval stimuli of different base pitches were randomly selected for the study phase (i.e., 72 study items), while the remaining 6 interval stimuli were presented during the test phase (with each participant randomly receiving 4 out of 6 interval stimuli; i.e., 48 test items in total).

Procedure. The procedure of Experiment 2 was identical to that in Experiment 1, with two modifications. First, during the study phase, participants learnt the interval stimuli solely through listening to them (i.e., without the reference songs and singing). On each study trial, an audio recording of the to-be-learnt interval stimulus was presented, with the name of its interval type displayed at the bottom of the screen. Participants were tasked to learn the auditory characteristics and names of the 12 interval types. The interval stimuli were presented in 12 study blocks of six trials each (i.e., 72 study trials in total, with six trials for each of the 12 interval types), in the overall study block format of BIIBBIIBBIIB where "B" denotes blocked learning and "I" denotes interleaved learning. Within each study block, the presentation order of the interval stimuli was randomized for each participant.

Table 2. Test performance in interleaved and blocked conditions by test block (Experiment 2).

Test block	Interleaved		Blocked	
	M	*SD*	*M*	*SD*
Block 1	.30	.22	.23	.18
Block 2	.36	.21	.22	.16
Block 3	.29	.23	.28	.22
Block 4	.31	.20	.24	.23

SD: standard deviation.
$N = 44$. Participants' test scores were computed as the proportion of correctly identified novel intervals.

Second, at the end of the test phase, participants additionally made a metacognitive judgment of their learning. Specifically, the differences between "blocked" and "interleaved" were reiterated to participants, who were then asked to indicate which strategy they thought had been more effective in helping them learn for the test. Participants were provided with three options to indicate their responses: *blocked, about the same*, or *interleaved*.

The initial exposure, study, and test phases lasted approximately 10, 30, and 20 min, respectively.

Results

As in Experiment 1, participants' test performance was scored as the proportion of novel intervals that they correctly identified in the blocked versus interleaved condition. Four participants who performed at or below chance level across both conditions on the test, as well as one participant whose test scores fell more than three standard deviations away from the mean across both conditions (i.e., an extreme outlier), were excluded from all analyses, leaving a final sample of 44.

The data were submitted to a 2×4 repeated measures ANOVA with *Learning Strategy* (interleaved or blocked) as the first independent variable, *Test Block* as the second independent variable for control purposes, and *Test Performance* as the dependent variable. There was a significant main effect of Learning Strategy, whereby interleaving ($M = .32$, $SD = .12$) produced significantly better test performance than blocking ($M = .24$, $SD = .14$), $F(1, 43) = 12.40$, $MSe = 0.04$, $p = .001$, $\eta_p^2 = .22$. There was neither a significant interaction between Learning Strategy and Test Block, $F(3, 129) = 1.62$, $MSe = 0.04$, $p = .19$, $\eta_p^2 = .04$, nor a significant main effect of Test Block, $F(3, 129) = 0.31$, $MSe = 0.03$, $p = .82$, $\eta_p^2 = .01$, suggesting that the interleaving effect and participants' overall test performance did not statistically differ across test blocks. The means and standard deviations of participants' test scores across the four test blocks appear in Table 2.

Analyzing participants' test performance for each of the 12 interval types across the interleaving and blocking conditions (Figure 2), we found that TT, M6, and m7 ranked amongst the most difficult intervals to identify on overall. Like our earlier observations in Experiment 1, this pattern of results largely overlaps with those in previous research on musical interval learning (e.g., Jeffries, 1967; Killam et al., 1975; Ponsatí et al., 2016, 2020; Samplaski, 2005).

The majority of the participants made inaccurate metacognitive judgments of each strategy's effectiveness. Overall, 73% (32 out of 44) of participants had better or comparable performance in the interleaved than blocked condition, but 75% (33 out of 44) of participants thought that blocking was as good as or better than interleaving (see Figure 3).

