Improving Diagnostic Accuracy of Lung Auscultation Through Interleaved Practice
A Quasi-Experimental Field Study

Onan, Erdem; Onan, Arif; Ozgun, Ezgi; Gundogdu, Semra; Bektas, Hicran; de Bruin, Anique B.H.

DOI
10.1002/acp.70063

Publication date
2025

Document Version
Final published version

Published in
Applied Cognitive Psychology

Applied Cognitive Psychology

WILEY

SHORT PAPER

Improving Diagnostic Accuracy of Lung Auscultation Through Interleaved Practice: A Quasi-Experimental Field Study

Erdem Onan[1,2] | Arif Onan[3] | Ezgi Ozgun[3] | Semra Gundogdu[4] | Hicran Bektas[4] | Anique B. H. de Bruin[1]

[1]School of Health Professions Education, Maastricht University, Maastricht, the Netherlands | [2]Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, the Netherlands | [3]Department of Medical Education, Akdeniz University, Antalya, Turkey | [4]Department of Nursing, Akdeniz University, Antalya, Turkey

Correspondence: Arif Onan (aonan@akdeniz.edu.tr)

Received: 13 September 2024 | **Revised:** 22 January 2025 | **Accepted:** 16 April 2025

Keywords: auscultation | clinical reasoning | desirable difficulties | diagnostic accuracy | interleaved practice

ABSTRACT

Health professions educators are increasingly encouraged to implement *desirable difficulties* in their instruction, such as interleaved practice. In practical context, however, there is limited empirical evidence regarding the (meta)cognitive benefits of desirable difficulties, and interleaved practice in particular, posing a challenge to theoretical propositions. In this quasi-experimental field study, we examined the effectiveness of interleaved practice in auscultation training for second-year nursing students, with a focus on their learning outcomes and relative monitoring accuracy. Over 3 weeks, we measured participants' immediate and delayed-test scores, monitoring accuracy, and metacognitive knowledge of blocked and interleaved practice. Results revealed that interleaved practice yielded better auscultation performance than blocked practice. Regarding metacognitive accuracy, however, we found no statistically significant benefit of interleaving. Many students were unaware of the learning benefits of interleaved practice and found it more effortful than blocking. Our findings indicate that interleaved practice is a viable instructional method that can be utilized in authentic environments.

1 | Introduction

Cognitive psychology has provided valuable insights to health professions educators about designing effective instruction. One insight that has drawn substantial attention is the concept of *desirable difficulties*. This term refers to learning conditions that make initial learning more effortful, thereby slowing down immediate performance but increasing chances of long-term learning and transfer (Bjork et al. 2013; Bjork and Bjork 2011) Accordingly, several recommendations are made (Cecilio-Fernandes et al. 2023; Nelson and Eliasz 2023) encouraging health professions educators to design instruction wherein students engage in desirable difficulties, such as retrieval practice (i.e., recalling information from long-term memory) and interleaved practice (i.e., introducing variability to the study sequence). However, a critical gap exists between theoretical propositions and empirical validation in authentic settings. For instance, although interleaved practice has been tested in controlled environments with educationally less relevant materials (e.g., bird species), its true efficacy in classrooms remains unexplored. In this field experiment, we examined the cognitive and metacognitive benefits of interleaved practice in auscultation training of nursing students.

1.1 | Research on Blocked and Interleaved Practice in Health Professions Education (HPE)

Interleaved practice concerns the strategic sequencing of to-be-learned information. This instructional method entails a mixed study sequence, in which students alternate between topics during a study session (ABC-ABC) (Kornell and Bjork 2008). This approach stands in contrast to blocked practice, in which students study one topic before moving to the next (AA-BB-CC) (Kornell and Bjork 2008). A growing body of research shows that interleaved practice yields better learning outcomes than blocked practice (for review studies, see Brunmair and Richter 2019; Firth et al. 2021). More specifically, the interleaving benefit on learning has been shown in domains such as problem-solving tasks in mathematics (Rohrer et al. 2020) and physics (Samani and Pan 2021), pattern recognition in chemistry (Eglington and Kang 2017), learning of science concepts (Sana and Yan 2022), and even when learning grammatical rules of foreign languages (Pan et al. 2025; Schweppe et al. 2025).

Researchers have proposed two hypotheses to explain the learning benefits of interleaved practice: the distributed practice (Foster et al. 2019) and the discriminative contrast hypotheses (Birnbaum et al. 2013). The distributed practice hypothesis suggests that interleaved practice benefits learning through spacing. Specifically, when students alternate exemplars from different categories, they introduce temporal gaps between successive exemplars of the same category. These temporal gaps require students to recall previously studied information when revisiting a category, a well-known retrieval process that benefits learning (Bjork and Bjork 2011). The discriminative contrast hypothesis suggests that interleaving facilitates learning by prompting students to recognize subtle differences between categories. Alternating exemplars from different categories provides students with opportunities to compare and contrast, making distinctions across categories more salient. As a result, interleaved practice becomes particularly beneficial when students learn hard-to-distinguish categories (Carvalho and Goldstone 2014). These mechanisms, together, underscore the potential of interleaved practice in HPE as a promising instructional method. That is, clinical reasoning, a fundamental skill in health care, heavily relies on recognizing patterns and distinguishing between similar symptoms based on prior exposure (Monteiro and Norman 2013; Schmidt and Mamede 2020). By leveraging the benefits of distributed practice, interleaved practice may help students to reinforce memory retrieval of previously encountered cases, while discriminative contrast supports their ability to identify subtle distinctions between clinical presentations.

