Optimizing Music Learning: The Contextual Interference Effect on Musical Style Learning

Xinjie (Jalor) Dong
Dr. Jessica Grahn, Rebekka Lagacé-Cusiac
MedSci 3999B: Experiential Learning

Introduction

Inductive category learning is the process of categorizing into groups based on shared features through exposure to examples (Kornell & Bjork, 2008). For instance, learning to identify different animal species relies on exposure to multiple examples, which helps form generalizable concepts about each species. This is a fundamental cognitive ability for learning since learners often construct knowledge by actively exploring examples rather than passively receiving information. Two schedules are widely known and can be used for inductive category learning: block and interleaved schedules. In a blocked schedule, one would repeatedly practice one concept before moving to the next (e.g., AAABBBCCC). In contrast, an interleaved schedule frequently mixes and alternates between concepts (e.g., ABCABCABC) throughout learning (Kornell & Bjork, 2008).

Previous studies have suggested an advantage in interleaved schedules in learning and long-term retention, which is referred to as a contextual interference effect. However, these studies mainly investigated visual and motor skills learning, which cannot differ from inductive category learning (Jaswal & Logie, 2013; Brady, 1985).

Interestingly, blocked practice remains widely used in many learning contexts despite its limitations in the long term. One primary reason is that depressed performance is often observed during interleaved practice, as constant task-switching introduces challenges to the learner. In contrast, blocked practice shows fast initial improvement through repetitive exposure to one concept and reduced cognitive load to avoid switching tasks (Czyż et al., 2024).

Research on the contextual interference effect in music education has been limited to applied performance, such as assessing playing speed, accuracy, and overall performance improvement (Mathias & Goldman, 2024; Carter & Grahn, 2016; Abushanab Bishara, 2013; Stambaugh). These studies all showed an advantage of interleaved practice in musicians when learning music, similar to the contextual interference effect observed in other fields. However, music education is not limited to practicing technical skills for performance; other skills, such as music perception, are also crucial in music learning (Lerch et al., 2020), and minimal research has been conducted in the other domains. Brunmair Richter (2019) has shown that the advantage in interleaved practice is not observed in all learning contexts. Thus, evidence of the effect of contextual interference in other domains of music education, such as music perceptive skills, is crucial to generalizing the effect in all music-learning contexts.

Music perceptive skills include differentiating among musical styles, historical eras, composers, and musical genres, which are fundamental skills to develop aural skills and musicianship in music education, and gaining these skills requires inductive category learning (Wong et al., 2020). For example, when learning to differentiate between composers' musical styles, one must compare and extract meaningful musical features such as the melody, harmony, and rhythms through exposure to exemplars. Wong et al. (2020) have previously investigated the effects of blocked and interleaved schedules on music style recognition in non-musicians. In music perception, specifically, learning musical styles from 12 composers spanning four major historical periods among people with minimal music training. The study was conducted in one session, including a study and test phases, aimed to learn and distinguish 12 composers spanning four major historical periods. During the study phase, participants were presented with 5-second excerpts from the composers in a blocked and interleaved schedule. After a 15-second distractor task, the test phase started, where participants were presented with novel excerpts to classify from the same composers. Authors found that participants more accurately classified composers learned in the interleaved schedule, yet they reported favoring the blocked schedule. Therefore, the results from this study are consistent with previous studies in other fields, such as painting style recognition. Thus, the findings in this study suggest that interleaved practice is potentially more effective in music education. However, it is limited to short, single-session experiments. Therefore, additional sessions should be introduced to deepen our understanding of how different practice schedules influence long-term learning and retention in music perceptive skills. Furthermore, they did not investigate the effect of musical experience as prior knowledge may benefit more from the blocked condition (Nemeth & Lipowsky, 2024). Additionally, Lagacé-Cusiac et al. (2025, submitted) have observed a larger effect of interleaved practice for musicians than non-musicians, while Wong et al. (2020) only investigated the effect in non-musicians.

Therefore, this study aims to use a longitudinal design to investigate the learning trajectory for interleaved and blocked practice in multiple training sessions. We are interested in exploring the difference in performance after learning the musical styles of jazz composers between the two practices. In addition, the effect of music expertise will be explored. It is hypothesized that interleaved practice will have a higher learning rate, leading to enhanced performance than blocked practice. We also hypothesize that musicians will learn faster than non-musicians, leading to faster-enhanced performance, especially during interleaved practice.

Methods

Participants

32 healthy adult participants aged 18 to 21 (Mean=19.58), including 22 females and 10 males, were recruited through the Mass Recruitment Email System at the University of Western Ontario. 20 of these participants were musicians with five or more years of music-playing experience, while 12 were non-musicians with less than 5 years of music-playing experience. All participants provided informed consent before participating with a compensation of $5 per session, and all procedures were approved by the Western University Non-Medical Research Ethics Board (REB approval code 109152).

Materials and Stimuli

Recordings by six jazz pianists were used: Chick Corea, Bill Evans, Oscar Peterson, Marian McPartland, Brad Mehldau, and Thelonious Monk.

The study included 44 excerpts from each pianist. To create the excerpts, improvising piano recordings were identified from each pianist and were cut into clips by Audacity ® (version 3.6.4) that became excerpts in the study. A maximum of 3 clips were selected from each recording. All excerpts were verified by an expert musician from the faculty of music at Western University and normalized to the sound level for control.

A questionnaire was created on Qualtrics to collect demographic data and participants' experiences during the study. Participants were asked to identify their gender, age, previous music experience, and preferred practice schedule (blocked, interleaved schedule, or no difference).