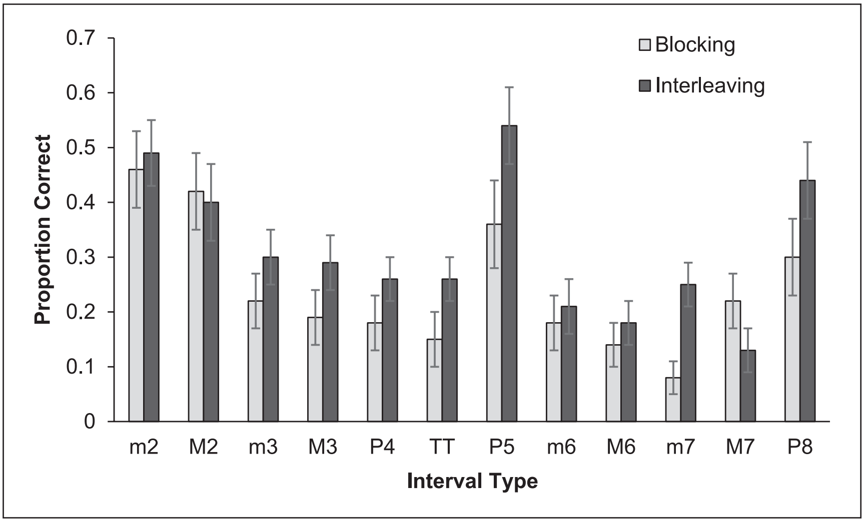

Figure 2. Learners' test performance for each interval type by learning condition in Experiment 2.

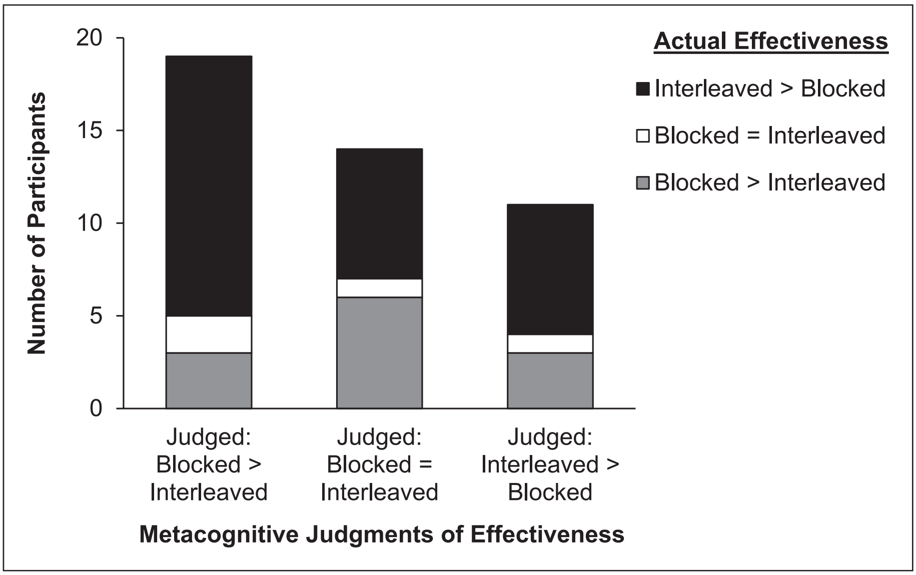

Figure 3. Learners' metacognitive judgments versus actual test performance in Experiment 2.

Unlike in Experiment 1, participants' test scores did not significantly correlate with their scores on the Gold-MSI, including the General Musical Sophistication factor and the five subscales, all $ps > .05$. Means and standard deviations of participants' scores on the Gold-MSI are available in the supplemental materials online. Considered in tandem with Experiment 1's findings, it appears that some elements of musical sophistication may be associated with musical interval learning within the context of additional elements such as reference songs and singing, but not when these elements are removed. Presumably, this may be because interleaving is a general cognitive-based learning strategy and does not predicate on learners' musical training or sophistication (see also Wong et al., 2020).

Discussion

Replicating the interleaving effect (e.g., Birnbaum et al., 2013; Kornell & Bjork, 2008; Taylor & Rohrer, 2010), novices were significantly better at identifying novel musical intervals that had been learnt interleaved rather than blocked. In the absence of reference songs and singing as supplementary aids during the aural training process, interleaving outperformed blocking in musical category induction.

Interestingly, however, most participants mistakenly believed that blocking had been equally or more useful for their learning. These inaccurate metacognitive judgments echo those observed in previous research (e.g., Birnbaum et al., 2013; Kornell & Bjork, 2008; Zulkiply et al., 2012) and can potentially be attributed to a sense of subjective fluency that learners developed during blocking, whereby repeated exposure to the same interval type created an illusion of successful learning (Yan et al., 2016).