Despite growing interest, the application of interleaved practice within HPE remains limited (Thompson and Hughes 2023). Nevertheless, the scarce research in this area highlights the potential benefits of interleaved practice. For instance, Hatala et al. (2003) examined the impact of study sequence on students' ability to interpret electrocardiogram records. Medical students were randomized into a *contrastive* and *non-contrastive* condition. In the contrastive condition, students followed an interleaved sequence and made comparisons between diagnoses. In the non-contrastive condition, students followed a blocked sequence. The results indicated that diagnostic accuracy was higher in the contrastive condition than in the non-contrastive condition.

Rozenshtein et al. (2016) found further evidence of the benefits of interleaved practice in a radiology training. First-and second-year medical students learned 12 types of chest patterns, using both methods. Their findings indicated that recognition of previously studied exemplars, as well as novel exemplars, was higher for the interleaved patterns than for the blocked patterns. Crucially, while second-year students outperformed first-year students, both cohorts derived greater benefits from interleaved practice, indicating that students with varying levels of prior knowledge benefit from interleaved practice. Although promising, it is important to note that prior investigations of interleaved practice focused primarily on visual materials—representing an essential yet limited element of HPE.

1.2 | Metacognitive Aspects

In addition to learning outcomes, there are significant metacognitive considerations involved in using blocked and interleaved practice, as these study techniques may affect students' metacognitive experiences (i.e., perceived effort and perceived learning) and the accuracy of their metacognitive judgments (de Bruin et al. 2023). For example, several studies (Janssen et al. 2023; Kirk-Johnson et al. 2019, Onan et al. 2022) indicated that although interleaved practice leads to better learning than blocked practice, students often perceive the opposite. Specifically, when using interleaved practice, they tend to experience higher effort and lower learning. These experiences, in turn, lead to a preference for blocked practice, preventing students from taking effective study decisions (de Bruin et al. 2023).

A second and less explored metacognitive consideration is how blocked and interleaved practice affect students' monitoring accuracy (i.e., how well students judge their understanding or progress toward a learning goal), which is essential to make effective regulation decisions (Kämmer et al. 2020; Nelson and Eliasz 2023). For instance, if students erroneously believe they have grasped a subject, they might prematurely cease their study efforts. Notably, perceived learning and monitoring accuracy refer to different aspects of metacognition. Perceived learning reflects how much information students feel they have learned, regardless of their actual learning; whereas monitoring accuracy reflects how closely those feelings (or judgments) align with actual learning. In the literature, monitoring accuracy is often expressed in two forms: absolute and relative accuracy (Schraw 2009). Absolute accuracy captures the exact difference between students' judgments and their actual learning of a specific piece of information. Relative accuracy, which is the focus of this study, captures students' ability to distinguish between well and poorly understood information. In the clinical context, which often involves time pressure and uncertainty, relative accuracy becomes especially critical, since health professionals frequently make likelihood judgments, such as for narrowing down diagnostic options or determining which tasks or patients require immediate attention. Accordingly, accurate comparison to determine the likelihood of competing or multiple options is critical for allocating cognitive resources effectively and for making informed decisions.

Arguably, interleaved practice might improve students' monitoring accuracy (Eglington and Kang 2017). By mixing learning materials, interleaved practice heightens cognitive engagement and disrupts fluent information processing (Kirk-Johnson et al. 2019; Onan et al. 2022). In contrast, blocked practice might create an illusion of learning by offering students a fluent learning experience: Immediately repeated exposure to the same type of information might induce a false sense of confidence in one's ability to recognize information later (Yan et al. 2016). Currently, the metacognitive benefits of interleaved practice remain largely unexplored in HPE research and beyond.

1.3 | Interleaved Practice in Auscultation Training

As mentioned, evidence concerning the benefits of interleaved practice is mostly confined to visual materials. HPE, however, encompasses more modalities, including the auditory modality. Auscultation (i.e., listening to internal body sounds) training serves as a prime example where the auditory modality takes center stage. This training is a critical component of clinical reasoning as it provides a noninvasive and cost-effective method to assess organ function, enabling health professionals to detect early signs of disease, monitor its progression, and make informed decisions about treatment. Several studies, however, suggested that students and professionals struggle to make correct diagnoses about auscultatory irregularities (Hafke-Dys et al. 2019; Moriki et al. 2023; Williams et al. 2009). For instance, Williams et al. (2009) examined the diagnostic accuracy of paramedic students from two Australian universities. Results revealed that students from both institutes had great difficulties in correctly categorizing common lung sounds (e.g., Crackles, Stridor, and Wheeze). These findings led the researchers to conclude that students need dedicated training early in their studies.

Can interleaved practice be used to improve auscultation skills? Although limited, prior research also suggests that the benefits of interleaved practice may extend beyond the visual modality (Abel 2023; Chen et al. 2015). For instance, Wong et al. (2020) examined the effectiveness of blocked versus interleaved practice in music education. In their study, students were tasked with learning the musical styles of 12 composers. For half of the composers, students practiced music pieces in a blocked fashion, while for the other half, they practiced in an interleaved fashion. Afterward, students were asked to classify a novel piece of music by the composers they had studied. The findings revealed that, despite overall low performance, students who engaged in interleaved practice demonstrated better classification accuracy than those who practiced in a blocked format (for a similar study, see Wong et al. 2021).

In the HPE domain, more direct evidence comes from Chen et al. (2015), who examined the effectiveness of interleaved practice in auscultation training of nursing students. Researchers recruited a small number of senior-level students ($N = 22$). Again, half of the students applied blocked practice while the other half applied interleaved practice. Results revealed that participants who followed an interleaved sequence performed better than students who followed a blocked sequence. Although these results are promising, evidence is lacking for the potential impact of interleaved practice in a larger sample and in authentic contexts, where learning takes place in a dynamic and noisy environment with a larger and more diverse group of students.

