Scaling Interleaved Mathematics Practice

Matlen, B. J.

Interleaving   learning engineering   Mathematics Education

Interleaved practice involves mixing different problem types during study rather than blocking them by category. It has been identified as a robust learning strategy that enhances both discrimination learning and long-term retention. Despite compelling laboratory evidence and successful classroom efficacy trials demonstrating large learning gains, interleaved practice is relatively uncommon in mainstream mathematics curricula. This disconnect between research and practice exemplifies a persistent challenge: learning principles validated in controlled settings often fail to make their way into classrooms. The present contribution reports on preliminary results of a cluster randomized study testing the efficacy of interleaved mathematics in a large, U.S. based sample. It also briefly discusses the inherent complexity in translating learning principles to authentic contexts.

Introduction

One product of cognitive science research are principles of learning - teaching and study strategies that have been shown to enhance learning in a range of settings and ages in both classroom and lab tests. One such principle is interleaved practice,

which involves mixing problems of different types during study rather than blocking them by category, and distributing problems of the same type over time. Interleaved practice has been found to improve both discrimination learning and long-term retention and a recent cluster randomized trial indicated substantial learning gains for interleaved practice over blocked practice in middle school mathematics (Rohrer et al. 2020). Still, interleaved practice remains relatively uncommon in mainstream mathematics curricula (Rohrer et al. 2020).

In the present study (ongoing), we are testing the efficacy of interleaved practice relative to blocked practice in a large, cluster randomized study. The study is designed to both replicate and extend the study conducted by Rohrer and colleagues (2020). Specifically, like Rohrer et al, the present study tests the efficacy of interleaving in middle school mathematics classrooms, randomly assigning classrooms within teachers to either an interleaved or blocked control group. However, unlike Rohrer et al, the study expands the sample beyond a single district and state to multiple districts and states in a U.S. nationally representative sample. The study also expands the intervention from a half year to a full year and implements interleaved practice on a digital platform - ASSISTments - which increases the intervention's ability to be implemented at scale in authentic contexts.

Learning principles like interleaving are typically framed as straightforward strategies, but the implementation of such strategies can be challenging as they require many contextual decisions that are not directly informed by past research (e.g., how many items are optimal within an interleaved assignment? Can students work in pairs? etc.). To instantiate the intervention in the present work we engaged in a structured design process guided by a logic model. The logic model is a structured description of the intervention informed by existing research, theory, and practical constraints of the use context, that can be revised and iterated upon in formative tests. Table 1 displays this working model.

Table 1. Logic Model of Interleaved Practice in the present study
The logic model of interleaved practice includes intervention components informed by prior research, materials and implementation features that define fidelity, and their connection to the interleaving intervention that this study is designed to replicate.

Intervention Component	Study Materials	Instantiation in Practice	Relation to past efficacy study
Interleaving problems of different types within assignments to support discrimination learning	Assignments created in ASSISTments that each consist of grade level appropriate problems that are interleaved	Students complete assignments that contain topics they've had prior instruction.	Core feature, held constant
		Students attempt each problem independently, writing out the corresponding solution steps	Core feature, held constant
		Students complete problems in ASSISTments	Design feature, varied
		Students complete assignments in class	Design feature, held constant

Revision of incorrect solutions after each assignment to support comprehension	Visual tool for showing problems and corresponding solutions, such as a whiteboard or screen	Teachers discuss problems with solutions available visually	Core feature, held constant
		Students instructed to correct both their work and answers	Core feature, held constant
		Teachers collect and verify students corrected their errors	Design feature, varied
Spaced practice of problems of same type to support long-term retention	Ordered assignments that consist of problems of the same type spaced across the assignments	Assignments are completed in the order of the study procedure, which are distributed over time so that long-term retention benefits from the spacing effect	Core feature, held constant

One value of this logic model-based approach for replication is that it makes explicit the relationship between implementation features that are considered "core" to the intervention and therefore held constant between this and the prior study, and design features that are not theoretically critical and that can therefore differ from past work to best align with the context. Early formative testing identified several design features that were varied to maximize feasibility and fidelity for this sample and context. This logic model-based approach to learning design shares many similarities to other learning engineering approaches such as the Nested Learning Engineering cycle (e.g., Craig et al, 2025) that involve understanding the problem in context, developing, testing, and revising the working solution.

Method and Results

This work reports on preliminary results of the ongoing study. The present sample (representing two of three planned cohorts) includes 813 sixth and seventh grade students across 51 classes, 29 teachers, 15 schools, and 8 districts. Random assignment occurred at the classroom level, with classes blocked within teachers. The post-assessment, developed collaboratively with the original efficacy study author, consisted of 12 isomorphic items to those in the review assignments, administered four weeks after the final intervention to measure robust learning. Hierarchical linear models accounting for students nested within classes revealed that students receiving interleaved practice (M = 0.09, SD = 1.03) outperformed blocked practice students (M = -0.08, SD = 0.96) on the delayed post-assessment, but this difference was just shy of statistical significance at $p < .05$, $F(1, 41.81) = 3.75$, $p = .05$, Hedges' $g = 0.18$. This work shows promise for the potential of interleaving to be effectively applied in authentic education contexts at a broad scale.

Acknowledgments

References

Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C. N. (2020). A randomized controlled trial of interleaved mathematics practice. Journal of Educational Psychology, 112(1), 40.

Rohrer, D., Dedrick, R. F., & Hartwig, M. K. (2020). The scarcity of interleaved practice in mathematics textbooks. Educational Psychology Review, 32(3), 873-883.

Craig, S. D., Avancha, K., Malhotra, P., C., J., Verma, V., Likamwa, R., Gary, K., Spain, R., & Goldberg, B. (2025). Using a Nested Learning Engineering Methodology to Develop a Team Dynamic Measurement Framework for a Virtual Training Environment. In International Consortium for Innovation and Collaboration in Learning Engineering (ICICLE) 2024 Conference Proceedings: Solving for Complexity at Scale (pp. 115-132). https://doi.org/10.59668/2109.21735

★★★★★