Journal of Educational Psychology

Deliberately Making and Correcting Errors in Mathematical Problem-Solving Practice Improves Procedural Transfer to More Complex Problems

Janson Boon Khang Yap and Sarah Shi Hui Wong

Online First Publication, March 14, 2024. https://dx.doi.org/10.1037/edu0000850

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

https://doi.org/10.1037/edu0000850

Deliberately Making and Correcting Errors in Mathematical Problem-Solving Practice Improves Procedural Transfer to More Complex Problems

Janson Boon Khang Yap[1] and Sarah Shi Hui Wong[1, 2]
[1] Department of Psychology, Faculty of Arts & Social Sciences, National University of Singapore
[2] Division of Social Sciences, Yale-NUS College

How can students effectively learn and transfer mathematical procedures to solve new problems? Here, we tested the effects of deliberately committing and correcting errors during mathematical problem-solving practice on transfer of the learned procedures. In two experiments, university students were instructed on mathematical algorithms (synthetic division and matrix multiplication) and solved practice problems during open-book study. Learners were then tested on flexibly adapting the algorithms to solve novel problems that were structurally more complex or embedded in "real-life" scenarios (i.e., mathematical word problems). Deliberately committing and correcting procedural errors during problem-solving practice yielded better transfer than errorless repeated practice (Experiment 1) or studying incorrect worked examples by finding, explaining, and correcting the errors that one's peers had made (Experiment 2). Yet, most learners failed to accurately predict or recognize the advantage of deliberate erring even after the test, instead misjudging this technique as less effective. This suggests that experiencing the benefit of deliberate erring is insufficient to dispel learners' metacognitive illusion that generating errors is not helpful for their learning. Overall, our results point to the critical role of first-hand errors in mathematical learning. Relative to avoiding errors or even studying others' errors and juxtaposing them with the correct solutions, guiding learners to deliberately commit and correct their own errors after instruction improves mathematical problem solving and transfer.

Educational Impact and Implications Statement

Transfer of learning lies at the heart of education, but is often difficult to achieve. Here, we show in the domain of mathematics that deliberately committing and correcting procedural errors during problem-solving practice enhances students' transfer of the learned procedures to solve novel, more challenging problems. Deliberate erring was not only more effective than errorless repeated practice, but also finding, explaining, and correcting others' errors in incorrect worked examples. These results expand our repertoire of approaches to harness the power of errors for improving mathematical problem solving and transfer in education.

Keywords: learning from errors, incorrect worked examples, mathematical problem solving, procedural knowledge, transfer of learning

Across many knowledge domains such as mathematics (Hiebert & Lefevre, 1986; Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001), programming (McGill & Volet, 1997), hypothesis testing (Howe et al., 2000; Wong et al., 2019), and musical performance (Palmer & Meyer, 2000), students must learn not only foundational concepts or "know that," but also the associated procedures or "know how" (Anderson, 1976, 1982; Booth, 2011; Knowlton et al., 2017). In mathematical problem solving, for instance, learners must develop both conceptual understanding of mathematical principles and procedural skill in executing action sequences to solve novel problems, which are often acquired through practice (Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001, 2015). Crucially, for learning to be meaningful, learners must be able to transfer their knowledge and skills to new contexts (Day & Goldstone, 2012; Mayer, 2002). Indeed, transfer of learning is widely regarded as an ultimate aim of education (McKeough et al., 1995; Perkins & Salomon, 1992), but is difficult to attain (Detterman, 1993; Renkl et al., 1996).

How can students effectively learn and transfer mathematical concepts and procedures to solve novel problems? Historically, behaviorist principles have suggested that exposure to errors may reinforce incorrect responses and should thus be avoided during learning

Tobias Richter served as action editor.
Sarah Shi Hui Wong https://orcid.org/0000-0003-4243-212X
The materials are available by emailing the corresponding author. The data and analysis code for this study are available at https://osf.io/us9jb/.

Janson Boon Khang Yap served as lead for data curation, formal analysis, and investigation. Sarah Shi Hui Wong served as lead for resources, supervision, and writing–review and editing and served in a supporting role for formal analysis. Janson Boon Khang Yap and Sarah Shi Hui Wong contributed equally to conceptualization, writing–original draft, and methodology.

Correspondence concerning this article should be addressed to Sarah Shi Hui Wong, Division of Social Sciences, Yale-NUS College, 28 College Avenue West, #01-501, Singapore 138533, Singapore. Email: psywshs@nus.edu.sg

(Skinner, 1958). But contrary to this notion, a growing body of research has shown that engaging with errors can improve learning (for reviews, see Metcalfe, 2017; Wong & Lim, 2019), including in the domain of mathematics (e.g., Booth et al., 2013; Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007; Siegler, 2002). Here, we test the extent that the novel approach of guiding students to deliberately commit and correct errors during mathematical problem-solving practice enhances learning and transfer.

Extant Approaches to Errors in Mathematical Learning

The traditional view of errors as aversive events to be avoided in learning (Ausubel, 1968; Bandura, 1986; Skinner, 1958) has become deeply entrenched in educational systems (Metcalfe, 2017). Notwithstanding variation across cultures (Stevenson & Stigler, 1992), some studies of teachers' error management behavior in classrooms have observed that students' errors tend to either be ignored or redirected to another student to produce the correct answer (Santagata, 2005; Tulis, 2013). Consequently, the student who erred loses the opportunity to correct their error (i.e., the "Bermuda triangle of error correction"; Oser & Spychiger, 2005). Teachers have also been found to make disparaging responses to their students' errors more often in mathematics classes than other domains such as German or economics (Tulis, 2013). Unsurprisingly, when students perceive the "error climate" in the classroom as negative (e.g., teachers' intolerance toward student errors), they may react to errors in less adaptive ways that harm their academic engagement, self-regulation, and learning (Soncini et al., 2022; Steuer et al., 2013; see also Pekrun et al., 2002). For instance, students may view errors as self-threatening events that evoke shame and frustration, especially when attributed to one's lack of ability (Brodbeck et al., 1993; Pekrun, 2006; Weiner, 1985).

Yet, attempting to avoid errors is futile since they are inevitable in most learning situations. Indeed, effective learning often involves "desirable difficulties" or challenging learning conditions that may lead to more errors as byproducts of meaningful processing, but actually yield deeper and more durable learning (Bjork, 1994). In line with this idea, research has revealed that errors can in fact benefit learning, especially when accompanied by corrective feedback (Metcalfe, 2017; Wong & Lim, 2019). For instance, in problem solving before instruction (PS-I; for reviews, see Kapur, 2016; Loibl et al., 2017; Sinha & Kapur, 2021), errors are elicited by having learners attempt to solve novel, challenging problems before receiving instruction. PS-I studies have found that this learning approach improves conceptual knowledge and transfer more than problem solving after instruction in domains such as mathematics, physics, and statistics, although both approaches typically do not differ in their effects on procedural knowledge (e.g., DeCaro & Rittle-Johnson, 2012; Kapur & Bielaczyc, 2012; Schwartz et al., 2011; Schwartz & Martin, 2004).

Alternatively, learners can engage with errors during or after instruction on to-be-learned information. For instance, research on incorrect worked examples has tested the approach of presenting learners with examples of incorrect solutions that they are told contain common errors or misconceptions, which they must then identify and explain (for reviews, see Barbieri et al., 2023; Booth et al., 2015). Presumably, studying incorrect examples prompts learners to think deeply about the correct concepts, revise their faulty knowledge structures, and fine-tune overly general problem-solving rules (Ohlsson, 1996a, 1996b; VanLehn, 1999). Consequently, learners are less likely to use incorrect procedures in the future (Siegler, 2002). Indeed, much work has shown that studying incorrect examples—either alone or with correct examples—improves conceptual understanding, procedural skill, and transfer in mathematics, relative to studying correct examples only or a problem-solving control (e.g., Adams et al., 2014; Barbieri & Booth, 2020; Booth et al., 2013; Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007; Heemsoth & Heinze, 2014; McLaren et al., 2015; Siegler, 2002; cf. Barbieri et al., 2023; Pillai et al., 2020).

Whereas abundant research has investigated PS-I and incorrect worked examples, errors can be approached in other ways to enhance mathematical learning and problem solving. As delineated in Wong and Lim's (2019) Prevention–Permission–Promotion (3P) framework, errors in learning can be observed, allowed, induced, or guided when not avoided. For instance, incorrect worked examples involve observing others' errors without making them oneself, whereas PS-I involves inducing errors by adding challenge to the learning task and withholding information needed to solve it. Alternatively, an oft-neglected approach is to guide errors by leading students to err in a systematic and strategic way. Because guided errors are explicitly incorporated as expected parts of the learning design and process, they may offer a means to offset or eliminate the ego threats associated with other errorful approaches (Lorenzet et al., 2005). In particular, students can be guided to deliberately commit and correct errors during problem-solving practice, even if they have already been instructed on the correct concepts and procedures.

Learning From Deliberate Errors

Intuitively, deliberately generating errors when one already knows the correct answers may seem to be redundant or, worse still, harm learning by producing interference. Defying such intuitions, though, deliberately committing and correcting errors in low-stakes contexts has been found to enhance learning in a phenomenon known as the *derring effect* (Wong & Lim, 2022a, 2022b). In the first demonstration of the derring effect (Wong & Lim, 2022b), learners studied scientific term-definition concepts (e.g., "Cocktail party effect is the selective enhancement of attention to filter out distractions") via open-book study by intentionally generating conceptually incorrect definitions with or without correction, or copying and underlining them. The key finding was that deliberate erring produced better memory for the concepts on a cued-recall test (e.g., "What is the cocktail party effect?") than errorless copying, with an additional benefit from correcting one's errors. Moreover, learners' deliberate errors rarely returned at test, constituting less than 5% of their incorrect test responses. Thus, deliberate erring improved learning of conceptual knowledge while incurring little interference cost.

Subsequent studies have extended the derring effect to meaningful learning outcomes such as applying conceptual knowledge from scientific expository texts to analyze related news events (Wong & Lim, 2022a), and even far transfer of learned concepts to different knowledge domains on short-answer inferential questions (Wong, 2023). Moreover, deliberate erring has been found to outperform not only copying, but also other errorless learning techniques such as concept-mapping and generating elaborations or examples of to-be-learned concepts (Wong & Lim, 2022a, 2022b), and even other errorful approaches such as spotting and correcting others' conceptual errors (Wong, 2023).