1.4 | The Present Study

The present study tested the effectiveness of blocked and interleaved practice in nursing students' auscultation training. First, we examined how these study techniques influenced students' diagnostic success, relative monitoring accuracy, and perceived learning when learning auscultation. Then, we examined students' knowledge and effort anticipation of blocked and interleaved practice. Research questions and hypotheses were as follows:

How do blocked and interleaved practice influence …

RQ 1: … nursing students' diagnosis of (ab)normal respiratory sounds?

- **Hypothesis 1.** *Interleaved practice would result in higher diagnostic accuracy than blocked practice. This difference in accuracy would be larger in the delayed test than in the immediate test.*

RQ 2: … nursing students' monitoring accuracy of their auscultation performance?

- **Hypothesis 2.** *Blocked practice would lead to higher perceived learning than interleaved practice. Due to inconclusive previous findings* (e.g., Eglington and Kang 2017; Foster et al. 2023), *we formulated no* a priori *hypothesis for monitoring accuracy.*

RQ 3: What is nursing students' perception of blocked and interleaved practice in terms of the effectiveness of these study techniques?

- **Hypothesis 3.** *Students would believe that blocked practice leads to better learning than interleaved practice.*

2 | Methods

2.1 | Transparency

The study was approved by the ethical review board of Akdeniz University, Faculty of Medicine: file number KAEK-782. This article's design, research questions, and hypotheses were preregistered, https://aspredicted.org/BBP_4P5. Additionally, we decided to explore students' effort perceptions of blocked and interleaved practice. This exploration was by omission not preregistered.

2.2 | Participants

We recruited two classes from the Nursing Department of Akdeniz University. Participants were second year undergraduate students ($N = 190$). Of the participants, 72% were female and 28% were male. The average age was 20.71 (SD = 1.75).

Participants were assigned to these classrooms on a single day by the admission office at the start of their undergraduate studies to optimize the use of limited resources (e.g., teaching staff, classroom sizes). The procedure for this assignment was as follows: First, each student received a unique ID (student number) based on the order of registration. Then, students with odd numbers were assigned to one classroom; those with even numbers were assigned to the other. This assignment was free from any academic criteria, such as entrance scores or high school GPA. All participants were full-time students, and there were no major differences in scheduling, such as one group having early morning classes and the other having late evening classes.

We conducted an a priori power analysis to determine the required sample size. Using G*power 3.1.9.7 (Faul et al. 2009), we estimated the required sample size based on a 2×2 repeated measures, within-between interaction with the following parameters: $\eta^2_p = 0.02$, $\alpha = 0.05$, $1 - \beta = 0.80$. This calculation yielded that we needed at least 100 participants to test our hypothesis (for details see the preregistration form).

2.3 | Design

In a quasi-experimental field study, the instructional method was manipulated as a between-subjects factor (blocked or interleaved practice). Participants' auscultation performance was measured twice (after 5 min and 1 week). Hence, we followed a 2×2 mixed factorial design.

2.4 | Materials

2.4.1 | (Ab)normal Respiratory Sounds

Throughout the study, participants were presented with 10 exemplars of each of six (ab)normal respiratory sounds: Normal Bronchial, Normal Vesicular, Fine Crackle, Coarse Crackle, Rhonchi, and Wheeze. These categories were selected in consultation with the course coordinators. A Doctor of Medicine (MD) recorded these exemplars during pulmonary auscultation, and two residents verified them. Six of the exemplars were used during the study phases. The remaining exemplars were used in the immediate and delayed tests.

2.4.2 | Study Units

We created six blocked and six interleaved study units. All units were stored as an mp4 file. Blocked units consisted of six exemplars from the same respiratory sound. Interleaved units consisted of six exemplars, one from six different respiratory sounds. Respiratory sounds were 10 s long. There was a 3-s break after each sound.

2.4.3 | Prior Knowledge and Classification Test

We measured participants' general knowledge about the respiratory system. This test included 10 multiple-choice questions with five options, one correct option and four lures (see Appendix A).

The immediate and delayed classification tests assessed participants' ability to correctly identify a novel respiratory sound. In a multiple-choice format, participants were asked to select the appropriate category from a list of all six categories. Each test consisted of 12 items, two items per respiratory sound. Participants received 1 point per correct answer.

2.4.4 | Category Learning Judgments (CLJs)

For each respiratory sound, participants evaluated their likelihood of identifying a new exemplar one week later (i.e., Please answer the following question from 0% to 100%. How likely do you think you will be able to identify this respiratory sound one week later?). These CLJs indicated students' perceived learning and were used to express students' relative accuracy, which is typically calculated through within-person gamma correlations (Nelson 1984) between CLJs and students' classification performance.

2.4.5 | Metacognitive Knowledge and Effort Ratings

Participants' knowledge of blocked and interleaved practice was measured using a written scenario (Morehead et al. 2016). The scenario was adapted to the context of auscultation training and described two professors to students. Professor A implemented blocked practice in their class, while Professor B implemented interleaved practice. Participants were asked to choose whose students would learn better: Professor A, Professor B, or equal. Participants rated the effort demands of each instructional method on a 9-point Likert scale (i.e., How much mental effort do you think Professor A/B's method requires to learn respiratory sounds?).

2.5 | Procedure

The study was integrated into a course on internal medicine, led by the third and fourth authors. Across 3 weeks (October 23, 2023 to November 10, 2023), participants attended three sessions, 1 week apart. Participants attended the study in one of two separate classes, with an identical procedure other than the implementation of strategies (Figure 1).

Session I started with a prior knowledge test, and then participants followed an introductory lecture on the respiratory system. This lecture provided a brief introduction about the respiratory sounds and their general characteristics. At the end of this lecture, participants listened to the respiratory sounds through either blocked or interleaved study units. To maximize the authentic learning situation, the study units were played from classroom speakers. The name of the respiratory sounds was simultaneously visible on a white board.