Experimental Setup

The PsychoPy software (Version 2024.2.4) was used to implement the study, including instructions and auditory stimuli, which were then uploaded to Palvolvia.org. Participants can access all sessions through Palvolvia.org using a web browser.

The study consisted of 4 sessions, each designed to occur over consecutive days from the first session. Each session included a practice and test phase, designed to be implemented over consecutive days from the first session. Session 1 was conducted in person in a quiet-controlled room using headphones to access the auditory stimuli at the Western Interdisciplinary Research Building. In contrast, other sessions were hosted online to be completed by the participants.

During each session, participants were exposed to 60 trials (10 excerpts x 6 pianists) for the practice phase and 24 trials (4 excerpts x 6 pianists) for the test phases, totaling 84 trials (14 excerpts x 6 pianists). During each practice phase, 36 trials (6 excerpts x 6 pianists) were novel to participants. Meanwhile, 24 trials (4 excerpts x 6 pianists) were recurring to participants, meaning they will be exposed to these trials repeatedly throughout the four sessions. All of the stimuli for the test phase were novel to participants for testing their learning. During each test session, participants were exposed to 60 trials (10 excerpts x 6 pianists), including 36 trials (6 excerpts x 6 pianists), with excerpts that were novel to participants, and 24 trials (4 excerpts x 6 pianists), with recurring excerpts where they had previously heard these excerpts from the learning sessions. The 24 recurring trials remained the same excerpts in both test sessions. At the end of Session 6, participants were asked to complete the demographic questionnaire.

Procedure

Each session included two blocks, one for each practice schedule condition: blocked and interleaved. Each block was then split into a practice phase and a test phase. Additionally, the six jazz pianists were split into two sets, each dedicated to one condition: one set consisted of Corea, Evans, and Peterson, and the other set consisted of Monk, Mehldau, and McPartland. Each pianist set was assigned to either the blocked or interleaved condition in a counterbalanced manner across participants. Additionally, the order of the learning schedule (blocked or interleaved) was counterbalanced across participants.

In the blocked condition, the participants listened to 10 excerpts performed by one pianist before moving on to the 10 excerpts performed by another, continuing until all three pianists from the composer set were heard (e.g., A…A, B…B, C…C). While the excerpt plays, the pianist’s name was shown on the screen, and participants were then prompted to select which pianist they had just listened to out of three pianists in the corresponding pianist set by clicking on the correct name on the screen. In the subsequent test phase, participants were presented with 4 excerpts by each of the three pianists they had just listened to in a randomized order. Participants were then asked to select the pianist’s name from the pianist list after each test excerpt played. They were then asked to rate their confidence between “guessing”, “ not very confident”, “somewhat confident”, and “very confident”. In the interleaved condition, participants listened to recordings of 10 excerpts by the other composer list, randomly interleaved such that no more than two of a pianist's excerpts were presented in succession (e.g., DEF, EDF, FED, …) in the practice phase with other procedures duplicating the blocked condition. They later entered the test phase similarly to the test phase in the blocked condition.

Statistical Analysis

4 participants (1 musician and 3 non-musicians) were removed from analyses due to observed inattention during practice. Additionally, 9 participants (5 musicians, 4 non-musicians) were excluded from analyses as they withdrew between sessions 1-4 or had incomplete data during their participation. Therefore, only 19 participants' data were analyzed, including 14 musicians and 5 non-musicians. All data from Palvolvia.org were exported and preprocessed, analyzed, and graphed using RStudio (version 2024.04.0+735) with tidyverse and afex packages. Accuracy was the dependent variable, calculated as the percentage of correct responses in each condition. A mixed-design ANOVA was conducted with Practice Schedule (Blocked, Interleaved) and Session (1–4) as within-subject factors and Musicianship (Musician, Non-Musician) as a between-subject factor. To further examine the influence of composer set difficulty, separate repeated-measures ANOVAs were performed within each pianist set (Set 1 and Set 2), comparing blocked and interleaved practice effects. Independent samples t-tests were also conducted to assess accuracy differences across composer sets.

For metacognitive analyses, participants' self-reported "most helpful" practice schedule was compared to their actual best-performing condition. A Fisher's Exact Test was used to assess the alignment between perceived and actual performance to account for small sample sizes.

Results

A 2 (Musicianship: Musician, Non-Musician) × 2 (Practice Schedule: Blocked, Interleaved) × 4 (Session: 1–4) mixed-design ANOVA was conducted. No significant effect of performance difference was found of practice schedules was found ($F(1, 18) = 0.91$, $p = 0.353$), nor a significant effect between practice schedules across the learning sessions ($F(2.78, 50.11) =$

0.52, p = 0.659). Table 1 showed the mean accuracy of participants across sessions. Figure 1 showed a visual representation of the trend in accuracy across sessions.