Some theoretical accounts have been proposed to explain the derring effect, although more research is needed to test the precise mechanisms. For instance, generating errors may enhance encoding of their subsequent correction (Hays et al., 2013; Kornell et al., 2009; Potts et al., 2019; Potts & Shanks, 2014), such that correcting one's deliberate errors yields better learning than leaving them uncorrected (Wong & Lim, 2022b). Unlike errors induced through guessing or PS-I, this enhanced encoding is not likely due to surprise or curiosity (Potts et al., 2019; Potts & Shanks, 2014), since learners already know the correct answers when deliberately erring. Rather, intentionally committing errors may draw learners' attention to the target response during correction for a unique and enduring episodic memory trace (e.g., Metcalfe & Huelser, 2020). For instance, when searching their prior knowledge to generate responses that are definitively wrong, learners may gain awareness of gaps in their knowledge and thus more actively process subsequent corrective information to repair their mental models (Wong, 2023; see also Chi, 2000; Loibl et al., 2017; VanLehn, 1999).

In addition, deliberate erring may induce mental processes that are not typically invited by the learning material but that enhance the quality of its processing (Wong, 2023; Wong & Lim, 2022a; see also McDaniel & Butler, 2011; McDaniel & Einstein, 1989 for discussions of material-appropriate processing). As compared to studying correct information only, exploring incorrect responses to a question may ironically weaken and cull those unproductive responses, thereby facilitating future retrieval of the correct answer (Kornell et al., 2009). Moreover, contrasting one's errors with their correction may highlight a concept's diagnostic properties, in turn aiding learners' detection of knowledge gaps and deeper understanding of inappropriate responses to be avoided for better transfer (Corral & Carpenter, 2020; Heemsoth & Heinze, 2014; Loibl et al., 2017; Loibl & Rummel, 2014).

Notwithstanding the promise of deliberate erring, an unresolved question remains: To what extent does deliberate erring improve procedural skill when learning how to apply solution procedures to solve novel problems, such as those in mathematics? Thus far, evidence for the derring effect has centered on the benefits of making deliberate conceptual errors (e.g., generating incorrect concepts) for conceptual knowledge and transfer (Wong, 2023; Wong & Lim, 2022a, 2022b). It remains unknown if deliberately committing and correcting procedural errors (e.g., executing incorrect procedures) improves the acquisition and transfer of learned procedures.

Some evidence from the limited research on guided errors hints at this possibility. For instance, when learning to use a software such as Microsoft PowerPoint, trainees have been found to benefit more from guided error training with click-by-click instructions that lead them into making and correcting trainer-identified common errors, relative to error-free training with errorless click-by-click instructions (Lorenzet et al., 2005). Notably, guided error training led to better performance on not only skill reproduction tasks (i.e., reproducing skills learned during training), but also skill generalization tasks (i.e., transferring skills to similar tasks that had not been explicitly covered in training). By extension, guiding learners to deliberately commit and correct self-generated procedural errors may enhance their transfer of the learned procedures to solve novel mathematical problems.

The Present Study

In the present pair of experiments, we tested the effects of deliberately committing and correcting errors during mathematical problem-solving practice on transfer of the learned procedures. University students were instructed on mathematical algorithms and practiced using them to solve practice problems. Experiment 1 compared deliberate erring against errorless repeated practice—learners practiced the algorithms by either deliberately committing and correcting procedural errors or correctly solving each practice problem twice. To test whether the derring effect is contingent on personally erring oneself, Experiment 2 pitted deliberate erring against the errorful approach of studying incorrect worked examples, whereby learners spotted, explained, and corrected their peers' deliberate procedural errors.

After their practice, learners underwent a transfer test that required applying the algorithms to solve novel problems that were structurally more complex. The test questions demanded adapting the learned procedures to execute additional steps that had not been explicitly required in the practice problems. In addition, some transfer questions were embedded in scenarios that described a "real-life" problem context (i.e., mathematical word problems). Such word problems are often challenging because they require learners not only to apply learned procedures, but also to construct a coherent mental representation of the problem situation to develop a corresponding solution plan (Kintsch & Greeno, 1985; Mayer & Hegarty, 1996; Pongsakdi et al., 2020; for a recent review, see Verschaffel et al., 2020).

Across both experiments, we further assessed students' metacognitive awareness of the learning methods' effects on their performance. Accurate metacognitive knowledge is vital for effective self-regulated learning because it causally influences learners' study choices (Metcalfe & Finn, 2008). Curiously, though, students are often unaware of the benefits of errors for their learning (Huelser & Metcalfe, 2012; Pan et al., 2020; Wong, 2023; Wong & Lim, 2022a, 2022b; Yang et al., 2017). Thus, in both experiments, we used a within-subjects design to examine the relative actual versus predicted effectiveness of the learning methods for each student. After experiencing the learning methods during their problem-solving practice, learners made a prediction of their test performance. In addition, learners rated the effectiveness of the learning methods after completing the transfer test.

Experiment 1

Experiment 1 provided the first test of the derring effect in the learning and transfer of mathematical problem-solving procedures. For generalizability purposes, we used two mathematical algorithms that similarly require sequential procedures involving basic arithmetic operations such as addition and subtraction: *synthetic division* (division of polynomials) and *matrix multiplication* (multiplication of matrices). Learners in our study were not expected to be highly familiar with these algorithms although they were proficient in basic arithmetic, as ascertained via measures of their prior knowledge of the algorithms and basic mathematical ability, respectively.

A within-subjects design was used whereby for each algorithm, learners received an instructional lesson then solved practice problems using one of two learning methods. Specifically, learners either deliberately committed a procedural error (i.e., incorrectly applied the action sequences for the algorithm) in their first solution attempt for each practice problem then corrected their error by accurately solving the problem in their second attempt (*procedure-error* method) or correctly solved each practice problem twice (*repeated-*

practice method). Thus, both learning methods involved practicing each problem twice—either correctly both times or once incorrectly then the second time correctly. To simulate naturalistic learning contexts in which students are typically able to refer to their textbooks or notes during practice, all learners engaged in open-book practice in which they were allowed to refer to printouts detailing the steps for each algorithm. This also ensured that learners' errors during practice were deliberately committed with knowledge of the correct procedures.

After completing the practice problems, learners made a prediction of their test performance. They were then tested on their transfer of the algorithms to solve novel, more complex problems. Thereafter, learners rated the effectiveness of both learning methods.

Method

Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow the American Psychological Association Journal Article Reporting Standards. Data and analysis code are available at https://osf.io/us9jb/. Materials for this study are available by emailing the corresponding author. Data were analyzed using SPSS Version 26. This study was not preregistered.

Participants

The participants were 46 university students (37 were female) between the ages of 18 and 54 ($M = 22.13$, $SD = 7.63$) from the National University of Singapore. Previous studies comparing deliberate conceptual erring against errorless learning reported medium-to-large effect sizes for transfer ranging from $d = 0.46$ to 0.77 (Wong, 2023; Wong & Lim, 2022a). Based on the most conservative effect size, a power analysis (G*Power; Faul et al., 2007) indicated that at least 40 participants were required for two-tailed within-subjects pairwise comparisons at 80% power and $\alpha = .05$. Outcomes reported below are based on data from 40 participants; six participants who failed to conform to the experimental instructions were excluded from analyses.[1]

Both experiments were conducted with ethics approval from our university's Institutional Review Board. All participants provided their written informed consent and received course credit for their participation.

Design

The single within-subjects factor was learning method: *procedure-error* (deliberate error commission and correction) versus *repeated-practice* (errorless control condition). Participants were first trained on both learning methods, then used them to practice two mathematical algorithms (synthetic division vs. matrix multiplication), respectively, within the same duration for each method. We counterbalanced the order in which participants used both learning methods during their problem-solving practice, as well as the pairing of algorithms with learning methods. Participants were randomly and evenly assigned to each of the four counterbalanced sequences.

The outcome of interest was participants' performance on a transfer test, which involved applying the learned algorithms to solve novel mathematical problems that were structurally more challenging or embedded in "real-life" scenarios (i.e., mathematical word problems). The transfer test was blocked by algorithm, with participants being tested first on the algorithm that they had practiced first during the studying phase, then the second algorithm.

Materials

Basic Mathematical Ability Test. We administered the Mathematical Prerequisites for Psychometrics scale (PMP; Galli et al., 2008, 2011) as a measure of participants' basic mathematical ability. The scale comprises 30 multiple-choice questions that assess one's proficiency in the mathematical basics necessary for successfully completing introductory statistics courses (e.g., addition, subtraction, multiplication, division with fractions, first-order equations, relations between numbers from -1 to 1). A sample question was: "Considering the following equation: $3x + 27 = 18$, which is the value of x?" For each question, participants were required to select the correct answer from four alternatives. Performance on the PMP at the beginning of a university introductory statistics course has been found to predict students' achievement at the end of the course (Galli et al., 2008, 2011).

Learning Methods Training. During the training phase to familiarize participants with both the procedure-error and repeated-practice methods (see Procedure section for the detailed instructions for each learning method), column addition served as the training algorithm for illustration purposes. Column addition involves adding multidigit numbers (e.g., "$536 + 427 = ?$") by arranging their digits in place value columns and "carrying" or regrouping digits that are transferred across columns. We used column addition as the training algorithm because it is an elementary arithmetic operation that university students would likely be familiar with, thus enabling participants to focus on mastering both learning methods without potential interference or excessive cognitive load from having to simultaneously learn a new or complex mathematical algorithm (see Sweller et al., 1998, 2019).

To explain column addition to participants, we created a brief self-paced PowerPoint lesson comprising three slides. A pool of six practice questions on column addition was also constructed, with three questions randomly assigned for participants to practice solving using the procedure-error versus repeated-practice methods, respectively. Participants were given a corresponding printout of the column addition lesson slides that they could refer to while solving the practice questions during the training phase.

Mathematical Algorithm Lessons. During the studying phase, the critical to-be-learned mathematical algorithms were synthetic division and matrix multiplication. Synthetic division is a "shorthand" alternative to long division for dividing polynomials by a linear factor, whereas matrix multiplication involves multiplying matrices by computing the dot product of various combinations of their rows and columns.

[1] Of the six participants who were excluded from analyses, three participants did not implement the learning methods with fidelity during the studying phase (e.g., failing to solve each practice question twice in the repeated-practice condition, or failing to generate, circle, and correct procedural errors in the procedure-error condition), whereas the remaining three participants inadvertently answered all practice questions incorrectly in the repeated-practice condition. Hence, for accuracy, we have focused on reporting the results from the cleaned data without these six participants, although we note that the results remain the same even when they are included in the analyses.