In Session II, participants first listened to the respiratory sounds, using the same units and the order of exemplars and categories as in Session I, and then provided CLJs. After a short break (~5 min), during which participants were provided with pen and paper for the classification test, they listened to the novel exemplars played through classroom speakers. The presentation order of these assessment exemplars differed from the order used in the study units. Each sound was played for 10 s,

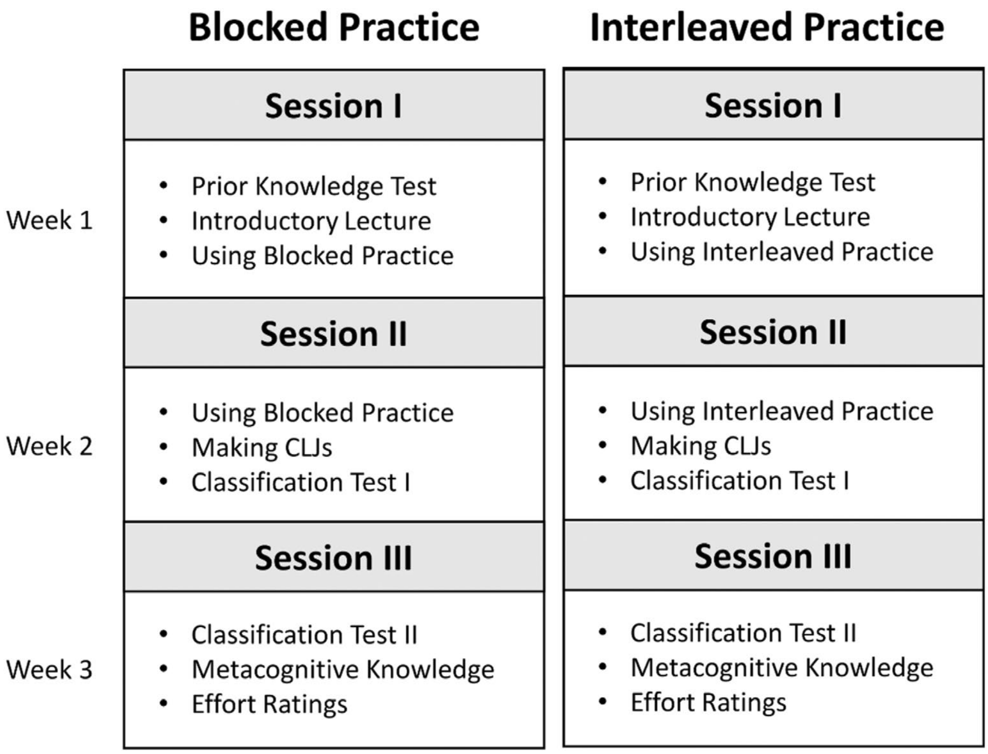

FIGURE 1 | Procedure of the study.

and examiners ensured that each student had sufficient time to complete their responses before the next sound was played.

During Session III, participants completed the delayed classification test[1] in a similar manner to the immediate classification test; note that the presentation order of the respiratory sounds differed from the order used in the immediate test. Subsequently, participants responded to learning scenarios and rated the anticipated effort demands of blocked and interleaved practice.

3 | Results

3.1 | Preliminary Analyses

Overall, 121 students (64%) participated in all three sessions ($n_{\text{Blocked}} = 67$; $n_{\text{Interleaved}} = 54$). The drop-out rate did not differ between classes, $\chi^2(1) = 1.43$, $p = 0.231$. Hence, we used the complete dataset for the subsequent analyses. Both groups were comparable with regard to their GPA, $t(119) = 0.028$, $p = 0.997$. Furthermore, there was no difference in their prior knowledge of the respiratory system between classes, $M_{\text{blocked}} = 4.71$, $SD_{\text{blocked}} = 2.10$; $M_{\text{interleaved}} = 4.46$, $SD_{\text{interleaved}} = 1.81$, $t(163.83) = 0.80$, $p = 0.42$. Thus, we excluded prior knowledge from the subsequent analyses.

3.2 | RQ 1: Auscultation Performance

Auscultation performance was analyzed in a 2×2 mixed ANOVA, with instructional method (blocked versus interleaved practice) and time (immediate and delayed tests) as independent variables. As shown in Figure 2, a significant main effect of instructional method was revealed, $F(1, 119) = 6.79$, $p = 0.010$, $\eta^2_p = 0.054$. Overall, interleaved practice ($M = 6.06$, $SD = 2.68$) led to better auscultation performance than blocked practice ($M = 5.05$, $SD = 2.24$). However, there was no main effect of time, $F(1, 119) = 0.52$, $p = 0.470$, $\eta^2_p = 0.004$, indicating that no substantial amount of forgetting occurred ($M_{\text{immediate}} = 5.58$, $SD_{\text{immediate}} = 2.31$; $M_{\text{delayed}} = 5.42$, $SD_{\text{delayed}} = 2.67$). The time × instructional method interaction was nonsignificant, $F(1, 119) = 0.16$, $p = 0.689$, $\eta^2_p = 0.001$.

3.3 | RQ 2: Relative Monitoring Accuracy

To calculate students' relative monitoring accuracy, we computed within-person gamma correlations (Nelson 1984) between students' CLJs and classification performance across different respiratory sounds. The classification test included two questions per abnormality, allowing students to either answer both questions correctly, both incorrectly, or partially correct (i.e., one correct, one incorrect). For each abnormality, we assigned three possible scores: 1 for both answers correct, 0 for both incorrect, and 0.5 for one correct answer. For eight participants, we were unable to calculate gamma correlation due to no variability in their CLJs or test scores; hence, they were omitted from the analysis.