Session	Blocked Accuracy	Interleaved Accuracy
1	45.37 ± 5.12%	40.28 ± 3.78%
2	47.84± 5.53%	48.49 ± 3.89%
3	42.24 ± 2.57%	36.89± 2.57%
4	36.45 ± 6.36%	36.45 ± 4.70%

Table 1. Mean accuracy scores (%) for blocked and interleaved conditions across four learning sessions (N = 19)

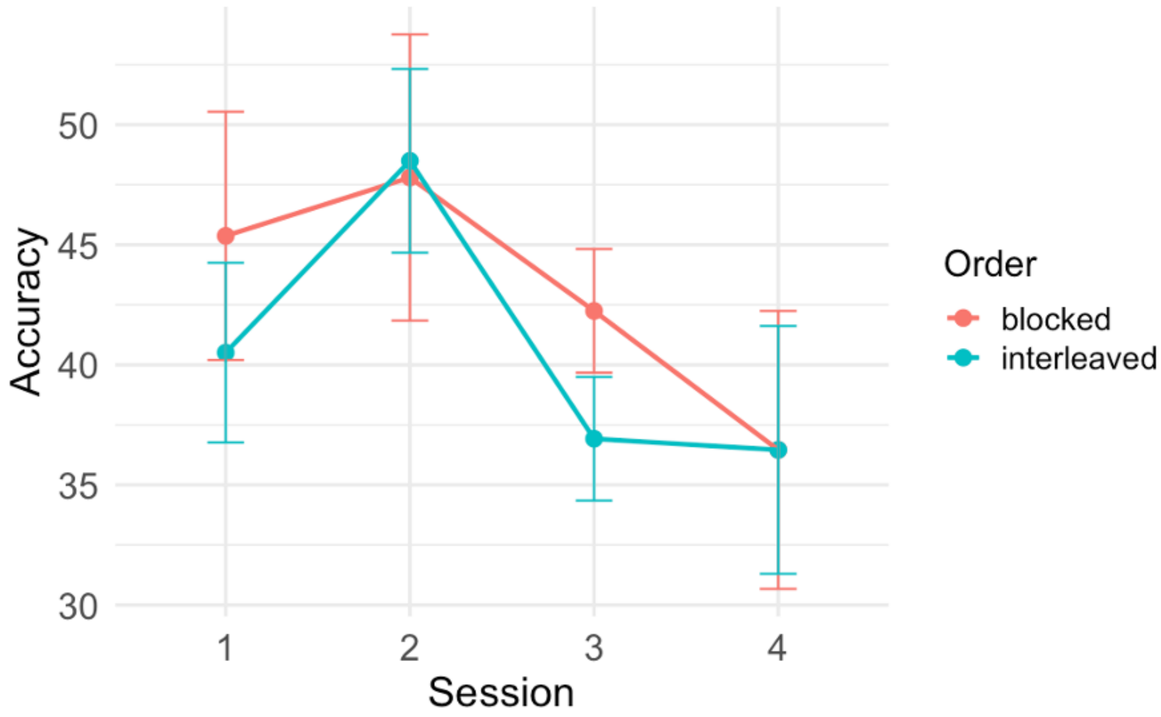

Figure 1. Mean accuracy (%) across four sessions in blocked and interleaved practice

Furthermore, no effect was found between the practice schedules and musicianship ($F(1, 18) = 1.37$, $p = 0.256$). The interactions between musicianship, session, and practice schedules also revealed no significance ($F(2.45, 39.26) = 0.80$, $p = .479$), suggesting musicians did not benefit from one condition throughout the learning sessions. (Table 2; Figure 2).

Session	Musicians-Blocked	Musicians-Interleaved	Non-Musicians - Blocked	Non-Musicians - Interleaved
1	47.9 ± 6.53%	45.8 ± 3.21%	38.9 ± 10.66%	44.4 ± 5.24%
2	51.4 ± 7.31%	51.4 ± 4.32%	37.5 ± 11.93%	45.8 ± 7.05%
3	50.7 ± 2.92%	43.8 ± 2.81%	34.7 ± 4.77%	34.7 ± 4.58%
4	44.4 ± 6.94%	53.5 ± 5.68%	29.2 ± 11.34%	20.8 ± 9.28%

Table 2. Mean accuracy scores (%) for musicians and non-musicians across sessions and conditions presented by musicianship group (Musicians: N = 14, Non-Musicians: N = 5), session (1–4), and practice condition (blocked or interleaved).