For each algorithm, we created a self-paced PowerPoint lesson comprising eight slides. In each lesson, participants were first introduced to the algorithm and an example of a problem that it could be used to solve. Next, the definitions of the common terms associated with each algorithm were explained: "dividend," "divisor," "coefficient," and "degree" for synthetic division; "row," "column," "dimension," and "dot product" for matrix multiplication. Then, participants were taught the procedures for using each algorithm. To facilitate participants' learning, they were given access to a printout of the lesson slides that they could refer to during the studying phase, but which the experimenter collected back before the final test.

Practice Questions. A pool of six practice questions was created for each algorithm, for a total of 12 practice questions (see samples in Table 1). For each algorithm, one practice question was randomly designated as a checkpoint question that was presented immediately after the lesson to ascertain that all participants understood how to use the algorithm. Participants practiced solving the checkpoint question and received corrective feedback including the working and solution, before they proceeded to independently practice solving the remaining five questions using their randomly assigned learning method (either procedure-error or repeated-practice) without feedback.

Transfer Test Questions. For each algorithm, a pool of 10 transfer test questions was created (i.e., 20 test questions in total; see samples in Table 1). The transfer test questions were of greater structural complexity than the practice questions and required participants to execute additional steps that had not been explicitly required in the solution procedures for the practice questions. Four of the transfer test questions for each algorithm were further embedded in scenarios that described a "real-life" problem context (i.e., mathematical word problems).

For instance, a transfer test question for synthetic division could involve a polynomial dividend of a higher degree than had been encountered during practice (e.g., up to x^5 in the test questions vs. up to x^3 only in the practice questions) or executing additional steps (e.g., inserting "0" as a placeholder for missing terms, factoring a nonmonic linear divisor and dividing the resulting quotient accordingly, whereas the practice questions had no missing terms and only involved monic linear divisors). For matrix multiplication, a transfer test question could involve multiplying matrices of larger dimensions than had been encountered during practice (e.g., up to a 3 × 3 matrix in the test questions vs. up to a 2 × 2 matrix only in the practice questions) or executing additional steps (e.g., sequentially multiplying up to three matrices or constructing matrices of appropriate dimensions based on information given in the question before multiplying them, whereas the practice questions involved multiplying only two matrices that were directly provided). Thus, to successfully solve the transfer test questions for both algorithms, participants had to flexibly adapt the procedures that they had learned during the studying phase.

Postlearning Questionnaires. A five-item postlearning questionnaire (adapted from Wong, 2023; Wong & Lim, 2022a) was administered after participants had practiced each algorithm in the studying phase. Specifically, participants (a) made a judgment of learning (JOL) on an 11-point scale from 0% to 100% (i.e., 0%, 10%, 20%, …, 100%) to predict their performance when later tested on the algorithm, (b) rated how interesting the lesson on the algorithm was (1 = *not at all*; 7 = *extremely*), (c) rated how understandable the lesson was (1 = *not at all*; 7 = *extremely*), (d) indicated their prior knowledge of the lesson material (1 = *not very well*; 7 = *very*

Table 1
Sample Practice and Transfer Test Questions for Synthetic Division and Matrix Multiplication

Algorithm	Sample practice questions	Sample transfer test questions	Adaptation(s) required for transfer test questions
Synthetic division	$(2x^3–5x^2+3x+7) \div (x–2) = ?$	$(3x^5–9x^4+6x^2–5x+4) \div (x–4) = ?$	• Involves a more complex dividend (up to x^5) • Requires inserting "0" as a placeholder for the missing term (x^3) during synthetic division
	$(x^3+2x^2–3x+4) \div (x–5) = ?$	$(3x^3–5x^2+13x–4) \div (2x+10) = ?$	• Involves the additional steps of factoring the nonmonic linear divisor $(2x+10)$ into a monic linear divisor $2(x+5)$ before synthetic division can be applied, then dividing the resulting quotient by the monic linear divisor's coefficient
	$(8x^2–2x–3) \div (x+4) = ?$	In liquids, Pressure = Density × Height × Gravity. The pressure in a liquid is represented by $p^3–8p^2+19p–12$. Given that $p–1$ and $p–3$ are the height and gravity, respectively, find the density of the liquid.	• Involves applying synthetic division in a "real-life" scenario (i.e., word problem) • Requires performing synthetic division twice sequentially for each divisor: $(p–1)$ and $(p–3)$
Matrix multiplication	$\begin{bmatrix} 3 & -1 \\ 2 & 5 \end{bmatrix} \times \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = ?$	$\begin{bmatrix} 2 & 0 \\ -1 & 4 \\ -3 & 2 \end{bmatrix} \times \begin{bmatrix} 3 & -1 & 2 \\ -2 & 4 & 0 \end{bmatrix} = ?$	• Involves multiplying more complex matrices with larger dimensions
	$\begin{bmatrix} 9 & 4 \\ 3 & 6 \end{bmatrix} \times \begin{bmatrix} 7 \\ 5 \end{bmatrix} = ?$	$\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = ?$	• Requires sequentially multiplying three matrices, rather than two only
	$\begin{bmatrix} 6 & 8 \end{bmatrix} \times \begin{bmatrix} 5 & 7 \\ 9 & 4 \end{bmatrix} = ?$	Costumes are needed for a play. Boys' costumes require eight units of fabric, five units of ribbon, and four units of sequins. Girls' costumes require six units of fabric, nine units of ribbon, and five units of sequins. The costs per unit of fabric, ribbon, and sequins are $6, $5, and $4, respectively. Find the costs of the costumes for boys versus girls.	• Involves applying matrix multiplication in a "real-life" scenario (i.e., word problem) • Requires constructing two matrices of appropriate dimensions based on the information given in the problem, then multiplying them

well), and (e) rated how easy it had been for them to learn the algorithm (1 = *not at all*; 7 = *extremely*). After participants had completed the test phase, they further rated how effective they thought each learning method had been in helping them learn the algorithms (1 = *not at all*; 7 = *extremely*).

Procedure

Before attending the experiment, participants completed the basic mathematical ability test via an online questionnaire. Upon arriving at the laboratory, participants were seated at individual cubicles and informed that they would be learning mathematical algorithms. They were told to study the algorithms to the best of their ability for a later test; the specific nature of the test was not revealed. The experiment comprised three phases administered via pen and paper: training, studying, and test. Throughout all three phases, participants performed all mathematical calculations by hand (i.e., without using a calculator). The total experimental duration was approximately 90 min.

Training Phase. During the training phase, all participants were first introduced to the training algorithm (column addition) via a PowerPoint lesson, and were instructed on both the repeated-practice and procedure-error methods. Then, participants were given 3 min to practice solving three column addition practice questions using each learning method. All participants were provided with a printout of the column addition lesson slides that they could refer to during their practice.

In the repeated-practice condition, participants were instructed to correctly solve each question twice and then check that their workings and answers for both attempts were accurate and consistent. In this way, participants engaged in errorless repeated practice for each question.

In the procedure-error condition, participants also solved each question twice, but once incorrectly. Specifically, participants were instructed to deliberately err in their first attempt by incorrectly writing down their working such that it contained a plausible procedural error (i.e., an error in applying or executing the action sequences for the algorithm). Then, participants were to circle the error that they had made, and correct it by writing down the correct working to solve the question in their second attempt. To ensure that participants understood what was required of them during the training phase, they were shown examples of procedural versus nonprocedural errors for the training algorithm. For instance, in column addition, neglecting to add the regrouped "1" from the ones to tens place is a procedural error, whereas writing "4 + 1 = 6" merely involves a calculation error. In line with extant research on incorrect guessing (Kang et al., 2011) and competitive incorrect responses (Little & Bjork, 2015; Little et al., 2012), participants were encouraged to make plausible procedural errors that were objectively incorrect but still believable. For instance, "5,467 + 192 = 5,559" (i.e., initiating the regrouping process but neglecting to add the regrouped "1" from the tens to hundreds column) is a more plausible procedural error than "5,467 + 192 = 55,159" (i.e., leaving the regrouped "1" in the tens column).

At the end of the 3-min practice duration for each learning method, all participants were presented with corrective feedback for each column addition practice question, including its working and solution. All participants solved the practice questions correctly.

Studying Phase. After the training phase, participants underwent the studying phase, which comprised two learning blocks. In each learning block, participants learned and practiced one of the critical mathematical algorithms (either synthetic division or matrix multiplication) using one of the learning methods (either repeated-practice or procedure-error). The order in which participants used both learning methods, as well as the pairing of algorithms with learning methods, was counterbalanced.

Specifically, in each learning block, participants were taught one of the critical mathematical algorithms via a PowerPoint lesson. At the end of the lesson, participants were given 1.5 min to practice solving a checkpoint question to ascertain that they understood how to use the algorithm, before receiving corrective feedback that included both the working and solution. Emulating naturalistic self-regulated learning environments, the studying phase was conducted under open-book conditions in which participants were provided with a printout of the algorithm lesson slides that they could refer to during their practice. Across both learning blocks, all except four participants correctly answered the checkpoint questions. For participants who incorrectly answered the checkpoint question, the experimenter verbally reinforced the corrective feedback to ensure participants' full understanding before they proceeded with the task.

Following which, all participants were given 12 min to independently practice solving five questions for the algorithm using their randomly assigned learning method (either the repeated-practice or procedure-error method). No feedback was provided for these practice questions. The instructions for both learning methods were identical to those in the training phase. Participants were further told that if they finished before the time was up, they should use the remaining time to check their responses. Thus, the total studying duration was equated across all conditions. All participants finished solving the practice questions within the given time. At the end of the 12-min learning block, participants completed a five-item postlearning questionnaire in which they made a JOL to predict their test performance, rated the lesson's interestingness and understandability, indicated their prior knowledge of the algorithm, and rated how easy it had been for them to learn the algorithm.

The same procedure was then repeated for the second learning block, in which participants used the second learning method to study the other critical algorithm. After completing both learning blocks, all participants were allowed to take a brief self-paced break before starting the test phase.

Test Phase. The test phase was conducted in two blocks by algorithm, corresponding to the same order in which participants had learned both algorithms during the studying phase. In each test block, participants were given 15 min to correctly solve 10 transfer test questions for one of the critical algorithms (either synthetic division or matrix multiplication). Participants completed the test questions without reference to their studying phase responses or printouts of the algorithm lesson slides, which the experimenter collected back before the test. At the end of the first 15-min test block, participants were allowed to take a brief self-paced break before they proceeded to solve the test questions for the second algorithm in the next 15-min test block. After completing both test blocks, participants rated the effectiveness of each learning method and provided their demographic information. All participants were then debriefed and thanked.