An independent samples t test[2] (see Figure 3) revealed that relative monitoring accuracy did not significantly differ as a function of blocked ($M = 0.19$; $SD = 0.59$) and interleaved practice ($M = 0.37$; $SD = 0.54$), $t(110) = 1.59$, $p = 0.114$, $d = 0.30$.

As for perceived learning, we compared the magnitude of CLJs, indicating overall confidence in one's ability to recall respiratory

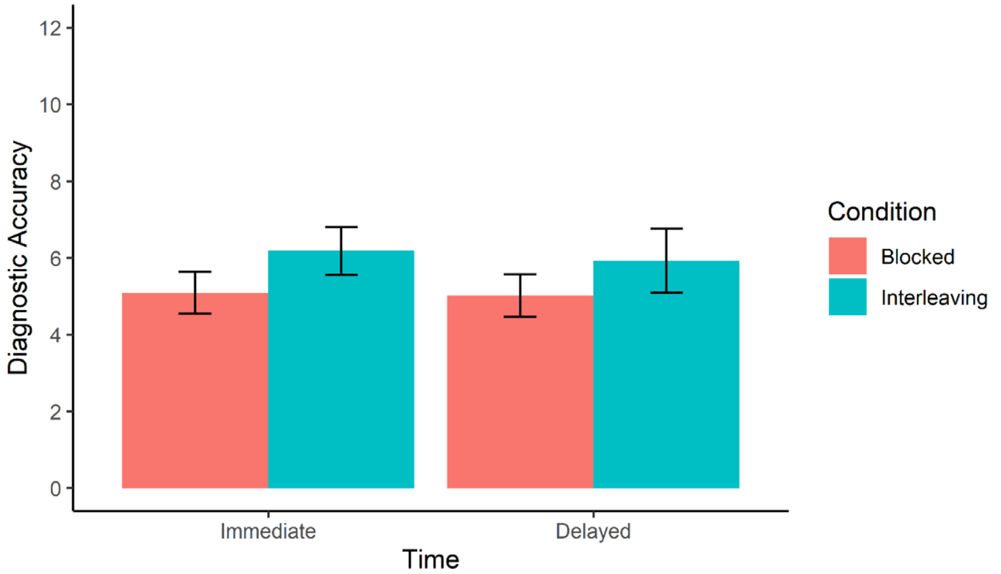

FIGURE 2 | Diagnostic accuracy as a function of blocked and interleaved practice. *Note:* Error bars represent the standard error.

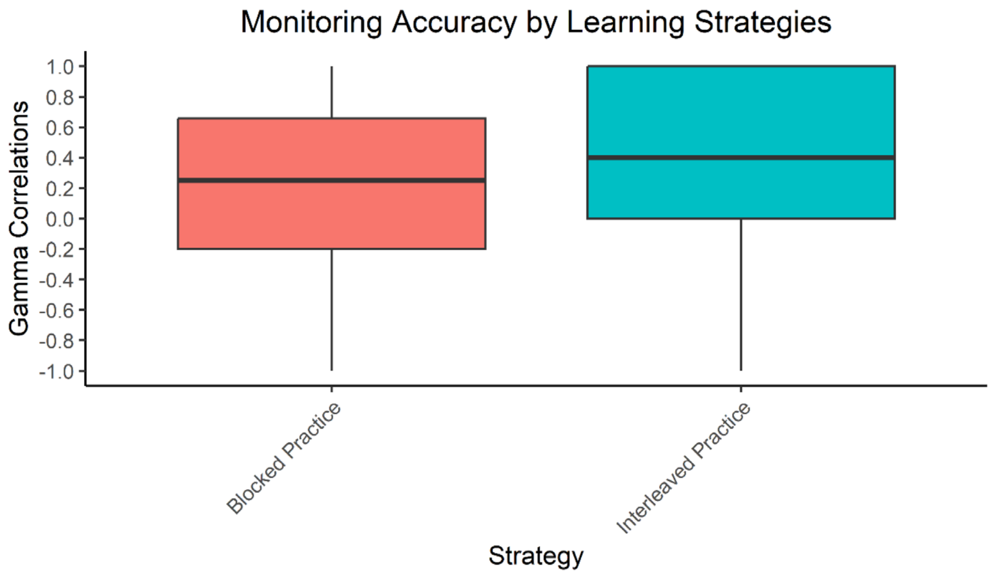

FIGURE 3 | Monitoring accuracy as a function of blocked and interleaved practice.

sounds as a function of blocked and interleaved practice. The results revealed that CLJs were higher for interleaved practice ($M = 66.90$) than for blocked practice ($M = 60.80$), $t\,(113) = 2.07$, $p = 0.040$, $d = 0.39$, suggesting that interleaving does not necessarily harm students' confidence in their auscultation performance.

3.4 | RQ 3: Metacognitive Knowledge and Effort Ratings

Overall, 56% believed that blocked practice would lead to better learning than interleaved practice (classroom blocked: 53% and classroom interleaved: 59%), while 42% believed the opposite (classroom blocked: 45% and classroom interleaved: 39%). The remaining participants (2%) indicated that both methods are equal in their effectiveness (classroom blocked: 1% and classroom interleaved 1%). Participants' responses to learning scenarios did not differ between classes, $\chi^2\,(1) = 0.41$, $p = 0.523$.

Finally, a paired-sample *t* test revealed that the perceived effort of interleaved practice ($M = 6.83$, SD = 1.94) was higher than that of blocked practice ($M = 5.12$, SD = 1.69), $t\,(120) = 5.94$, $p < 0.001$. For exploratory reasons, we calculated the correlations between effort ratings and CLJs. There was no association between perceived effort and CLJs, neither for interleaved practice, $r = 0.09$, $p = 0.256$, nor for blocked practice, $r = 0.03$, $p = 0.712$.