General discussion

Across two experiments, we investigated the effects of interleaved versus blocked learning on musical novices' category induction of 12 melodic ascending interval types. When combined with reference songs and singing, blocking and interleaving produced comparable performance on a test requiring learners to classify novel instances of the studied interval types (Experiment 1). However, when the effect of presentation schedule was investigated alone in the absence of reference songs and singing, interleaving yielded superior interval identification performance than blocking (Experiment 2). Yet, learners were largely unaware of the advantages of interleaving, and instead inaccurately judged blocking to be equally or more useful.

Whereas previous studies on the interleaving effect in category induction have focused primarily on the visual domain (e.g., Birnbaum et al., 2013; Kornell & Bjork, 2008), our findings contribute to an emerging area of research that has further demonstrated the benefits of interleaving in the auditory domain (e.g., Wong et al., 2020; Zulkiply et al., 2012). While cognitive psychology studies have often endorsed interleaving (for a review, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013), this learning technique has received comparatively less attention in extant research and classroom practices in musical interval learning (see the Introduction). Our study addressed this critical gap by examining the specific contexts in which blocking versus interleaving may enhance musical interval learning. Specifically, we found that blocking was potentially as effective as interleaving when bolstered with other techniques that are often delivered in combination in real-world educational settings (e.g., reference songs and singing; Ponsatí et al., 2016, 2020). However, when learners could not rely on supplementary aids such as reference songs and singing but were compelled to learn the various interval types in isolation as a function of their presentation schedule, the robust interleaving effect typically observed in empirical literature was replicated. Altogether, the extent to which interleaving benefits musical interval learning may vary depending on the specific learning context in question.

Educational implications

Our findings highlight the effectiveness of adopting an interleaved schedule during aural training when musical novices learn melodic intervals by listening to exemplars of the interval types. As a learning technique that is relatively easy and inexpensive to implement, interleaving can be readily adopted in music classrooms when introducing students to various melodic intervals. Instead of repeatedly drilling students on a particular interval before going on to drill the next one (i.e., blocking), teachers can expose students to examples of various intervals in intermixed succession. In view of our finding that certain intervals—TT, m6, M6, m7, and M7—tend to be more difficult for novices to identify (see also Jeffries, 1967; Killam et al., 1975; Ponsatí et al., 2016, 2020; Samplaski, 2005), teachers may also consider giving these intervals special care. In this manner, students can be guided to better appreciate and learn the differences among the intervals by capitalizing on the discriminative contrast that interleaving affords (Birnbaum et al., 2013; Carvalho & Goldstone, 2017; Goldstone & Steyvers, 2001; Kang & Pashler, 2012).

In particular, we developed and found support for a single-session programmed interleaving intervention that trained musical novices on all 12 common melodic ascending interval types, going beyond previous studies that have trained novices to learn only some of the 12 intervals over extended periods of time (e.g., Little et al., 2019; Smith et al., 1994). This highlights the potential of the present musical interval training program as both an effective and efficient learning tool. For instance, mobile applications such as *Auralbook* enable students to use their mobile devices to develop their aural skills through singing, clapping rhythms, and identifying musical stylistic features, with feedback provided by the applications (Chen, 2015). Likewise, the interval learning procedure in our study can be readily adapted into a software or integrated with current mobile applications that students can access at their own pace and convenience to support their learning. Music educators and students may also profit from using commercially available aural training applications that incorporate elements of interleaving in their music interval recognition and identification programs.

In addition, music educators could explicitly guide their students to adopt interleaving when learning to aurally identify melodic intervals, depending on the learning context at hand. As is evident from participants' inaccurate metacognitive judgments in Experiment 2, novices may not always be aware of the strategies that are most useful for their learning. Moreover, such metacognitive illusions have proven to be particularly persistent and difficult to dispel due to learners' erroneous a priori beliefs and a misleading sense of subjective fluency that blocking encourages (Yan et al., 2016). As such, it is essential that music educators teach their students how and when to apply promising techniques informed by the cognitive science of learning, such as interleaved practice.