Results

Scoring

Participants' responses in the studying and test phases were scored by awarding either 1, 0.5, or 0 marks for each question, with a maximum possible score of 5 for the studying phase and 10 for the transfer test. Specifically, participants were awarded 0.5 marks for correctly applying the algorithm procedures to solve each question (i.e., method marks), with an additional 0.5 marks awarded for accuracy in the intermediate steps and final answer following a correct method (i.e., accuracy marks). Accuracy marks were not awarded unless participants had first earned the method marks for a given question. Thus, to receive any marks (i.e., at least 0.5 marks) for any particular question, participants had to first demonstrate an understanding of the procedures for correctly applying the algorithm. Similar scoring schemes have been used in international standardized tests such as the General Certificate of Education Advanced Level (GCE A-Level) mathematics examination (Cambridge Assessment International Education, 2022).

For instance, a participant would earn one mark (i.e., full marks) for a question if their response correctly applied the algorithm procedures to solve the specific question at hand, and provided an accurate working and answer. Marks were not deducted for up to two minor arithmetic slips (e.g., calculation errors) in a response, provided that these did not suggest a lack of understanding in applying the algorithm. Conversely, a participant would earn 0.5 marks for a question if their response correctly applied the algorithm procedures but contained several arithmetic slips (e.g., three or more calculation errors, provided that these did not exceed 50% of the response to produce an egregiously wrong answer), or if the participant provided a fully correct working that demonstrated their understanding of the algorithm procedures but did not derive the final answer (e.g., writing down a fully correct working but failing to compute the final answer). Conversely, a participant would not earn any marks (i.e., zero marks) for a question if they incorrectly applied the algorithm procedures or did not attempt the question (i.e., no response).

Two raters independently scored all 40 scripts. Interrater reliability was excellent, intraclass correlation (ICC) = .95, 95% confidence interval (CI) [.93, .97], based on a two-way random-effects model. Discrepancies were reviewed and resolved to reach 100% agreement.

Basic Mathematical Ability and Prior Knowledge

Participants' mean basic mathematical ability score on the PMP was 25.93 ($SD = 3.78$) out of a total possible score of 30. Correlational analyses indicated that participants' PMP scores were positively associated with their transfer test performance in both the repeated-practice condition, $r(38) = .50$, $p = .001$, and procedure-error condition, $r(38) = .58$, $p < .001$.

We further analyzed participants' self-reported familiarity with the critical algorithms. Participants reported relatively low prior knowledge of the algorithms on overall, with no significant difference across the repeated-practice ($M = 3.20$, $SD = 1.76$) and procedure-error ($M = 3.15$, $SD = 1.69$) conditions, $t(39) = -0.12$, $p = .909$, $d = -0.02$, 95% CI [−0.33, 0.29].

Lesson Interestingness, Understandability, and Ease of Learning

Participants' ratings of how interesting the algorithm lessons were did not significantly differ across both learning conditions, $t(39) = -1.00$, $p = .323$, $d = -0.16$, 95% CI [−0.47, 0.16]. Neither was there a significant difference in participants' ratings of how understandable the algorithm lessons were, $t(39) = -0.56$, $p = .578$, $d = -0.09$, 95% CI [−0.40, 0.22], nor how easy it had been for them to learn the algorithms across both conditions, $t(39) = -0.14$, $p = .891$, $d = -0.02$, 95% CI [−0.33, 0.29]. Table 2 shows the means and *SD*s.

Studying Phase Performance

We ascertained that across both learning conditions, participants were highly successful in solving the five practice questions during the studying phase, with no significant difference between the repeated-practice ($M = 4.79$, $SD = 0.56$) and procedure-error ($M = 4.66$, $SD = 0.54$) conditions, $t(39) = -1.40$, $p = .168$, $d = -0.22$, 95% CI [−0.53, 0.09]. Thus, any subsequent differences in participants' performance on the transfer test could not be attributed to differing success rates during their initial practice.

Transfer Test Performance

As predicted, the procedure-error method ($M = 5.43$, $SD = 1.87$) produced superior transfer test performance than the repeated-practice method ($M = 4.14$, $SD = 1.70$), $t(39) = 4.84$, $p < .001$, $d = 0.77$, 95% CI [0.41, 1.11]. Attesting to the derring effect (Wong & Lim, 2022a, 2022b), deliberately committing and correcting procedural errors during practice outperformed errorless repeated practice (Figure 1A).

Breaking down participants' total score on the transfer test, we further analyzed their performance on the six transfer questions that were structurally more complex versus the four transfer questions that were embedded in "real-life" scenarios (i.e., mathematical word problems). We found that the derring effect generalized across both question types. Specifically, the procedure-error method ($M = 3.41$, $SD = 1.22$) yielded an advantage over the repeated-practice method ($M = 2.61$, $SD = 1.20$) for the transfer questions of increased structural complexity, $t(39) = 3.57$, $p = .001$, $d = 0.56$, 95% CI [0.23, 0.90]. In addition, the procedure-error method ($M = 2.01$, $SD = 1.14$) outperformed the repeated-practice method ($M = 1.53$, $SD = 0.98$) on the word problems in the transfer test, $t(39) = 2.27$, $p = .029$, $d = 0.36$, 95% CI [0.04, 0.68].

Metacognitive Judgments

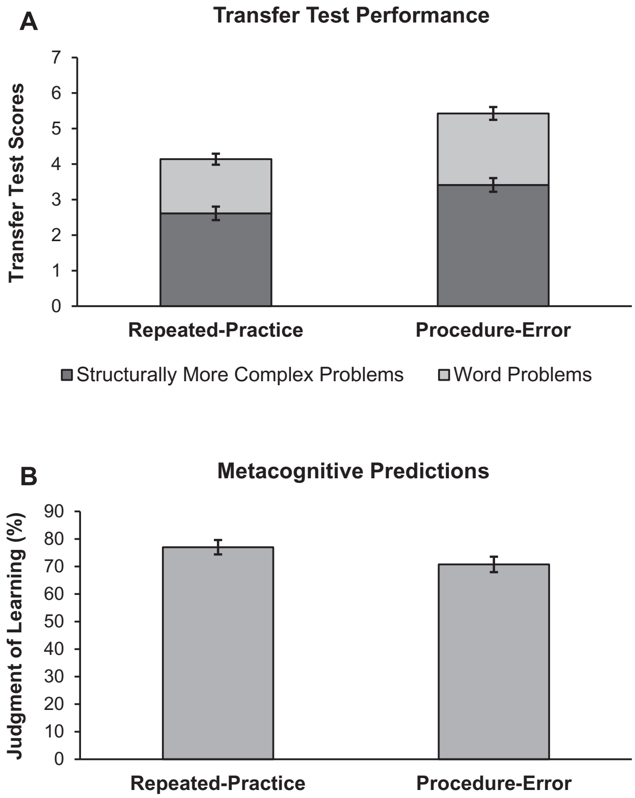
In contrast to their actual performance, participants inaccurately predicted in their JOLs that their test performance would be significantly better in the repeated-practice than procedure-error condition, $t(39) = -2.61$, $p = .013$, $d = -0.41$, 95% CI [−0.73, −0.09]. Figure 1B displays participants' JOLs across learning conditions. This metacognitive illusion persisted even after participants had experienced the benefits of deliberate erring for their test performance.

Table 2
Mean Scores on Postlearning Questionnaires (Experiment 1)

	Repeated-practice		Procedure-error	
Variable	*M*	*SD*	*M*	*SD*
Prior knowledge	3.20	1.76	3.15	1.69
Lesson interestingness	4.80	1.54	4.60	1.65
Lesson understandability	5.73	0.99	5.63	1.19
Ease of learning	5.18	1.11	5.15	1.05
Judgment of learning	77.00	16.52	70.75	17.74
Method effectiveness	5.28	1.54	3.98	1.92

Note. $N = 40$.

Figure 1
Transfer Test Performance and Metacognitive Predictions (Experiment 1)



Note. (A) The mean transfer test scores across question types and learning conditions; the total possible score for the transfer test was 10. (B) Participants' predictions of their test performance (i.e., their metacognitive judgments of learning). Error bars indicate *SE*s.

When rating the effectiveness of both learning methods after the test, participants still incorrectly judged the repeated-practice method as more effective than the procedure-error method, $t(39) = -3.76$, $p = .001$, $d = -0.60$, 95% CI [−0.93, −0.26]. Means and *SD*s of participants' metacognitive ratings are presented in Table 2.

We further examined the predicted versus actual effectiveness of both learning methods for participants' transfer performance. Overall, 29 out of 40 (73%) learners performed better in the procedure-error condition than repeated-practice condition. Yet, 34 out of 40 (85%) learners predicted that the repeated-practice method would be just as effective as or even more effective than the procedure-error method. Even after the transfer test, the advantage of deliberate erring went largely unappreciated—33 out of 40 (83%) learners rated the repeated-practice method as just as effective as or even more effective than the procedure-error method for their test performance. Taken together, participants' pre- and posttest metacognitive judgments were at odds with how effective both learning methods had been. Table 3 shows the number of participants who actually performed better after deliberate erring than errorless repeated practice, the number who showed the opposite pattern, and the number who performed similarly across both conditions. For each of these three performance outcomes, Table 3 also shows the number of participants who made the corresponding pretest metacognitive predictions (JOLs) and posttest metacognitive judgments (effectiveness ratings).

Discussion

Experiment 1 tested and found evidence for the benefit of deliberately committing and correcting procedural errors in mathematical problem-solving and transfer. When learners intentionally erred while practicing mathematical algorithms, they were subsequently more successful in applying those algorithms to solve novel problems that were structurally more complex or embedded in "real-life" scenarios, relative to avoiding errors during their practice. Notably, this advantage occurred even when the errorless control condition involved correctly solving each practice problem twice. Thus, despite receiving more exposure to the correct solutions during errorless repeated practice, learners still transferred better at test when they had deliberately solved each practice problem wrongly then corrected it once.

However, the vast majority of learners failed to accurately predict the advantage of deliberate erring. Even after experiencing the effects of both learning methods for their test performance, most learners continued to believe that errorless repeated practice had been more effective. These results align with those in previous research that students are often woefully unaware of the benefits of deliberate erring, even after profiting from this technique (Wong, 2023; Wong & Lim, 2022a, 2022b).

Experiment 2

Experiment 2 was conducted to address two critical questions from Experiment 1's findings. First, why does deliberate erring improve transfer? Previous research has shown that even comparing a mixture of correct and incorrect solutions promotes transfer more than studying correct solutions only (e.g., Große & Renkl, 2007; Loibl & Leuders, 2019). For instance, juxtaposing correct versus incorrect solutions may prompt learners to detect and repair flaws or gaps in their mental models for deeper learning (Heemsoth & Heinze, 2014; Loibl & Leuders, 2019; Loibl & Rummel, 2014). Hence, it is possible that learners' better transfer in the procedure-error condition was merely due to being exposed to incorrect workings and/or comparing them with the correct workings during error correction. If this were indeed the case, then one would expect that learners could reap similar gains from simply being presented with others' errors and juxtaposing them with the correct solutions, even without having erred themselves.