4 | Discussion

This study is the first to show that interleaved practice improves the learning of auditory materials within an authentic HPE setting with a large number of students. Supporting our first hypothesis, we found that interleaved practice resulted in higher diagnostic accuracy than blocked practice when nursing students learned to identify (ab)normal respiratory sounds. Notably, we observed this effect in an actual classroom led by teachers, showing the applicability of interleaved practice outside of experimentally controlled environments. An unexpected finding was that no significant forgetting occurred. Possibly, our repetitive approach might have flattened the forgetting curve, as students applied blocked and interleaved practice twice, one week apart. This repetition can be considered a form of spacing, which is known to promote long-term learning (Carpenter et al. 2022).

Regarding relative monitoring accuracy, we observed a small, numerical, but not statistically significant advantage of interleaved practice ($d = 0.30$, $p = 0.11$). Therefore, this finding should be approached with caution, and future research should aim to replicate these results with larger sample sizes to ensure sufficient statistical power. Regarding perceived learning, an unexpected finding was that CLJs were higher in the interleaved practice condition than in the blocked practice condition. This finding is striking because prior research showed that students' perceived learning is typically higher with blocked practice, while their actual learning is higher with interleaved practice, especially in learning tasks involving visual materials (Kirk-Johnson et al. 2019; Onan et al. 2022). Interestingly, in the context of the auditory modality, this metacognitive illusion seems to diminish; specifically, Abel (2023) also found no difference in CLJs when students learned auditory stimuli (bird sounds), using blocked and interleaved practice (also see limitations and future research). If this trend in CLJs remains in favor of interleaved practice, having students engage in both blocked and interleaved practice and reflect on their experiences may encourage them to use more interleaving.

Supporting our third hypothesis, we found that most students were unaware of the general learning benefits of interleaved practice, replicating prior research. Potentially, their preference for blocked practice might stem from how students interpret the effort demands of instructional methods: The higher their perceived or anticipate effort, the less effective they perceive the method to be (Kirk-Johnson et al. 2019; Onan et al. 2022). It is essential to correct such misinterpretations because they may cultivate students' erroneous beliefs about the efficacy of instructional methods and hinder effective study decisions (de Bruin et al. 2023; Onan et al. 2024). Together, these findings emphasize the need for targeted strategy trainings in nursing education—and HPE in general. Such trainings are essential because HPE students are faced with a continuous challenge of staying abreast of rapidly expanding medical knowledge, while experiencing time pressure (Nelson and Eliasz 2023). Against this background, supporting educators to implement desirable difficulties in their teaching can enhance auscultation training outcomes and prepare students to build long-term knowledge of (ab)normal respiratory sounds.

5 | Limitations

Auscultation is a complex skill that goes beyond recognizing (ab)normal sounds. Equally important is that students integrate the findings into a broader clinical context, considering patient history and physical examination. Due to the nature of this study, we primarily focused on the isolated recognition of (ab)normal sounds, omitting the comprehensive evaluation required in a real-world clinical setting.

A lack of randomization poses an inherent challenge in quasi-experimental studies. We implemented several safeguards to control for potential biases and establish comparability among students. Specifically, we observed no differences in prior knowledge and GPA across student groups. Furthermore, both classes shared the same instructors. Future studies, however, should replicate our findings through the implementation of true randomized controlled trials.

A third consideration is that the present study employed a between-subjects design, exposing students to either blocked or interleaved practice, rather than both. This design choice may affect students' CLJs, as they are unable to directly compare their visceral experiences. For example, Janssen et al. (2023) found that the difference in perceived learning between blocked and interleaved practice, when learning visual categories, diminished in a between-subjects design compared to a within-subjects design, yet remained significant in favor of blocked practice. However, as mentioned, Abel (2023) reported no difference when learning auditory stimuli in a between-subjects design. Future research should further investigate these metacognitive judgment dynamics, considering how study design and the modality of stimuli influence perceived learning.

6 | Conclusion

This study highlights the applicability of interleaved practice in authentic learning settings. Our findings further show that the learning benefits of interleaved practice extend to the auditory domain. It is notable that a significant proportion of students appear to be unaware of the learning benefits of interleaved practice. Educators and institutions should consider incorporating strategy trainings to familiarize students with the cognitive and metacognitive advantages of desirably difficult instructional methods and learning strategies.

Author Contributions

Erdem Onan: conceptualization, methodology, formal analysis, visualization, writing – original draft, data curation, investigation. **Arif Onan:** conceptualization, methodology, writing – review and editing, project administration, data curation, investigation, resources. **Ezgi Ozgun:** investigation, formal analysis, project administration, data curation, conceptualization, writing – review and editing. **Semra Gundogdu:** conceptualization, investigation, methodology, writing – review and editing, resources. **Hicran Bektas:** conceptualization, methodology, investigation, writing – review and editing, resources, validation. **Anique B. H. de Bruin:** conceptualization, methodology, funding acquisition, writing – review and editing, supervision.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] Due to an experimenter error, we did not collect delayed judgments. Therefore, we deviated from the preregistration form for the second research question, as we could only examine the influence of blocked and interleaved practice on monitoring accuracy, but not the timing of judgments.

[2] Since delayed CLJs were not obtained, gamma correlations were further analyzed using an independent samples *t* test instead of the preregistered two-way mixed analysis of variance.

References

Abel, R. 2023. "Interleaving Effects in Blindfolded Perceptual Learning Across Various Sensory Modalities." *Cognitive Science* 47, no. 4: e13270. https://doi.org/10.1111/cogs.13270.

Birnbaum, M. S., N. Kornell, E. L. Bjork, and R. A. Bjork. 2013. "Why Interleaving Enhances Inductive Learning: The Roles of Discrimination and Retrieval." *Memory & Cognition* 41, no. 3: 392–402. https://doi.org/10.3758/s13421-012-0272-7.