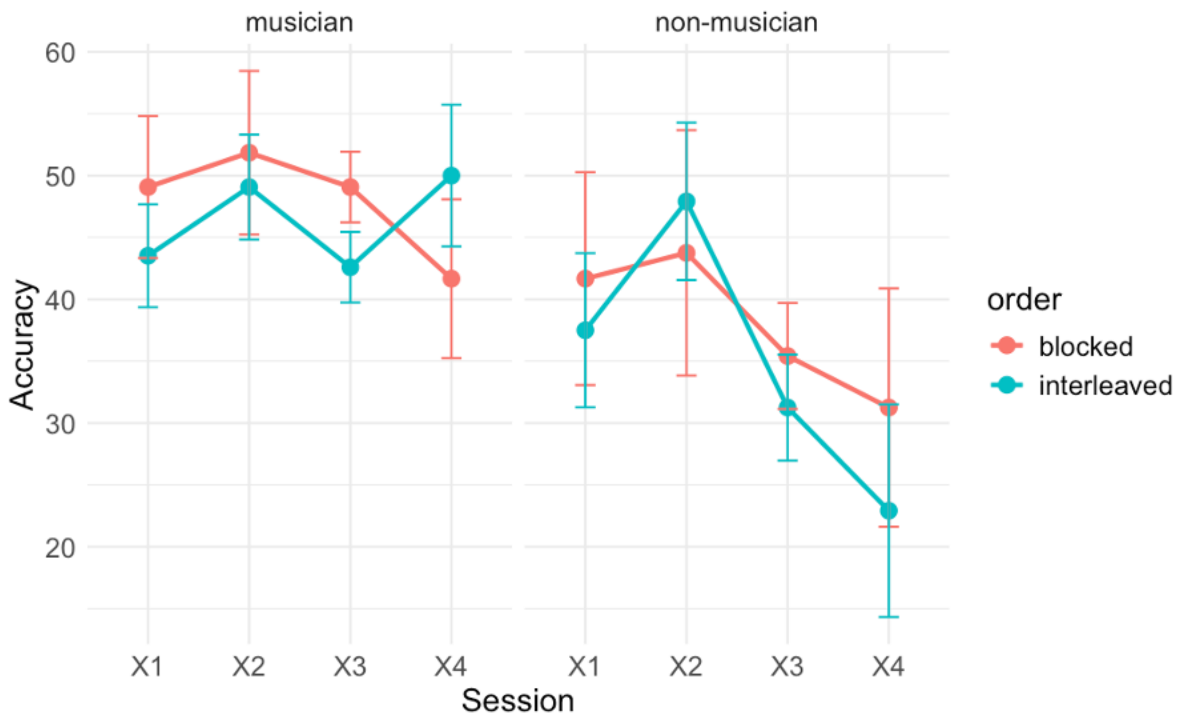

Figure 2. Mean accuracy (%) across four learning sessions by musicianship group (musician vs. non-musician) and practice condition (blocked vs. interleaved) in musicians and non-musicians.

Exploratory: The Effect of Composer Sets

Participants reported that one set of pianists was much easier to learn than the others, which may elicit different effects for the different sets. To test this hypothesis, we conducted two independent T-tests to explore whether an effect between blocked and interleaved was generally seen between the 2 pianist sets. The effect of practice schedules was insignificant for both pianist sets (set 1: ($t(95.95) = -0.85$, $p = 0.398$); set 2: ($t(73.24) = 0.39$, $p = 0.697$) Shown in Figure 3, the difference in the reported difficulty for the composer set in overall accuracy between the blocked and interleaved conditions. Although not significant, accuracy was higher for pianist set 1 in the interleaved condition, while performance for pianist set 2 decreased slightly. The effect of practice schedules was investigated separately within each pianist set to explore the impact of stimulus difficulty. In pianist set 1, no significant differences were found between practice schedules, yielding $F(1, 16) = 0.04$, $p = 0.837$ after a mixed-design ANOVA on the interactions of 2 (Musicianship: Musician, Non-Musician) × 2 (Practice Schedule: Blocked, Interleaved) × 4 (Session: 1–4). There was also no significant effect in pianist set 2, yielding $F(1, 16) = 0.08$, $p = 0.778$ after conducting the same statistical test. However, the performance trends across sessions differed between the 2 composer sets shown in Figure 4 and Figure 5.

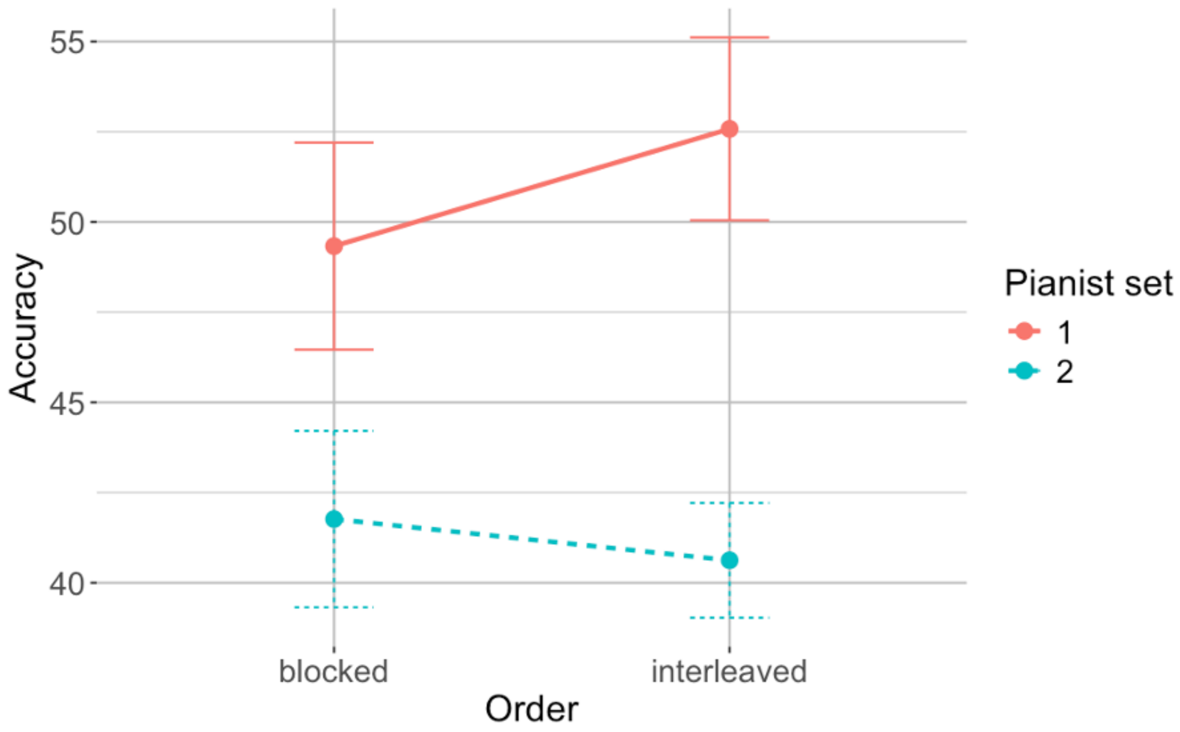

Figure 3. Mean accuracy (%) in blocked and interleaved conditions for pianist sets 1 and 2.