Future directions

Extant research has suggested that novices are capable of recognizing and identifying musical intervals after receiving training (Little et al., 2019). Extending this line of work, we found that, relative to blocking, interleaving enhanced novices' ability to discriminate among various melodic interval types and even explicitly label them after solely listening to exemplars of each interval type during training. To further test the parameters of the interleaving effect observed in the present study, future work can explore the effects of presentation schedule on novices' aural identification of harmonic intervals, which is a relatively more challenging task than melodic interval identification (e.g., Killam et al., 1975; Samplaski, 2005).

In addition, while intervals form basic pitch relations in music, it is important to note that they take on meaning in musical environments, such that the same interval may serve varying scale-degree functions in different contexts and thus be perceived differently (Karpinski, 2000; Rogers, 2004). That is, listeners' experience of intervals is shaped by the broader tonal context in which they occur (Bruner, 1984). For instance, musicians have been found to be more accurate in interval discrimination and labeling when these intervals are presented in a melodic context than in isolation (Wapnick, Bourassa, & Sampson, 1982). Accordingly, it will be valuable for future research to investigate the extent that the interleaving advantage for novices in our study transfers from intervals that are presented acontextually to those situated in larger-scale musical contexts. Further developing this area of study, it will also be pedagogically relevant to examine the effects of interleaving in novices' learning of progressively larger and more complex musical units such as harmony involving chord progressions, musical form, and musical styles (see Wong et al., 2020).

Conclusion

The question of how best to cultivate learners' aural skill of musical interval identification is one of the important implications for the development of sophisticated musicianship. Our results suggest that the answer to the question of which presentation schedule constitutes the "best" practice in melodic interval learning may depend on the nature of the learning context at hand. Developing a deeper understanding of the cognitive psychology of music—when exactly blocking versus interleaving is more (vs. less) effective—will, in turn, promote more nuanced applications of these strategies to enhance music education.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research of this article was supported in part by student research funds awarded by the National University of Singapore to the second author.

ORCID iDs

Sarah Shi Hui Wong https://orcid.org/0000-0003-4243-212X
Stephen Wee Hun Lim https://orcid.org/0000-0003-3636-7587

Supplemental material

Supplemental material for this article is available online.

References

Abushanab, B., & Bishara, A. J. (2013). Memory and metacognition for piano melodies: Illusory advantages of fixed- over random-order practice. *Memory & Cognition*, *41*, 928–937. doi:10.3758/s13421-013-0311-z

Austin, J. R., & Berg, M. H. (2006). Exploring music practice among sixth-grade band and orchestra students. *Psychology of Music*, *34*, 535–558. doi:10.1177/0305735606067170

Baker, D. J., Ventura, J., Calamia, M., Shanahan, D., & Elliott, E. M. (2018). Examining musical sophistication: A replication and theoretical commentary on the Goldsmiths Musical Sophistication Index. *Musicae Scientiae*. Advance online publication. doi:10.1177/1029864918811879

Barry, N. H. (1992). The effects of practice strategies, individual differences in cognitive style, and gender upon technical accuracy and musicality of student instrumental performance. *Psychology of Music*, *20*, 112–123. doi:10.1177/0305735692202002

Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), *Acquisition of skill* (pp. 215–244). New York: Academic Press.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100–130. doi:10.1016/j.cognition.2005.11.007

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory and Cognition*, *41*, 392–402. doi:10.3758/s13421-012-0272-7

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

Boberg, R. M. (1975). Ear-opening experiences with rhythm and pitch. *Music Educators Journal*, *64*, 32–39. doi:10.2307/3394921

Bruner, C. L. (1984). The perception of contemporary pitch structures. *Music Perception*, *2*, 25–39. doi:10.2307/40285280

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

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 215–264). San Diego, CA: Academic Press.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *The Journal of the Acoustical Society of America*, *63*, 456–468. doi:10.1121/1.381737

Buttram, J. B. (1969). The influence of selected factors on interval identification. *Journal of Research in Music Education*, *17*, 309–315. doi:10.2307/3343879

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

Carter, C. E., & Grahn, J. A. (2016). Optimizing music learning: Exploring how blocked and interleaved practice schedules affect advanced performance. *Frontiers in Psychology*, *7*, Article 1251. doi:10.3389/fpsyg.2016.01251

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1699–1719. doi:10.1037/xlm0000406

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380. doi:10.1037/0033-2909.132.3.354