Bjork, E. L., and R. A. Bjork. 2011. "Making Things Hard on Yourself, but in a Good Way: Creating Desirable Difficulties to Enhance Learning." In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, edited by M. A. Gernsbacher, R. W. Pew, L. M. Hough, and J. R. Pomerantz, 56–64. Worth Publishers.

Bjork, R. A., J. Dunlosky, and N. Kornell. 2013. "Self-Regulated Learning: Beliefs, Techniques, and Illusions." *Annual Review of Psychology* 64: 417–444. https://doi.org/10.1146/annurev-psych-113011-143823.

Brunmair, M., and T. Richter. 2019. "Similarity Matters: A Meta-Analysis of Interleaved Learning and Its Moderators." *Psychological Bulletin* 145, no. 11: 1029–1052. https://doi.org/10.1037/bul0000209.

Carpenter, S. K., S. C. Pan, and A. C. Butler. 2022. "The Science of Effective Learning With Spacing and Retrieval Practice." *Nature Reviews Psychology* 1, no. 9: 496–511. https://doi.org/10.1038/s44159-022-00089-1.

Carvalho, P. F., and R. L. Goldstone. 2014. "Putting Category Learning in Order: Category Structure and Temporal Arrangement Affect the Benefit of Interleaved Over Blocked Study." *Memory & Cognition* 42: 481–495. http://dx.doi.org/10.3758/s13421-013-0371-0.

Cecilio-Fernandes, D., R. Patel, and J. Sandars. 2023. "Using Insights From Cognitive Science for the Teaching of Clinical Skills: AMEE Guide No. 155." *Medical Teacher* 45: 1214–1223.

Chen, R., L. Grierson, and G. Norman. 2015. "Manipulation of Cognitive Load Variables and Impact on Auscultation Test Performance." *Advances in Health Sciences Education* 20, no. 4: 935–952. https://doi.org/10.1007/s10459-014-9573-x.

de Bruin, A. B. H., F. Biwer, L. Hui, E. Onan, L. David, and W. Wiradhany. 2023. "Worth the Effort: The Start and Stick to Desirable Difficulties (S2D2) Framework." *Educational Psychology Review* 35, no. 2: 41. https://doi.org/10.1007/s10648-023-09766-w.

Eglington, L. G., and S. H. Kang. 2017. "Interleaved Presentation Benefits Science Category Learning." *Journal of Applied Research in Memory and Cognition* 6, no. 4: 475–485.

Faul, F., E. Erdfelder, A. Buchner, and A. G. Lang. 2009. "Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses." *Behavior Research Methods* 41, no. 4: 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Firth, J., I. Rivers, and J. Boyle. 2021. "A Systematic Review of Interleaving as a Concept Learning Strategy." *Review of Education* 9, no. 2: 642–684. https://doi.org/10.1002/rev3.3266.

Foster, N. L., M. L. Mueller, C. Was, K. A. Rawson, and J. Dunlosky. 2019. "Why Does Interleaving Improve Math Learning? The Contributions of Discriminative Contrast and Distributed Practice." *Memory & Cognition* 47, no. 6: 1088–1101. https://doi.org/10.3758/s13421-019-00918-4.

Foster, N. L., M. L. Mueller, J. Dunlosky, and L. Finkenthal. 2023. "What Is the Impact of Interleaving Practice and Delaying Judgments on the Accuracy of Category-Learning Judgments?" *Journal of Experimental Psychology: Applied* 29, no. 2: 374–385.

Hafke-Dys, H., A. Bręborowicz, P. Kleka, J. Kociński, and A. Biniakowski. 2019. "The Accuracy of Lung Auscultation in the Practice of Physicians and Medical Students." *PLoS One* 14, no. 8: e0220606. https://doi.org/10.1371/journal.pone.0220606.

Hatala, R. M., L. R. Brooks, and G. R. Norman. 2003. "Practice Makes Perfect: The Critical Role of Mixed Practice in the Acquisition of ECG Interpretation Skills." *Advances in Health Sciences Education* 8, no. 1: 17–26. https://doi.org/10.1023/A:1022687404380.

Janssen, E. M., T. van Gog, L. van de Groep, et al. 2023. "The Role of Mental Effort in Students' Perceptions of the Effectiveness of Interleaved and Blocked Study Strategies and Their Willingness to Use Them." *Educational Psychology Review* 35, no. 3: 85.

Kämmer, J. E., W. E. Hautz, and M. März. 2020. "Self-Monitoring Accuracy Does Not Increase Throughout Undergraduate Medical Education." *Medical Education* 54, no. 4: 320–327. https://doi.org/10.1111/medu.14057.

Kirk-Johnson, A., B. M. Galla, and S. H. Fraundorf. 2019. "Perceiving Effort as Poor Learning: The Misinterpreted-Effort Hypothesis of How Experienced Effort and Perceived Learning Relate to Study Strategy Choice." *Cognitive Psychology* 115: 101237. https://doi.org/10.1016/j.cogpsych.2019.101237.

Kornell, N., and R. A. Bjork. 2008. "Learning Concepts and Categories: Is Spacing the "Enemy of Induction"?" *Psychological Science* 19, no. 6: 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x.

Monteiro, S. M., and G. Norman. 2013. "Diagnostic Reasoning: Where We've Been, Where We're Going." *Teaching and Learning in Medicine* 25: S26–S32. https://doi.org/10.1080/10401334.2013.842911.

Morehead, K., M. G. Rhodes, and S. DeLozier. 2016. "Instructor and Student Knowledge of Study Strategies." *Memory* 24, no. 2: 257–271. https://doi.org/10.1080/09658211.2014.1001992.