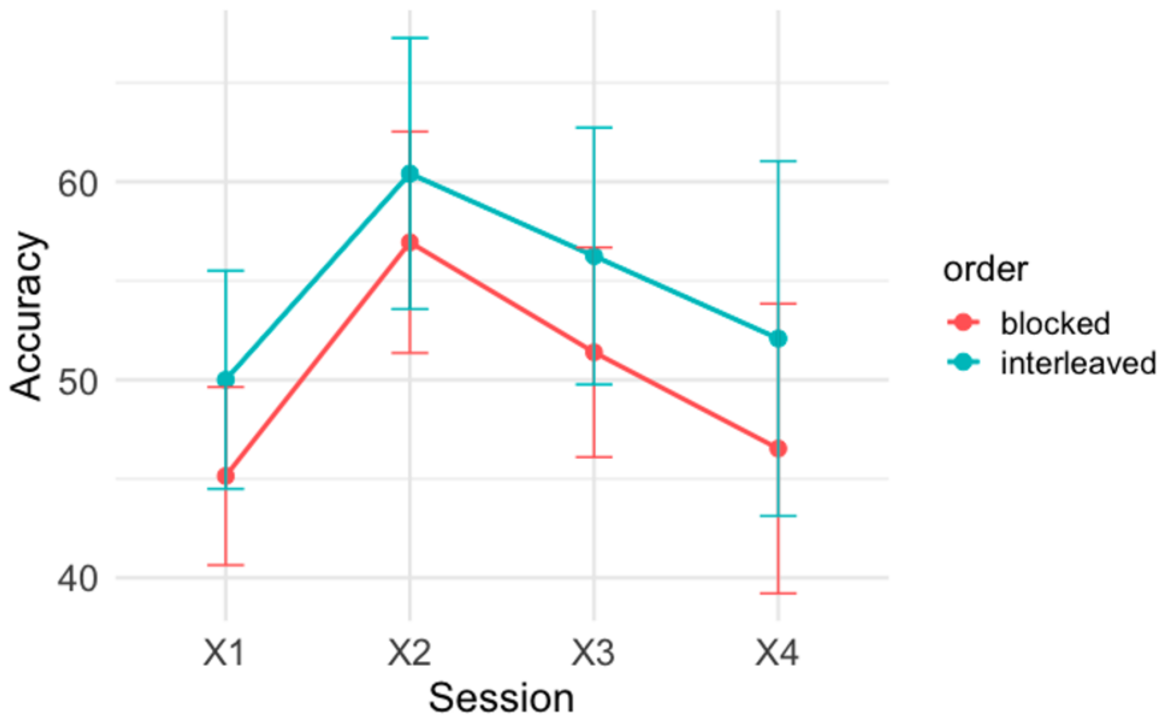

Figure 4. Mean accuracy for pianist set 1 across sessions in blocked and interleaved conditions

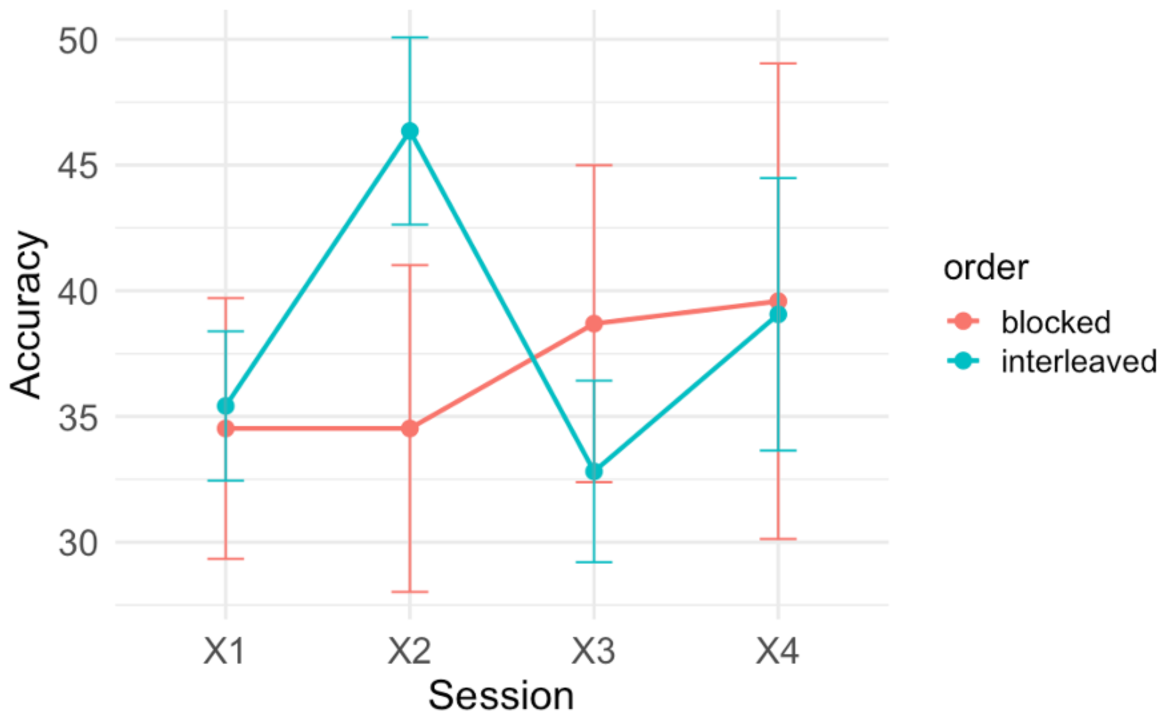

Figure 5. Mean accuracy for the pianist set 2 across sessions in blocked and interleaved conditions