Chen, C. W. J. (2015). Mobile learning: Using application Auralbook to learn aural skills. *International Journal of Music Education*, *33*, 244–259. doi:10.1177/0255761414533308

Dalla Bella, S., Peretz, I., & Aronoff, N. (2003). Time course of melody recognition: A gating paradigm study. *Perception & Psychophysics*, *65*, 1019–1028. doi:10.3758/BF03194831

Degrave, P., & Dedonder, J. (2019). A French translation of the Goldsmiths Musical Sophistication Index, an instrument to assess self-reported musical skills, abilities and behaviours. *Journal of New Music Research*, *48*, 138–144. doi:10.1080/09298215.2018.1499779

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. doi:10.1177/1529100612453266

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. doi:10.1016/j.jarmac.2017.07.005

Garner, A. M. (2009). Singing and moving: Teaching strategies for audiation in children. *Music Educators Journal*, *95*, 46–50. doi:10.1177/0027432109335550

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116–139. doi:10.1037//0096-3445.130.1.116

Gordon, E. E. (1985). Research studies in audiation: I. *Bulletin of the Council for Research in Music Education*, *84*, 34–50.

Jeffries, T. B. (1967). The effects of order of presentation and knowledge of results on the aural recognition of melodic intervals. *Journal of Research in Music Education*, *15*, 179–190. doi:10.2307/3343858

Jeffries, T. B. (1970). A further investigation of certain learning aspects in the aural recognition of melodic intervals. *Journal of Research in Music Education*, *18*, 399–406. doi:10.2307/3344507
Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103. doi:10.1002/acp.1801
Karpinski, G. S. (2000). *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians*. New York: Oxford University Press.
Killam, R. N., Lorton, P. V., & Schubert, E. D. (1975). Interval recognition: Identification of harmonic and melodic intervals. *Journal of Music Theory*, *19*, 212–234. doi:10.2307/843589
Klonoski, E. (1998). Teaching pitch internalization processes. *Journal of Music Theory Pedagogy*, *12*, 81–96.
Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*, 585–592. doi:10.1111/j.1467-9280.2008.02127.x
Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, *126*, 159–179. doi:10.1037//0033-2909.126.1.159
Lee, Y.-S., Janata, P., Frost, C., Martinez, Z., & Granger, R. (2015). Melody recognition revisited: Influence of melodic Gestalt on the encoding of relational pitch information. *Psychonomic Bulletin and Review*, *22*, 163–169. doi:10.3758/s13423-014-0653-y
Leon-Guerrero, A. (2008). Self-regulation strategies used by student musicians during music practice. *Music Education Research*, *10*, 91–106. doi:10.1080/14613800701871439
Leung, Y., & Dean, R. T. (2018). Learning unfamiliar pitch intervals: A novel paradigm for demonstrating the learning of statistical associations between musical pitches. *PLoS ONE*, *13*(8), e0203026. doi:10.1371/journal.pone.0203026
Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2020). Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*, *48*, 376–388. doi:10.1177/0305735618801997
Lin, H.-R., Kopiez, R., Müllensiefen, D., & Wolf, A. (2019). The Chinese version of the Gold-MSI: Adaptation and validation of an inventory for the measurement of musical sophistication in a Taiwanese sample. *Musicae Scientiae*. Advance online publication. doi:10.1177/1029864919871987
Little, D. F., Cheng, H. H., & Wright, B. A. (2019). Inducing musical-interval learning by combining task practice with periods of stimulus exposure alone. *Attention, Perception, & Psychophysics*, *81*, 344–357. doi:10.3758/s13414-018-1584-x
Madsen, C. K., & Geringer, J. M. (2000–2001). A focus of attention model for meaningful listening. *Bulletin of the Council for Research in Music Education*, *147*, 103–108.
Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. *Human Movement Science*, *9*, 241–289. doi:10.1016/0167-9457(90)90005-X
Maynard, L. M. (2006). The role of repetition in the practice sessions of artist teachers and their students. *Bulletin of the Council for Research in Music Education*, *167*, 61–72.
McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*, 1263–1271. doi:10.1111/j.1467-9280.2008.02235.x
Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 978–984. doi:10.1037/xlm0000216
Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*, e89642. doi:10.1371/journal.pone.0089642
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. doi:10.1111/modl.12581
Ottman, R. W. (1956). *A statistical investigation of the influence of selected factors on the skill of sight-singing* (Unpublished doctoral dissertation). University of North Texas College of Music, Denton.
Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (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. doi:10.1037/edu0000336