Second and relatedly, whereas Experiment 1 showed that deliberate erring outperforms errorless learning, does this advantage hold over other errorful methods that benefit mathematical learning? Notably, studying others' errors in incorrect worked examples similarly involves engaging with errors during or after instruction on to-be-learned material and has been found to improve mathematical transfer more than studying correct examples only or a problem-solving control (e.g., Adams et al., 2014; Barbieri & Booth, 2020; Booth et al., 2013; Durkin & Rittle-Johnson, 2012; Große & Renkl, 2007; McLaren et al., 2015). Is there anything special about deliberately committing and correcting one's own errors?

Although studies that directly compare errorful learning approaches are scarce, some limited evidence suggests that learners could gain more from personally making errors and correcting them (e.g., Metcalfe & Xu, 2018; Sadler & Good, 2006). For instance, Wong (2023) found that deliberately committing and correcting conceptual

Table 3

Frequency Count (and Percentage) of Participants Showing Different Patterns of Metacognitive Ratings and Actual Transfer Test Performance (Experiment 1)

	Performance outcome		
Metacognitive ratings versus actual performance	Procedure-error > repeated-practice	Procedure-error = repeated-practice	Procedure-error < repeated-practice
Metacognitive ratings			
Pretest predictions (JOLs)	6 (15%)	17 (43%)	17 (43%)
Posttest judgments (effectiveness ratings)	7 (18%)	7 (18%)	26 (65%)
Actual transfer test performance	29 (73%)	2 (5%)	9 (23%)

Note. $N = 40$. JOL = judgment of learning.

errors yielded better far transfer of conceptual knowledge than a spot-and-fix condition in which learners spotted and corrected their peers' deliberate errors then generated their own correct answers. However, the spot-and-fix condition did not require learners to explicitly explain the errors that they had identified. Although self-explanation does not seem to be a prerequisite for benefiting from studying worked examples (Corral & Carpenter, 2020), it could enhance the effectiveness of this learning approach in supporting inference-making and generalization of knowledge for better transfer (Barbieri et al., 2019; Chi et al., 1989; Loibl & Leuders, 2019; Rittle-Johnson, 2006; for reviews, see Atkinson et al., 2000; Bisra et al., 2018; Chi, 2022; Fiorella, 2023; Renkl, 2014; Rittle-Johnson et al., 2017).

Hence, Experiment 2 directly compared deliberate erring (*procedure-error* method) against the errorful learning approach of studying incorrect worked examples, in which learners identified, explained, and corrected others' errors (*spot-explain-fix* method). Specifically, each learner was presented with the incorrect solution attempts that a peer had deliberately made on the practice problems in Experiment 1.[2] This yoked procedure enabled us to control for the quality of errors that learners were exposed to across the procedure-error and spot-explain-fix conditions (e.g., Wong, 2023). Using the spot-explain-fix method, learners identified their peers' errors in the incorrect worked examples, explained why those solutions were incorrect and what should have been done instead, then corrected the errors by solving the practice problem correctly. Thus, Experiment 2 allowed us to explore the benefits of personally committing and correcting errors, over and above the general benefits of comparing incorrect and correct solution procedures.

Method

Participants

The participants were 43 university students (29 were female) between the ages of 19 and 43 ($M = 21.42$, $SD = 4.70$) from the National University of Singapore who did not take part in Experiment 1. Outcomes reported below are based on data from 42 participants; one participant who failed to follow the experimental instructions was excluded from subsequent analyses. A power analysis (G*Power; Faul et al., 2007) indicated that this sample size afforded sufficient sensitivity to detect medium within-subjects effects ($d \geq 0.44$) for two-tailed pairwise comparisons at 80% power and $\alpha = .05$.

Design

The single within-subjects factor was learning method: *procedure-error* (deliberate error commission and correction) versus *spot-explain-fix* (identifying, explaining, and correcting others' deliberate errors). As in Experiment 1, the learning outcome of interest was participants' transfer test performance.

Materials and Procedure

Experiment 2 used identical materials and procedures as Experiment 1, except that the repeated-practice condition was replaced with a spot-explain-fix condition in which participants identified, explained, and corrected the errors in others' incorrect workings when practicing the algorithms. Specifically, for each practice question in the spot-explain-fix condition, each participant was presented with a corresponding incorrect working that a peer had generated for that question in Experiment 1's procedure-error condition. Thus, in this yoked design, each participant was exposed to a different set of deliberate errors that a peer from Experiment 1 had made. Participants were instructed to identify and circle the error in the incorrect working presented. Next, participants were prompted to write down an explanation for why their peer's working was incorrect and what should have been done instead. Then, participants corrected the erroneous working by writing down the correct working to solve the given practice question. Similar procedures have been used in studies on learning from incorrect worked examples (e.g., Barbieri & Booth, 2020; Booth et al., 2013; Heemsoth & Heinze, 2014). Hence, whereas the procedure-error condition involved generating deliberately incorrect solutions and correcting them, the spot-explain-fix condition involved spotting and explaining others' deliberate errors plus generating correct solutions.

Results

Scoring

Two raters independently scored all 42 scripts in the same way as in Experiment 1. Interrater reliability was excellent, ICC = .96, 95%

[2] It was not possible to present learners with the deliberate errors generated by a peer in Experiment 2's procedure-error condition while still preserving a within-subjects design similar to that in Experiment 1. Alternatively, instructor-selected common errors could be presented in the incorrect worked examples, but we note that this would have varied the quality of errors that learners experienced in the spot-explain-fix versus procedure-error conditions. Upon balanced consideration, we used the deliberate errors from Experiment 1's procedure-error condition in the incorrect worked examples for Experiment 2's spot-explain-fix condition, given that both experiments had identical learning materials and procedures with participants sampled from the same population (see Wong, 2023 for similar procedures).

CI [.94, .98], based on a two-way random-effects model. Discrepancies were reviewed and resolved to reach 100% agreement.

Basic Mathematical Ability and Prior Knowledge

The PMP scores of three participants were missing due to experimenter error. Participants' mean basic mathematical ability score on the PMP was 26.28 ($SD = 3.01$) out of a total possible score of 30. Participants' PMP scores were not significantly associated with their transfer test performance in the spot-explain-fix condition, $r(37) = .26$, $p = .105$, but positively correlated with their transfer test performance in the procedure-error condition, $r(37) = .33$, $p = .043$.

As in Experiment 1, participants reported relatively low prior knowledge of the critical algorithms on overall, with no significant difference across the spot-explain-fix ($M = 3.21$, $SD = 1.91$) and procedure-error ($M = 3.05$, $SD = 1.95$) conditions, $t(41) = -0.47$, $p = .643$, $d = -0.07$, 95% CI [−0.37, 0.23].

Lesson Interestingness, Understandability, and Ease of Learning

Participants' ratings of how interesting the algorithm lessons were did not significantly differ across the spot-explain-fix and procedure-error conditions, $t(41) = -0.72$, $p = .474$, $d = -0.11$, 95% CI [−0.41, 0.19]. Neither did participants' ratings of the algorithm lessons' understandability differ across conditions, $t(41) = -1.00$, $p = .323$, $d = -0.15$, 95% CI [−0.46, 0.15]. There was also no significant difference in participants' ratings of how easy it had been for them to learn the algorithms in both conditions, $t(41) = -1.73$, $p = .092$, $d = -0.27$, 95% CI [−0.57, 0.04]. Table 4 shows the means and *SD*s.

Studying Phase Performance

We ascertained that participants were highly successful in solving the practice questions during the studying phase, with no significant difference between the spot-explain-fix ($M = 4.87$, $SD = 0.31$) and procedure-error ($M = 4.75$, $SD = 0.57$) conditions, $t(41) = -1.33$, $p = .193$, $d = -0.20$, 95% CI [−0.51, 0.10]. Thus, any subsequent differences in participants' transfer test performance could not be attributed to differing rates of success during their initial practice.

Transfer Test Performance

The key finding in Experiment 2 was that the procedure-error method ($M = 5.94$, $SD = 1.98$) produced better transfer test performance than the spot-explain-fix method ($M = 4.71$, $SD = 2.26$), $t(41) = 4.09$, $p < .001$, $d = 0.63$, 95% CI [0.30, 0.96]. Hence, personally generating and correcting one's own deliberate errors was more beneficial for transfer than spotting, explaining, and correcting others' deliberate errors (Figure 2A).

Decomposing participants' total score on the transfer test, we further analyzed their performance on the six transfer questions that were structurally more complex versus the four transfer questions that were embedded in "real-life" scenarios (i.e., word problems). On the transfer questions of increased structural complexity, the procedure-error method ($M = 3.63$, $SD = 1.44$) yielded superior performance than the spot-explain-fix method ($M = 3.05$, $SD = 1.48$), $t(41) = 2.07$, $p = .045$, $d = 0.32$, 95% CI [0.01, 0.63]. Likewise, on the word problems, the procedure-error method ($M = 2.31$, $SD = 1.41$) outperformed the spot-explain-fix method ($M = 1.67$, $SD = 1.20$), $t(41) = 2.55$, $p = .014$, $d = 0.39$, 95% CI [0.08, 0.71].