Metacognitive Judgements

To examine participants' metacognitive awareness of practice effectiveness, we compared their self-reported "most helpful" condition from the post-experiment questionnaire with the condition in which they performed best. Only 8 out of 16 participants (50%) correctly identified the condition where their accuracy was higher (Table 3). These findings suggest that participants were not reliably aware of which learning schedule helped them most, and judgments of learning were largely misaligned with actual outcomes. A Fisher's Exact Test was also conducted, which confirmed the result, $p = 1.00$. These findings suggest that participants were not reliably aware of which learning schedule helped them most, and judgments of learning were largely misaligned with actual outcomes.

Accuracy (Best Performance)	Blocked	Equal	Interleaved
Blocked	6	1	1
Equal	1	0	0
Interleaved	5	0	2

Table 3. Contingency table comparing participants’ self-reported “most helpful” learning schedule (columns) with the condition in which they performed best (rows*)*. Each cell represents the number of participants falling into each response category.

Discussion

This study examined the effects of blocked and interleaved practice schedules on musical style recognition across multiple learning sessions, with a secondary focus on how musical expertise influenced performance. Contrary to our hypotheses, we found no statistically significant differences between practice schedules across the four sessions. The mean accuracy between blocked and interleaved conditions decreased and converged by Session 4, as shown in Figure 1, indicating that neither schedule produced a clear advantage in this long-term music style recognition. These results are inconsistent with other studies showing significant benefit of interleaved practice on music style recognition(Wong et al., 2020; 2021); (Lagacé-Cusiac et al., 2025, submitted). However, our findings align with Abel (2023), who similarly found no advantage of interleaved practice when learning to categorize bird calls—another form of auditory category learning. Abel has suggested that when a category is too similar, the effect of interleaved practice may disappear since interleaving is thought to enhance learning by highlighting contrasts between categories. In our study, although interleaved practice yielded slightly higher accuracy in Sessions 1 and 2, performance declined in Sessions 3 and 4, yielding similar accuracy to the blocked practice. Therefore, the results seen in our study may have occurred due to similarities in styles between composers (Abel, 2023). Additionally, the interleaved practice may introduce higher cognitive load and reduced motivation over repeated sessions without feedback, which may have impaired learning and lead to performance declines (Chen et al.,2021).

Musical Expertise and Learning

Contrary to our hypothesis, neither musicians nor non-musicians showed significant improvement with interleaved practice compared to blocked practice. A trend was observed in which musicians performed better in the interleaved condition than in the blocked condition. At the same time, non-musicians showed a drop in interleaved performance over time (Figure 2). This is consistent with the idea that interleaved practice may benefit skilled learners more by strengthening retrieval and discrimination processes (Brady, 2008; Richter et al., 2023). In contrast, non-musicians may lack the foundational frameworks needed to manage the cognitive demands of interleaving, resulting in cognitive overload and reduced retention previously seen in motor skills learning(Shea & Morgan, 1979; Magill & Hall, 1990). Musicians also attend more to fine musical details (melody, rhythm, harmony), while non-musicians focus on broader aspects like mood or pleasantness (Gromko, 1993), which may contribute to differences in the process of inductive category learning. Despite the trends that match previous findings, the lack of significant effect is attributed to a limited sample size of musicians and non-musicians.

Stimulus Difficulty and Composer Set Variability

The exploratory analysis conducted to investigate the effect of the practice schedule on the two pianist sets independently was done solely based on participant feedback and was not statistically significant. However, if one pianist set was inherently too difficult to distinguish, the potential effects of practice schedules may be masked (Gromko, 1993). Indeed, different trends were observed between the pianist sets. As shown in Figure 4, the interleaved condition in one of the pianist sets (pianist set 1) showed a general increase in accuracy compared to the blocked, although not statistically significant. Based on this fact, we further compared the practice schedules across sessions in pianist sets 1 and 2 separately. In pianist set 1, we observed a higher accuracy in interleaved practice across sessions, although not significant (Figure 4), which is consistent with the advantage of interleaved practice. In pianist set 2 (Figure 5), the interleaved group showed a spike in performance at Session 2 but a sharp drop by Session 3, eventually converging with the blocked condition in Session 4. This variability contrasts the studied benefit of interleaving as it should consistently benefit more than blocked practice. It may reflect the instability of learning effects under interleaved conditions in this particular set due to being more difficult than pianist set 1.

Metacognitive awareness

We found that participants were not reliably aware of which learning schedule best supported their performance. Only 50% correctly identified the condition in which they performed best, and a Fisher's Exact Test confirmed no significant alignment between perceived helpfulness and actual performance. This suggests that metacognitive judgments were largely inaccurate, particularly in distinguishing between learning schedules (Metcalfe, 2009). While participants were not systematically better in either the interleaved or blocked condition, many still reported blocked practice as the most helpful. Interestingly, even among participants whose performance was not highest in the blocked condition, the blocked practice remained their preferred schedule. This aligns with prior research on metacognitive illusions, where learners misjudge their learning under more challenging conditions, such as interleaved practice, which can feel harder despite offering long-term benefits (Abushanab & Bishara, 2013). The findings reinforce the idea that perceived ease may bias learners toward blocked schedules, even when these do not necessarily enhance retention. a Fisher's Exact

Test confirmed no significant alignment between perceived helpfulness and actual performance. This suggests that metacognitive judgments were largely inaccurate, particularly in distinguishing between learning schedules.