Ponsatí, I., Miranda, J., Amador, M., & Godall, P. (2020). Aural identification of harmonic intervals: An observational study at the Girona Music Conservatory (Catalonia, Spain). *Psychology of Music*, *48*, 448–462. doi:10.1177/0305735618809868

Ponsatí, I., Miranda, J., Amador, M., & Godall, P. (2018). Aural identification of harmonic intervals: An observational study at the Girona Music Conservatory (Catalonia, Spain). *Psychology of Music*. Advance online publication. doi:10.1177/0305735618809868

Rogers, M. R. (2004). *Teaching approaches in music theory: An overview of pedagogical philosophies* (2nd ed.). Carbondale: Southern Illinois University Press.

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

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

Rohwer, D., & Polk, J. (2006). Practice behaviors of eighth-grade instrumental musicians. *Journal of Research in Music Education*, *54*, 350–362. doi:10.1177/002242940605400407

Russo, F. A., & Thompson, W. F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin & Review*, *12*, 1068–1075. doi:10.3758/BF03206445

Samplaski, A. (2005). Interval and interval class similarity: Results of a confusion study. *Psychomusicology*, *19*, 59–74. doi:10.1037/h0094040

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*, 84–98. doi:10.1037/edu0000119

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*, 411–417. doi:10.1016/j.tics.2008.07.006

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

Siegel, J. A., & Siegel, W. (1977). Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, *21*, 143–152. doi:10.3758/BF03198717

Smith, J. D., Kemler Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition*, *52*, 23–54. doi:10.1016/0010-0277(94)90003-5

Spohn, C. L. (1963). Programming the basic materials of music for self-instructional development of aural skills. *Journal of Research in Music Education*, *11*, 91–98. doi:10.2307/3344147

Stambaugh, L. A. (2011). When repetition isn't the best practice strategy: Effects of blocked and random practice schedules. *Journal of Research in Music Education*, *58*, 368–383. doi:10.1177/0022429410385945

Stambaugh, L. A., & Nichols, B. E. (2020). The relationships among interval identification, pitch error detection, and stimulus timbre by preservice teachers. *Journal of Research in Music Education*, *67*, 465–480. doi:10.1177/0022429419885931

Tarratus, E. A., Jr., & Spohn, C. L. (1967). Cooperative research in programed learning: Taped interval discrimination skills. *Journal of Research in Music Education*, *15*, 210–214. doi:10.2307/3343861

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837–848. doi:10.1002/acp.1598

Thompson, W. F. (2013). Intervals and scales. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 107–140). San Diego, CA: Elsevier. doi:10.1016/B978-0-12-381460-9.00004-3

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. doi:10.3758/s13421-010-0063-y

Wapnick, J., Bourassa, G., & Sampson, J. (1982). The perception of tonal intervals in isolation and in melodic context. *Psychomusicology: A Journal of Research in Music Cognition*, *2*, 21–37. doi:10.1037/h0094264

Wasserman, M. (1974). *The design and validation of a self-instructional learning package for the acquisition of ascending melodic interval skills: A developmental process* (Doctoral dissertation). Available from ProQuest Dissertations and Theses databases. (Order No. 7429879)

Williamon, A. (2004). *Musical excellence: Strategies and techniques to enhance performance*. Oxford, UK: Oxford University Press.

Wong, S. S. H., & Lim, S. W. H. (2017). Mental imagery boosts music compositional creativity. *PLoS ONE*, *12*(3), e0174009. doi:10.1371/journal.pone.0174009

Wong, S. S. H., Low, A. C. M., Kang, S. H. K., & Lim, S. W. H. (2020). Learning music composers' styles: To block or to interleave? *Journal of Research in Music Education*. Advance online publication. doi:10.1177/0022429420908312

Wright, C. R. (2016). *Aural and the university music undergraduate*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, *145*, 918–933. doi:10.1037/xge0000177

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics*, *26*, 384–395. doi:10.3758/BF03204164

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215–221. doi:10.1016/j.learninstruc.2011.11.002