Metacognitive Judgments

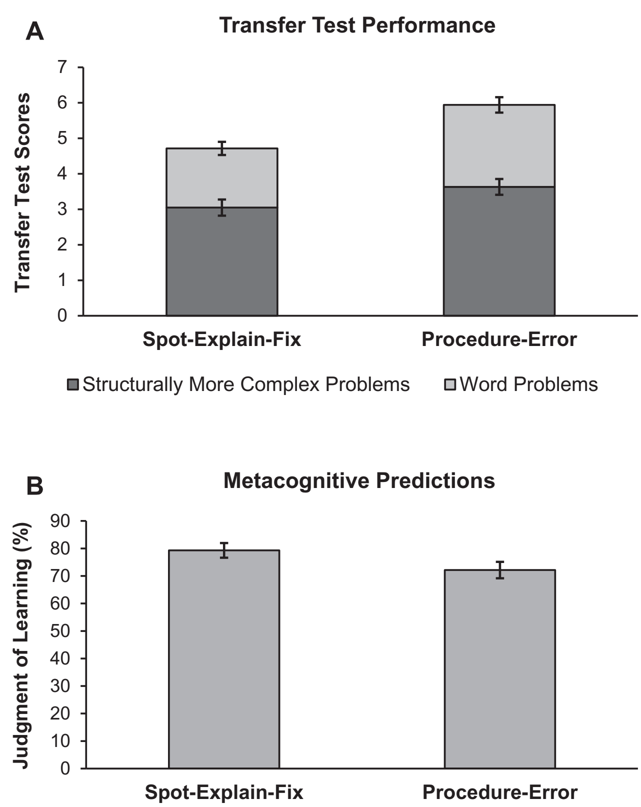
However, participants inaccurately predicted in their JOLs that their performance would be better in the spot-explain-fix than procedure-error condition, $t(41) = -2.82$, $p = .007$, $d = -0.44$, 95% CI [−0.75, −0.12]. Figure 2B displays participants' JOLs

Table 4
Mean Scores on Postlearning Questionnaires (Experiment 2)

Variable	Spot-explain-fix *M*	Spot-explain-fix *SD*	Procedure-error *M*	Procedure-error *SD*
Prior knowledge	3.21	1.91	3.05	1.95
Lesson interestingness	4.64	1.36	4.50	1.45
Lesson understandability	5.88	1.02	5.71	1.13
Ease of learning	5.40	1.08	5.05	1.29
Judgment of learning	79.29	17.31	72.14	19.32
Method effectiveness	4.86	1.37	4.07	1.46

Note. $N = 42$.

Figure 2
Transfer Test Performance and Metacognitive Predictions (Experiment 2)



Note. (A) The mean transfer test scores across question types and learning conditions; the total possible score for the transfer test was 10. (B) Participants' predictions of their test performance (i.e., their metacognitive judgments of learning). Error bars indicate *SE*s.

across learning conditions. Even after experiencing the benefits of deliberate erring for their test performance, participants still misjudged the spot-explain-fix method as more effective than the procedure-error method, $t(41) = -2.48$, $p = .017$, $d = -0.38$, 95% CI [−0.69, −0.07]. Means and *SD*s of participants' metacognitive ratings are presented in Table 4.

Overall, 29 out of 42 (69%) learners performed better in the procedure-error than spot-explain-fix condition. Yet, 37 out of 42 (88%) learners predicted that the spot-explain-fix method would be just as effective as or even more effective than the procedure-error method. Learners' unawareness of the deliberate erring advantage persisted even after the transfer test, whereby 35 out of 42 (83%) learners rated the spot-explain-fix method as just as effective as or even more effective than the procedure-error method for their test performance. Thus, as in Experiment 1, participants' pre- and posttest metacognitive judgments stood in stark contrast to how effective both methods had actually been (Table 5).

Discussion

Experiment 2's findings dovetail with those in previous research showing the benefits of learning from self-generated errors (e.g., Metcalfe & Xu, 2018; Sadler & Good, 2006). Extending Experiment 1's results, we found that deliberately committing and correcting procedural errors was more beneficial for mathematical problem solving and transfer than finding, explaining, and correcting others' deliberate errors in incorrect worked examples. This effect occurred even though both methods similarly involved exposure to errors and comparing them with the correct solutions during error correction, with error quality controlled for across conditions via a yoked procedure. Moreover, the spot-explain-fix method had been supplemented with self-explanation that previous research has shown to facilitate transfer (e.g., Rittle-Johnson, 2006), whereas the procedure-error method did not require that learners explicitly self-explain their deliberate errors. Taken together, these results suggest a new level of robustness to the derring effect in mathematical problem-solving and transfer: Personally committing and correcting one's own deliberate errors offers a unique advantage over learning from others' deliberate errors (Wong, 2023).

As in Experiment 1, though, there was a large disconnect between learners' metacognitive judgments and actual test performance. Besides mispredicting that the procedure-error method would yield worse performance than the spot-explain-fix method, learners persisted in believing that the procedure-error method was less effective even after profiting from it on the transfer test.

General Discussion

Most educators would like for their students to be able to transfer knowledge they have learned to new contexts or problems (Bransford & Schwartz, 1999; Haskell, 2001; McKeough et al., 1995; Perkins & Salomon, 1992), but this has often proven difficult to achieve (Detterman, 1993; Renkl et al., 1996). Here, we show in the domain of mathematics that guiding learners to deliberately commit and correct procedural errors during problem-solving practice after instruction enhances transfer of learned procedures.

In Experiment 1, deliberate erring was more effective than traditional errorless repeated practice in enabling learners to apply and flexibly adapt learned mathematical algorithms to solve novel, more complex problems. This result offers the first evidence that the derring effect extends to procedural transfer, thus establishing its generalizability beyond conceptual learning and transfer (Wong, 2023; Wong & Lim, 2022a, 2022b).

Experiment 2's findings go some way toward clarifying the locus of the derring effect. Specifically, our results indicate that the deliberate erring advantage in Experiment 1 was not merely due to exposing learners to incorrect solutions and/or juxtaposing them with the correct ones during error correction. Rather, Experiment 2 provided evidence that personally committing and correcting deliberate errors is vital to unleashing the full potential of learning from errors. When learners intentionally made and corrected their own errors, they transferred better than when they spotted, explained, and corrected their peers' errors (see also Wong, 2023). This finding corroborates the advantage of learning from self- over other-generated errors in extant research on other error types, including those that are allowed to occur spontaneously (Metcalfe & Xu, 2018; Sadler & Good, 2006) or induced before direct instruction (Kapur, 2014a, 2014b; cf. Hartmann et al., 2021, 2022).

On closer inspection of these effects, the transfer task serves as a window into what had been learned (McDaniel, 2007). Notably, across both experiments, deliberate erring reliably benefited performance on not only the transfer questions that were structurally more complex, but also the mathematical word problems. To successfully solve word problems, learners must construct a mental representation of the problem and formulate a corresponding solution plan, besides applying learned mathematical procedures (Kintsch & Greeno, 1985; Mayer & Hegarty, 1996; Pongsakdi et al., 2020). Thus, one implication is that deliberate erring may have fostered recognition and encoding of the target knowledge's deep features, which facilitated transfer to new problems (Loibl et al., 2017). When deliberately generating their own incorrect solutions, learners' attention may have

Table 5

Frequency Count (and Percentage) of Participants Showing Different Patterns of Metacognitive Ratings and Actual Transfer Test Performance (Experiment 2)

Metacognitive ratings versus actual performance	Performance outcome: Procedure-error > spot-explain-fix	Procedure-error = spot-explain-fix	Procedure-error < spot-explain-fix
Metacognitive ratings			
Pretest predictions (JOLs)	5 (12%)	18 (43%)	19 (45%)
Posttest judgments (effectiveness ratings)	7 (17%)	12 (29%)	23 (55%)
Actual transfer test performance	29 (69%)	3 (7%)	10 (24%)

Note. $N = 42$. JOL = judgment of learning.

been more strongly directed toward the defining features of the to-be-learned procedures and problem structures, relative to studying others' errors or simply following the procedural steps during errorless practice (e.g., Terwel et al., 2009). Examining these potential processes more directly represents an interesting avenue for future work.

In spite of the benefits of deliberate erring, the majority of learners in both experiments misjudged this technique as less effective than it actually was, even after experiencing its effects on their transfer test performance. Such stable metacognitive illusions have similarly been observed in previous deliberate erring research (Wong, 2023; Wong & Lim, 2022a, 2022b) and align more broadly with learners' unawareness of the benefit of generating errors (Huelser & Metcalfe, 2012; Pan et al., 2020; Yang et al., 2017).

Although learners in our study were not asked to explain their metacognitive judgments, it is possible that these may have been influenced by theory-based and/or experience-based cues (Koriat, 1997; Koriat et al., 2004). For instance, students may hold a priori theories or beliefs that committing errors is bad for their learning (Pan et al., 2020; Yang et al., 2017). Furthermore, generating errors may be subjectively experienced as less fluent than correctly practicing learned procedures or studying others' errors from a "safe distance," thus lowering JOLs (e.g., Potts & Shanks, 2014). Indeed, learners may rely on an "easily learned, easily remembered" heuristic when making metacognitive judgments (Koriat, 2008), such that greater subjective ease of processing can lead to overconfidence or illusions of knowing (e.g., Begg et al., 1989; Hertzog et al., 2003; Undorf & Erdfelder, 2011; see Finn & Tauber, 2015 for a review). Such miscalibrations bear potential costs for self-regulated learning, given that accurately monitoring one's learning is crucial for supporting effective and efficient control of what and how to study (Bjork et al., 2013; Dunlosky & Rawson, 2012; Metcalfe & Finn, 2008).

Educational Implications and Future Directions

In demonstrating the benefits of guiding deliberate errors, our findings offer new practical insights for how errors can be effectively approached in mathematical learning to improve problem solving and transfer. For instance, after students have been instructed on to-be-learned mathematical algorithms, they could be guided to deliberately commit and correct procedural errors when solving practice problems in assignments or class discussions. Whereas students in our study independently studied and practiced the algorithms, this learning process can potentially be supplemented with explicitly teaching for transfer. Some evidence suggests that transfer on mathematical problems is boosted when teachers not only instruct students on a solution method, but also teach abstraction and metacognitive skills for transfer (e.g., Fuchs et al., 2003). Such instruction could include explicitly teaching the concept of transfer and the types of superficial problem features that alter a problem such that it appears "new" even when its structure remains the same, while prompting students to search novel problems for superficial changes to identify familiar deep structures (e.g., Chi & VanLehn, 2012; Fuchs et al., 2003).

To inform educational recommendations that are more precisely attuned to learner characteristics, further work is needed to examine the role of prior knowledge in the effectiveness of deliberate erring. In our study, learners had relatively low prior knowledge of the to-be-learned mathematical algorithms. For these learners, deliberate erring during problem-solving practice was more beneficial than errorless repeated practice or studying incorrect worked examples. This result aligns with previous studies on deliberate conceptual errors that observed the derring effect even when students had little prior knowledge about the to-be-learned concepts (Wong, 2023; Wong & Lim, 2022a, 2022b). However, we note that learners in our study had high basic mathematical ability in already knowing how to carry out fundamental arithmetic operations such as addition and subtraction that the algorithms required. It remains to be explored whether high—or, at least, some—prior domain knowledge modulates the derring effect. It is conceivable that some degree of domain knowledge enables learners to engage productively with deliberate erring when generating sensible or meaningful errors, while avoiding cognitive overload (Sweller et al., 1998, 2019). If so, learners with low prior domain knowledge may require more support to benefit fully from deliberate erring (see also Barbieri & Booth, 2016; Große & Renkl, 2007).