Limitations and Future Directions

One limitation of the present study was the potential variability in difficulty between the two composer sets. As mentioned previously, one set may have been easier to distinguish than the other, possibly masking the effects of the practice schedule. In particular, the composer set with more significant feature overlap may have been too complicated for participants to extract consistent stylistic cues, thereby limiting learning outcomes regardless of practice condition. Additionally, each participant was assigned the same composer set to either the blocked or interleaved condition throughout the study. While the composer-to-condition assignment was counterbalanced across participants, this fixed pairing could have introduced stimulus bias, favoring one condition if a composer set was inherently easier or more memorable. Future studies may utilize a different study design by randomizing composers into the two conditions across sessions to prevent fixed set-condition pairings without using set composer sets to increase generalizability.

Furthermore, jazz improvisation is a highly nuanced and stylistically complex domain; thus, the findings may not be generalized to other areas of music education. Therefore, concerns arise about the representativeness of the stimuli. Future research should utilize genres that have known stylistic (e.g., classical or pop) to control for rhythm, harmony, and timbre. Computational music analysis tools (e.g., Music Information Retrieval) may extract and control measurable musical features, allowing for greater stimulus validity in learning targets.

Some excerpts utilized in the study also suffered from poor audio quality, such as containing background noise. These inconsistencies may have introduced confounding variables, as participants could have learned to associate irrelevant auditory cues—such as noise artifacts—with specific composers rather than relying on musical features. Future studies should ensure consistent audio quality across all stimuli to improve internal validity. A pilot study may help identify whether non-musical features unintentionally influence classification accuracy.

Another potential limitation lies in the learning environment set by the study design. Sessions 2–4 were conducted online and completed independently by participants. This lack of environmental control may have led to variability in attention, audio playback quality, and overall engagement, potentially compromising learning consistency across participants. Furthermore, the study was initially designed for consistent learning across consecutive days (Sessions 1–4), but in practice, few participants followed this schedule. Some participants had long and inconsistent delays between sessions, potentially disrupting learning and weakening the effects of practice schedules. Future studies should utilize stricter controls, such as all in-person sessions, to maintain standardization. Additionally, comparisons between online and in-person learning conditions could be investigated with a greater sample size to understand better how other factors affect perceptual learning and retention.

Finally, the current study only measured short- to mid-term learning and retention, with sessions designed to occur over consecutive days. Prior literature (Brady, 1998; Rohrer & Taylor, 2007) suggested that the benefits of interleaved practice often emerge over longer retention intervals, such as 2 to 4 weeks or more. As such, our design may not have fully captured the long-term advantages of interleaved learning. Future research can implement longer follow-up testing intervals to assess durable learning effects better and determine whether interleaved practice leads to delayed but more robust retention gains.

Implications

Based on our findings, the contextual interference effect may not be inherently effective for learning musical styles that are difficult to distinguish even after long periods of exposure. These findings challenge the assumption that interleaved practice is universally better in music education contexts. While interleaving may enhance retention and discrimination in technical or performance-based skills (Carter & Grahn, 2016; Mathias & Goldman, 2024), it may not always be optimal for perceptual tasks such as identifying musical styles or aural recognition — especially when stimuli are difficult or when learners lack prior expertise. For music educators, this implies a need to tailor practice schedules to task type and student background. A hybrid approach may be introduced where learners start with blocked exposure to build familiarity, followed by interleaved sessions to promote flexible recognition — this may better support learners across different levels of musical experience. Furthermore, the observed metacognitive misalignment highlights the need for explicit instruction about effective learning strategies. Learners often equate fluency with effectiveness and may avoid more challenging approaches like interleaving. Therefore, educators play a crucial role in helping students understand that some challenges can be desirable and that effortful learning often leads to better long-term outcomes.

Beyond music, our findings contribute to the broader conversation about the contextual interference effect in long-term, cognitively demanding learning tasks. Interleaved practice has enhanced performance in domains like motor learning, sports training, and visual categorization (Brady, 1998; Shea & Morgan, 1979). However, these effects are often tested in short-term contexts or under highly controlled conditions. Our results suggest that in realistic, longitudinal training scenarios, the effectiveness of interleaving may depend on factors such as task difficulty, stimulus similarity, learner motivation, and feedback availability(Magill & Hall, 1990; Wulf & Shea, 2002). Thus, our findings suggest that interleaving is not universally optimal across all domains or learners. Instead, it should be used strategically, in combination with blocked or scaffolded approaches, depending on learner expertise, task complexity, and the stage of learning.

Conclusion

This study explored the effects of blocked and interleaved practice schedules on musical style learning across multiple sessions in musicians and non-musicians. We found no benefit of interleaved practice, suggesting that interleaved practice may not be universally superior in all music learning contexts. Participants, however, tended to underestimate the benefits of interleaved learning, highlighting the presence of metacognitive illusions. This misalignment between perceived and actual learning outcomes reinforces the importance of educating learners on effective, evidence-based learning strategies.

Our findings suggest that the contextual interference effect may not generalize seamlessly to all forms of music education, particularly in long-term perceptual learning tasks involving complex auditory categorization. While interleaved practice may still hold value, especially for learners with higher expertise, educators should consider adopting hybrid practice schedules that begin with blocked exposure and transition to interleaved formats. Future research with larger sample sizes, controlled stimulus design, and extended retention intervals is needed to clarify the conditions under which interleaved practice is most effective. These insights are crucial for optimizing music education and refining how contextual interference is applied across diverse learning domains.