Moriki, D., D. Koumpagioti, M. Kalogiannis, et al. 2023. "Physicians' Ability to Recognize Adventitious Lung Sounds." *Pediatric Pulmonology* 58, no. 3: 866–870. https://doi.org/10.1002/ppul.26266.

Nelson, A., and K. L. Eliasz. 2023. "Desirable Difficulty: Theory and Application of Intentionally Challenging Learning." *Medical Education* 57, no. 2: 123–130. https://doi.org/10.1111/medu.14916.

Nelson, T. O. 1984. "A Comparison of Current Measures of the Accuracy of Feeling-Of-Knowing Predictions." *Psychological Bulletin* 95, no. 1: 109–133. https://doi.org/10.1037/0033-2909.95.1.109.

Onan, E., F. Biwer, W. Wiradhany, and A. B. de Bruin. 2024. "Instruction Meets Experience: Using Theory-and Experience-Based Methods to Promote the Use of Desirable Difficulties." *Learning and Instruction* 93: 101942.

Onan, E., W. Wiradhany, F. Biwer, E. M. Janssen, and A. B. H. de Bruin. 2022. "Growing out of the Experience: How Subjective Experiences of Effort and Learning Influence the Use of Interleaved Practice." *Educational Psychology Review* 34, no. 4: 2451–2484. https://doi.org/10.1007/s10648-022-09692-3.

Pan, S. C., S. R. Flores, M. E. Kaku, and W. H. E. Lai. 2025. "Interleaved Practice Enhances Grammar Skill Learning for Similar and Dissimilar Tenses in Romance Languages." *Learning and Instruction* 95: 102045.

Rohrer, D., R. F. Dedrick, M. K. Hartwig, and C. N. Cheung. 2020. "A Randomized Controlled Trial of Interleaved Mathematics Practice." *Journal of Educational Psychology* 112, no. 1: 40–52.

Rozenshtein, A., G. D. Pearson, S. X. Yan, A. Z. Liu, and D. Toy. 2016. "Effect of Massed Versus Interleaved Teaching Method on Performance of Students in Radiology." *Journal of American College of Radiology* 13, no. 8: 979–984.

Samani, J., and S. C. Pan. 2021. "Interleaved Practice Enhances Memory and Problem-Solving Ability in Undergraduate Physics. npj Science of." *Learning* 6, no. 1: 32.

Sana, F., and V. X. Yan. 2022. "Interleaving Retrieval Practice Promotes Science Learning." *Psychological Science* 33, no. 5: 782–788.

Schmidt, H. G., and S. Mamede. 2020. "How Cognitive Psychology Changed the Face of Medical Education Research." *Advances in Health Sciences Education* 25: 1025–1043. https://doi.org/10.1007/s10459-020-10011-0.

Schraw, G. 2009. "A Conceptual Analysis of Five Measures of Metacognitive Monitoring." *Metacognition and Learning* 4: 33–45.

Schweppe, J., A. Lenk-Blochowitz, M. Pucher, and A. Ketzer-Nöltge. 2025. "Interleaved Practice in Foreign Language Grammar Learning: A Field Study." *Journal of Educational Psychology* https://psycnet.apa.org/record/2025-63706-001.

Thompson, C. P., and M. A. Hughes. 2023. "The Effectiveness of Spaced Learning, Interleaving, and Retrieval Practice in Radiology Education: A Systematic Review." *Journal of the American College of Radiology* 20: 1092–1101.

Williams, B., M. Boyle, and P. O'Meara. 2009. "Can Undergraduate Paramedic Students Accurately Identify Lung Sounds?" *Emergency Medicine Journal* 26, no. 8: 580–582. https://doi.org/10.1136/emj.2008.058552.

Wong, S. S. H., A. C. M. Low, S. H. Kang, and S. W. H. Lim. 2020. "Learning Music Composers' Styles: To Block or to Interleave?" *Journal of Research in Music Education* 68, no. 2: 156–174.

Wong, S. S. H., S. Chen, and S. W. H. Lim. 2021. "Learning Melodic Musical Intervals: To Block or to Interleave?" *Psychology of Music* 49, no. 4: 1027–1046.

Yan, V. X., E. L. Bjork, and R. A. Bjork. 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, no. 7: 918–933. https://doi.org/10.1037/xge0000177.

Appendix A

Example Questions From the Prior Knowledge Test

Please answer the following questions* to the best of your knowledge.

1. During expiration, the diagram ...
   a. Rises by contraction.
   b. Contracts and flattens.
   c. Rises by relaxation.
   d. Relaxes and flattens.
   e. No opinion.

2. Thoracic volume is ...
   a. The maximum volume reached by the lungs when breathing (inspiration).
   b. The volume of air entering or leaving the lungs during quiet breathing.
   c. Volume of air exhaled after a deep breath (inspiration).
   d. The volume of air remaining in the lungs after a deep exhalation (expiration).
   e. No opinion.

3. During breathing (inspiration), ...
   a. Diaphragm and intercostal muscles contract.
   b. The diaphragm and intercostal muscles relax.
   c. The intercostal muscles contract as the diaphragm relaxes.
   d. Intercostal muscles relax as the diaphragm contracts.
   e. No opinion.

4. During exhalation (expiration), ...
   a. Diaphragm and intercostal muscles contract.
   b. Diaphragm and intercostal muscles relax.
   c. The intercostal muscles contract as the diaphragm relaxes.
   d. Intercostal muscles relax as the diaphragm contracts.
   e. No opinion.

5. Tidal volume is ...
   a. The maximum volume reached by the lungs when breathing (inspiration).
   b. The volume of air entering or leaving the lungs during quiet breathing.
   c. Volume of air exhaled after a deep breath (inspiration).
   d. The volume of air remaining in the lungs after a deep exhalation (expiration).
   e. No opinion.

* Questions are translated from Turkish.