Our data also raise new questions and possibilities about the reach of the derring effect. Whereas we focused on testing the transfer of learned mathematical procedures within the same knowledge domain, recent evidence suggests that deliberate erring can also aid far transfer of learned concepts across different knowledge domains (Wong, 2023). Synthesizing these findings, it could thus be fruitful for future work to explore the derring effect in far transfer of learned procedures across knowledge domains, such as between isomorphic topics in algebra and physics (e.g., transferring procedures in algebraic arithmetic-progression problems to solve unfamiliar but analogous constant-acceleration problems in physics; Bassok, 1990; Bassok & Holyoak, 1989).

Relatedly, conceptual and procedural knowledge in mathematics have been viewed to develop iteratively, with gains in conceptual understanding supporting the generation of correct procedures, and gains in procedural knowledge in turn increasing conceptual understanding (Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001, 2015). By extension, it is possible that deliberately committing and correcting procedural errors in mathematics may support learners' conceptual knowledge development, thereby boosting their transfer when constructing solutions for new problems. Future work could address this prospect by measuring conceptual and procedural knowledge independently. Whereas procedural knowledge in mathematics has almost always been measured via problem-solving accuracy, conceptual knowledge of mathematical principles can be assessed in more varied ways such as sorting examples into categories, generating or selecting definitions for concepts, and evaluating unfamiliar procedures (for a review, see Rittle-Johnson, 2019). Given the iterative relations between conceptual and procedural knowledge, it may be challenging for a measure to target only a single type of knowledge but not the other. Thus, such measures are thought to predominantly—rather than exclusively—assess one type of knowledge (Rittle-Johnson, 2019). With the use of multiple measures for each type of knowledge, future work would be better poised to examine convergent and discriminant validity when disentangling conceptual and procedural knowledge (Schneider & Stern, 2010).

Another open question relates to the emotional and motivational effects of deliberate erring. When unintended, errors can cause considerable chagrin and emotional strain, especially when learners struggle to correct them over a prolonged duration (Brodbeck et al., 1993; Ivancic & Hesketh, 1995/1996). Conversely, because deliberate errors are framed as intentional and expected parts of the learning process, they may incur less emotional cost. Of course, learners can still reap learning gains from effective study techniques despite experiencing greater frustration or stress during the learning process, as when studying incorrect worked examples (Richey et al., 2019) or taking practice tests (Wenzel & Reinhard, 2021; cf. Hinze & Rapp, 2014). Learners can also be taught to positively reappraise and even embrace discomfort and difficulties in learning for their personal growth (Woolley & Fishbach, 2022; Zepeda et al., 2020). But to the extent that the negative side effects of errors can be mitigated, learners may be more amenable to actively engaging with them. Moreover, better emotional self-regulation when learning from errors leads to better transfer (Keith & Frese, 2005). To pursue this line of questioning, future work could examine how deliberate erring impacts learners' emotional and motivational states and, in turn, their learning.

At the same time, effective learning strategies can only exert their benefits when they are actually used. If students do not recognize or believe that deliberate erring is beneficial and thus choose not to use this learning strategy, then they cannot profit from it. Indeed, this perplexing problem also plagues several other effective but underutilized learning techniques such as retrieval practice (e.g., Karpicke et al., 2009; Rivers, 2021), interleaving (e.g., Wong et al., 2020, 2021; Yan et al., 2016, 2017), and learning-by-teaching (e.g., Fiorella & Mayer, 2013; Lim et al., 2021; Wong et al., 2023). To support learners' self-regulated use of deliberate erring, it is worth investigating why they perceive this learning strategy as less effective. As our data clearly show, personal experience with deliberate erring is not enough to dispel learners' metacognitive illusions. Rather, based on the sources of learners' (mis)beliefs, future training protocols could explore how learners can be guided not only to acquire knowledge about effective strategies such as deliberate erring, but also to develop a belief that these strategies benefit them, while committing to and actually formulating concrete plans for implementing them in their learning (McDaniel & Einstein, 2020; McDaniel et al., 2021).

Conclusion

The wisdom of learning from our errors has long been incontrovertible, although deliberate errors are only beginning to be studied systematically. The present study is a first inquiry into enhancing mathematical learning using such errors. We found that deliberately committing and correcting procedural errors during problem-solving practice improved students' transfer of the learned mathematical procedures. This benefit generalized across transfer questions that were structurally more complex or embedded in "real-life" scenarios, and held over errorless repeated practice and spotting, explaining, and correcting others' errors in incorrect worked examples. Our findings highlight the rich and colorful nuances of different errorful learning methods and expand our repertoire of approaches to harness the power of errors for improving mathematical problem solving and transfer in education.

References

Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, *36*, 401–411. https://doi.org/10.1016/j.chb.2014.03.053

Anderson, J. R. (1976). *Language, memory, and thought*. Lawrence Erlbaum Associates.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Barbieri, C. A., & Booth, J. L. (2016). Support for struggling students in algebra: Contributions of incorrect worked examples. *Learning and Individual Differences*, *48*, 36–44. https://doi.org/10.1016/j.lindif.2016.04.001

Barbieri, C. A., & Booth, J. L. (2020). Mistakes on display: Incorrect examples refine equation solving and algebraic feature knowledge. *Applied Cognitive Psychology*, *34*(4), 862–878. https://doi.org/10.1002/acp.3663

Barbieri, C. A., Miller-Cotto, D., & Booth, J. L. (2019). Lessening the load of misconceptions: Design-based principles for algebra learning. *Journal of the Learning Sciences*, *28*(3), 381–417. https://doi.org/10.1080/10508406.2019.1573428

Barbieri, C. A., Miller-Cotto, D., Clerjuste, S. N., & Chawla, K. (2023). A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review*, *35*(1), Article 11. https://doi.org/10.1007/s10648-023-09745-1

Bassok, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 522–533. https://doi.org/10.1037/0278-7393.16.3.522

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 153–166. https://doi.org/10.1037/0278-7393.15.1.153

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*(5), 610–632. https://doi.org/10.1016/0749-596X(89)90016-8

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review*, *30*(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

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

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Booth, J. L. (2011). Why can't students get the concept of math? *Perspectives on Language and Literacy*, *37*(2), 31–35.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24–34. https://doi.org/10.1016/j.learninstruc.2012.11.002

Booth, J. L., McGinn, K. M., Young, L. K., & Barbieri, C. (2015). Simple practice doesn't always make perfect: Evidence from the worked example effect. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 24–32. https://doi.org/10.1177/2372732215601691

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*(1), 61–100. https://doi.org/10.3102/0091732X024001061

Brodbeck, F. C., Zapf, D., Prümper, J., & Frese, M. (1993). Error handling in office work with computers: A field study. *Journal of Occupational and Organizational Psychology*, *66*(4), 303–317. https://doi.org/10.1111/j.2044-8325.1993.tb00541.x

Cambridge Assessment International Education. (2022). *Cambridge international AS & A level mathematics (9709): Past papers, examiner reports* *and specimen papers*. https://www.cambridgeinternational.org/programmes-and-qualifications/cambridge-international-as-and-a-level-mathematics-9709/past-papers/

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Lawrence Erlbaum Associates.

Chi, M. T. H. (2022). The self-explanation principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 381–393). Cambridge University Press. https://doi.org/10.1017/9781108894333.040

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chi, M. T. H., & VanLehn, K. A. (2012). Seeing deep structure from the interactions of surface features. *Educational Psychologist*, *47*(3), 177–188. https://doi.org/10.1080/00461520.2012.695709

Corral, D., & Carpenter, S. K. (2020). Facilitating transfer through incorrect examples and explanatory feedback. *Quarterly Journal of Experimental Psychology*, *73*(9), 1340–1359. https://doi.org/10.1177/1747021820909454

Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, *47*(3), 153–176. https://doi.org/10.1080/00461520.2012.696438

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*(4), 552–568. https://doi.org/10.1016/j.jecp.2012.06.009

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1–24). Ablex Publishing.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, *22*(3), 206–214. https://doi.org/10.1016/j.learninstruc.2011.11.001

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review*, *27*(4), 567–586. https://doi.org/10.1007/s10648-015-9313-7

Fiorella, L. (2023). Making sense of generative learning. *Educational Psychology Review*, *35*(2), Article 50. https://doi.org/10.1007/s10648-023-09769-7

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, *38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., Hosp, M., & Jancek, D. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology*, *95*(2), 293–305. https://doi.org/10.1037/0022-0663.95.2.293

Galli, S., Chiesi, F., & Primi, C. (2008). The construction of a scale to measure mathematical ability in psychology students: An application of the Rasch Model. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, *15*(1), 3–18.

Galli, S., Chiesi, F., & Primi, C. (2011). Measuring mathematical ability needed for "non-mathematical" majors: The construction of a scale applying IRT and differential item functioning across educational contexts. *Learning and Individual Differences*, *21*(4), 392–402. https://doi.org/10.1016/j.lindif.2011.04.005

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612–634. https://doi.org/10.1016/j.learninstruc.2007.09.008

Hartmann, C., van Gog, T., & Rummel, N. (2021). Preparatory effects of problem solving versus studying examples prior to instruction. *Instructional Science*, *49*(1), 1–21. https://doi.org/10.1007/s11251-020-09528-z

Hartmann, C., van Gog, T., & Rummel, N. (2022). Productive versus vicarious failure: Do students need to fail themselves in order to learn? *Applied Cognitive Psychology*, *36*(6), 1219–1233. https://doi.org/10.1002/acp.4004

Haskell, R. E. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. Academic Press. https://doi.org/10.1016/B978-0-12-330595-4.X5000-8

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 290–296. https://doi.org/10.1037/a0028468

Heemsoth, T., & Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science*, *42*(4), 639–657. https://doi.org/10.1007/s11251-013-9302-5

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 22–34. https://doi.org/10.1037/0278-7393.29.1.22

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203063538

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology*, *28*(4), 597–606. https://doi.org/10.1002/acp.3032

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattray, C. (2000). Hypothesis testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, *10*(4), 361–391. https://doi.org/10.1016/S0959-4752(00)00004-9

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, *40*(4), 514–527. https://doi.org/10.3758/s13421-011-0167-z

Ivancic, K., & Hesketh, B. (1995/1996). Making the best of errors during training. *Training Research Journal*, *1*, 103–125.