Appendix 1. ANOVA Table (Type III Tests) of mixed-design ANOVA on the interactions of 2 (Musicianship: Musician, Non-Musician) × 2 (Practice Schedule: Blocked, Interleaved) × 4 (Session: 1–4)

Effect	Num Df	Den Df	MSE	F	ges	Pr(>F)
Musicianship	1.000	18.000	576.50	2.9543	0.036119	0.1028
Order	1.000	18.000	508.41	0.9084	0.010059	0.3532
Musicianship:Order	1.000	18.000	508.41	1.3745	0.015142	0.2563
Session	1.993	35.873	418.70	1.6219	0.028916	0.2117
Musicianship:Session	1.993	35.873	418.70	0.5952	0.010809	0.5563
Order:Session	2.784	50.112	217.57	0.5170	0.006843	0.6590
Musicianship:Order:Session	2.784	50.112	217.57	0.1889	0.002511	0.8913

Appendix 2. ANOVA Table (Type III Tests) of mixed-design ANOVA on the interactions of 2 (Musicianship: Musician, Non-Musician) × 2 (Practice Schedule: Blocked, Interleaved) × 4 (Session: 1–4) in pianist set 1 only.

Effect	Num Df	Den Df	MSE	F	ges	Pr(>F)
Order	1.0000	16.000	561.99	0.0436	0.000927	0.8372
Musicianship	1.0000	16.000	561.99	1.6617	0.034119	0.2157
Order:Musicianship	1.0000	16.000	561.99	1.1558	0.023980	0.2983
Session	2.4535	39.257	444.36	0.7792	0.031134	0.4895
Order:Session	2.4535	39.257	444.36	0.4195	0.017006	0.7009
Musicianship:Session	2.4535	39.257	444.36	0.7867	0.031424	0.4858
Order:Musicianship:Session	2.4535	39.257	444.36	0.8017	0.032004	0.4785

Appendix 3. ANOVA Table (Type III Tests) of mixed-design ANOVA on the interactions of 2 (Musicianship: Musician, Non-Musician) × 2 (Practice Schedule: Blocked, Interleaved) × 4 (Session: 1–4) in pianist set 2 only.

Effect	Num Df	Den Df	MSE	F	ges	Pr(>F)
Order	1.0000	16.000	320.35	0.0823	0.002367	0.77794
Musicianship	1.0000	16.000	320.35	3.7382	0.097340	0.07108
Order:Musicianship	1.0000	16.000	320.35	0.7173	0.020273	0.40952
Session	2.0366	32.585	183.50	0.6025	0.019872	0.55633
Order:Session	2.0366	32.585	183.50	1.1676	0.037807	0.32451
Musicianship:Session	2.0366	32.585	183.50	0.6549	0.021565	0.52884
Order:Musicianship:Session	2.0366	32.585	183.50	1.1589	0.037535	0.32719

References

Abel, R. (2023). Interleaving Effects in Blindfolded Perceptual Learning Across Various Sensory Modalities. Cognitive Science, 47: e13270.

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.

Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. Quest, 50(3), 266–293.

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

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

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

Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. Educational Psychology Review, 33(4), 1499–1522.

Czyż, S. H., Wójcik, A. M., Solarská, P., & Kiper, P. (2024). High contextual interference improves retention in motor learning: systematic review and meta-analysis. Scientific Reports, 14(1), 15974.

Do, L. A., & Thomas, A. K. (2023). The underappreciated benefits of interleaving for category learning. Journal of Intelligence, 11(8), 153.

Gromko, J. E. (1993). Perceptual differences between expert and novice music listeners: A multidimensional scaling analysis. Psychology of Music, 21(1), 34–47.

Jaswal, S., & Logie, R. H. (2013). The contextual interference effect in visual feature binding: What does it say about the role of attention in binding? Quarterly Journal of Experimental Psychology, 66(4), 687–704.

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? Psychological Science, 19(6), 585–592.

Lagacé-Cusiac, R., Kinghorn, E., Grahn, J. A., De Souza, J., & Carter, C. E. (2025). The effect of blocked and interleaved practice on musical style learning in musicians and non-musicians. Manuscript submitted for publication.

Lerch, A., Arthur, C., Pati, A., & Gururani, S. (2020). An Interdisciplinary Review of Music Performance Analysis. Transactions of the International Society for Music Information Retrieval, 3(1), 221–245.

Magill, R. A., & Hall, K. G. (1990). A review of the contextual interference effect in motor skill acquisition. Human Movement Science, 9(3-5), 241–289.

Mathias, T., & Goldman, A. (2024). How does increasing contextual interference in a musical practice session affect acquisition and retention? Journal of Research in Music Education, 72(4), 357–375.

Metcalfe J. (2009). Metacognitive Judgments and Control of Study. Current directions in psychological science, 18(3), 159–163.

Nemeth, L., & Lipowsky, F. (2024). The role of prior knowledge and need for cognition for the effectiveness of interleaved and blocked practice. European Journal of Psychology of Education, 39, 907–929.

Richter, T., Nemeth, L., Berger, R., Ferri, R. B., Hänze, M., & Lipowsky, F. (2022). Using interleaving to promote inductive learning in educational contexts: Promises and challenges. Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 54(4), 164–175.

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

Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. Journal of Educational Psychology, 107(3), 900–908.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. Journal of Experimental Psychology: Human Learning and Memory, 5(2), 179–187.

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, 68(2), 156–174.

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