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*(1), 48–59. https://doi.org/10.1037/a0021977

Kapur, M. (2014a). Comparing learning from productive failure and vicarious failure. *Journal of the Learning Sciences*, *23*(4), 651–677. https://doi.org/10.1080/10508406.2013.819000

Kapur, M. (2014b). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology*, *90*(4), 677–691. https://doi.org/10.1037/0021-9010.90.4.677

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, *92*(1), 109–129. https://doi.org/10.1037/0033-295X.92.1.109

Knowlton, B. J., Siegel, A. L. M., & Moody, T. D. (2017). Procedural learning in humans. In H. Eichenbaum (Ed.), *Memory systems, Volume 3 of Learning and memory: A comprehensive reference* (J. H. Byrne, Series ed.) (2nd ed., pp. 295–312). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21085-7

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*(2), 416–428. https://doi.org/10.3758/MC.36.2.416

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989–998. https://doi.org/10.1037/a0015729

Lim, K. Y. L., Wong, S. S. H., & Lim, S. W. H. (2021). The "silent teacher": Learning by teaching via writing a verbatim teaching script. *Applied Cognitive Psychology*, *35*(6), 1492–1501. https://doi.org/10.1002/acp.3881

Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition*, *43*(1), 14–26. https://doi.org/10.3758/s13421-014-0452-8

Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, *23*(11), 1337–1344. https://doi.org/10.1177/0956797612443370

Loibl, K., & Leuders, T. (2019). How to make failure productive: Fostering learning from errors through elaboration prompts. *Learning and Instruction*, *62*, 1–10. https://doi.org/10.1016/j.learninstruc.2019.03.002

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review*, *29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction*, *34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

Lorenzet, S. J., Salas, E., & Tannenbaum, S. I. (2005). Benefiting from mistakes: The impact of guided errors on learning, performance, and self-efficacy. *Human Resource Development Quarterly*, *16*(3), 301–322. https://doi.org/10.1002/hrdq.1141

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice*, *41*(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4

Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29–53). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203053270

McDaniel, M. A. (2007). Transfer: Rediscovering a central concept. In H. L. Roediger, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 267–270). Oxford University Press.

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork* (pp. 175–198). Psychology Press.

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review*, *1*(2), 113–145. https://doi.org/10.1007/BF01326639

McDaniel, M. A., & Einstein, G. O. (2020). Training learning strategies to promote self-regulation and transfer: The knowledge, belief, commitment, and planning framework. *Perspectives on Psychological Science*, *15*(6), 1363–1381. https://doi.org/10.1177/1745691620920723

McDaniel, M. A., Einstein, G. O., & Een, E. (2021). Training college students to use learning strategies: A framework and pilot course. *Psychology Learning & Teaching*, *20*(3), 364–382. https://doi.org/10.1177/1475725721989489

McGill, T. J., & Volet, S. E. (1997). A conceptual framework for analyzing students' knowledge of programming. *Journal of Research on Computing in Education*, *29*(3), 276–297. https://doi.org/10.1080/08886504.1997.10782199

McKeough, A., Lupart, J., & Marini, A. (Eds.). (1995). *Teaching for transfer: Fostering generalization in learning*. Lawrence Erlbaum. https://doi.org/10.4324/9781315044736

McLaren, B. M., Adams, D. M., & Mayer, R. E. (2015). Delayed learning effects with erroneous examples: A study of learning decimals with a web-based tutor. *International Journal of Artificial Intelligence in Education*, *25*(4), 520–542. https://doi.org/10.1007/s40593-015-0064-x

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. https://doi.org/10.3758/PBR.15.1.174

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, *138*, Article 107296. https://doi.org/10.1016/j.neuropsychologia.2019.107296

Metcalfe, J., & Xu, J. (2018). Learning from one's own errors and those of others. *Psychonomic Bulletin & Review*, *25*(1), 402–408. https://doi.org/10.3758/s13423-017-1287-7

Ohlsson, S. (1996a). Learning from error and the design of task environments. *International Journal of Educational Research*, *25*(5), 419–448. https://doi.org/10.1016/S0883-0355(97)81236-0

Ohlsson, S. (1996b). Learning from performance errors. *Psychological Review*, *103*(2), 241–262. https://doi.org/10.1037/0033-295X.103.2.241

Oser, F., & Spychiger, M. (2005). *Lernen ist schmerzhaft. Zur Theorie des Negativen Wissens und zur Praxis der Fehlerkultur* [Learning is painful. On the theory of negative knowledge and the practice of error culture]. Beltz.

Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, *11*(1), 63–68. https://doi.org/10.1111/1467-9280.00216

Pan, S. C., Sana, F., Samani, J., Cooke, J., & Kim, J. A. (2020). Learning from errors: Students' and instructors' practices, attitudes, and beliefs. *Memory*, *28*(9), 1105–1122. https://doi.org/10.1080/09658211.2020.1815790

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In T. Husén & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed., pp. 425–441). Pergamon Press.

Pillai, R. M., Loehr, A. M., Yeo, D. J., Hong, M. K., & Fazio, L. K. (2020). Are there costs to using incorrect worked examples in mathematics education? *Journal of Applied Research in Memory and Cognition*, *9*(4), 519–531. https://doi.org/10.1016/j.jarmac.2020.06.007

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text

comprehension, and arithmetic skills. *ZDM*, *52*(1), 33–44. https://doi.org/10.1007/s11858-019-01118-9
Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1023–1041. https://doi.org/10.1037/xlm0000637
Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–667. https://doi.org/10.1037/a0033194
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. https://doi.org/10.1111/cogs.12086
Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3
Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M. A. L., Star, J. R., Baker, R. S., & McLaren, B. M. (2019). More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education*, *139*, 173–190. https://doi.org/10.1016/j.compedu.2019.05.012
Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15. https://doi.org/10.1111/j.1467-8624.2006.00852.x
Rittle-Johnson, B. (2019). Iterative development of conceptual and procedural knowledge in mathematics learning and instruction. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 124–147). Cambridge University Press. https://doi.org/10.1017/9781108235631.007
Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*(1), 175–189. https://doi.org/10.1037/0022-0663.91.1.175
Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM Mathematics Education*, *49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z
Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x
Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346
Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review*, *33*(3), 823–862. https://doi.org/10.1007/s10648-020-09578-2
Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1
Santagata, R. (2005). Practices and beliefs in mistake-handling activities: A video study of Italian and US mathematics lessons. *Teaching and Teacher Education*, *21*(5), 491–508. https://doi.org/10.1016/j.tate.2005.03.004
Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology*, *46*(1), 178–192. https://doi.org/10.1037/a0016701
Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. https://doi.org/10.1037/a0025140
Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. https://doi.org/10.1207/s1532690xci2202_1
Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). Cambridge University Press. https://doi.org/10.1017/CBO9780511489709.002
Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, *91*(5), 761–798. https://doi.org/10.3102/00346543211019105
Skinner, B. F. (1958). Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self-instruction. *Science*, *128*(3330), 969–977. https://doi.org/10.1126/science.128.3330.969
Soncini, A., Visintin, E. P., Matteucci, M. C., Tomasetto, C., & Butera, F. (2022). Positive error climate promotes learning outcomes through students' adaptive reactions towards errors. *Learning and Instruction*, *80*, Article 101627. https://doi.org/10.1016/j.learninstruc.2022.101627
Steuer, G., Rosentritt-Brunn, G., & Dresel, M. (2013). Dealing with errors in mathematics classrooms: Structure and relevance of perceived error climate. *Contemporary Educational Psychology*, *38*(3), 196–210. https://doi.org/10.1016/j.cedpsych.2013.03.002
Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. Simon & Schuster.
Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205
Terwel, J., van Oers, B., van Dijk, I., & van den Eeden, P. (2009). Are representations to be provided or generated in primary mathematics education? Effects on transfer. *Educational Research and Evaluation*, *15*(1), 25–44. https://doi.org/10.1080/13803610802481265
Tulis, M. (2013). Error management behavior in classrooms: Teachers' responses to student mistakes. *Teaching and Teacher Education*, *33*, 56–68. https://doi.org/10.1016/j.tate.2013.02.003
Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1264–1269. https://doi.org/10.1037/a0023719
VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of cascade. *Journal of the Learning Sciences*, *8*(1), 71–125. https://doi.org/10.1207/s15327809jls0801_3
Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4
Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548
Wenzel, K., & Reinhard, M.-A. (2021). Does the end justify the means? Learning tests lead to more negative evaluations and to more stress experiences. *Learning and Motivation*, *73*, Article 101706. https://doi.org/10.1016/j.lmot.2020.101706
Wong, S. S. H. (2023). Deliberate erring improves far transfer of learning more than errorless elaboration and spotting and correcting others' errors. *Educational Psychology Review*, *35*(1), Article 16. https://doi.org/10.1007/s10648-023-09739-z
Wong, S. S. H., Chen, S., & Lim, S. W. H. (2021). Learning melodic musical intervals: To block or to interleave? *Psychology of Music*, *49*(4), 1027–1046. https://doi.org/10.1177/0305735620922595
Wong, S. S. H., Lim, K. Y. L., & Lim, S. W. H. (2023). To ask better questions, teach: Learning-by-teaching enhances research question generation more than retrieval practice and concept-mapping. *Journal*

of Educational Psychology, *115*(6), 798–812. https://doi.org/10.1037/edu0000802

Wong, S. S. H., & Lim, S. W. H. (2019). Prevention–permission–promotion: A review of approaches to errors in learning. *Educational Psychologist*, *54*(1), 1–19. https://doi.org/10.1080/00461520.2018.1501693

Wong, S. S. H., & Lim, S. W. H. (2022a). Deliberate errors promote meaningful learning. *Journal of Educational Psychology*, *114*(8), 1817–1831. https://doi.org/10.1037/edu0000720

Wong, S. S. H., & Lim, S. W. H. (2022b). The derring effect: Deliberate errors enhance learning. *Journal of Experimental Psychology: General*, *151*(1), 25–40. https://doi.org/10.1037/xge0001072

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. https://doi.org/10.1177/0022429420908312

Wong, S. S. H., Ng, G. J. P., Tempel, T., & Lim, S. W. H. (2019). Retrieval practice enhances analogical problem solving. *Journal of Experimental Education*, *87*(1), 128–138. https://doi.org/10.1080/00220973.2017.1409185

Woolley, K., & Fishbach, A. (2022). Motivating personal growth by seeking discomfort. *Psychological Science*, *33*(4), 510–523. https://doi.org/10.1177/09567976211044685

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*(7), 918–933. https://doi.org/10.1037/xge0000177

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied*, *23*(4), 403–416. https://doi.org/10.1037/xap0000139

Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1073–1092. https://doi.org/10.1037/xlm0000363

Zepeda, C. D., Martin, R. S., & Butler, A. C. (2020). Motivational strategies to engage learners in desirable difficulties. *Journal of Applied Research in Memory and Cognition*, *9*(4), 468–474. https://doi.org/10.1016/j.jarmac.2020.08.007

Received July 7, 2023
Revision received November 23, 2023
Accepted November 28, 2023 ■