Educational Psychology Review (2023) 35:16
https://doi.org/10.1007/s10648-023-09739-z

INTERVENTION STUDY

Deliberate Erring Improves Far Transfer of Learning More Than Errorless Elaboration and Spotting and Correcting Others' Errors

Sarah Shi Hui Wong[1]

Accepted: 21 December 2022

Abstract

Transfer of learning is a fundamental goal of education but is challenging to achieve, especially where far transfer to remote contexts is at stake. How can we improve learners' flexible application of knowledge to distant domains? In a counterintuitive phenomenon termed the *derring effect*, deliberately committing and correcting errors in low-stakes contexts enhances learning more than avoiding errors. Whereas this benefit has been demonstrated with tests in domains similar to those in the initial learning task, the present set of three experiments ($N = 120$) investigated whether deliberate erring boosts far transfer of conceptual knowledge to dissimilar domains. Undergraduates studied scientific expository texts either by generating conceptually correct responses or by deliberately generating conceptually erroneous responses then correcting them. Deliberate erring improved not only retention (Experiment 1), but also far transfer on inferential test questions that required applying the learned concepts to remote knowledge domains (e.g., from biology/vaccines to geography/forest management techniques; Experiment 2). This advantage held even over a control that further involved spotting and correcting the same errors that one's peers had deliberately made (Experiment 3). Yet, learners failed to predict or recognize the benefits of deliberate erring even after the test. Altogether, these results suggest that the derring effect is specific to generating incorrect, but not correct, elaborations. Neither does mere exposure to others' errors nor juxtaposing these errors with the correct responses suffice. Rather, guiding learners to personally commit and correct deliberate errors is vital for enhancing generalization and far transfer of learning to distant knowledge domains.

Keywords Deliberate errors · Elaboration · Error correction · Learning from errors · Metacognition · Transfer

✉ Sarah Shi Hui Wong
psywshs@nus.edu.sg

[1] Department of Psychology, Faculty of Arts & Social Sciences, National University of Singapore, Block AS4, 9 Arts Link, Singapore 117570, Singapore

Published online: 07 February 2023

What do vaccines have in common with interventions against persuasive attacks and fake news? Widely considered one of the most important innovations in human history, vaccines confer immunity to disease through deliberate exposure to weakened strains of the disease-causing pathogen. These strains do not cause a full-blown infection, but stimulate the body's production of antibodies to guard against future attacks. Notably, this immunization analogy can also be applied to many areas of human life beyond the biomedical domain. For instance, inoculation theory in attitudinal research holds that cognitive resistance against persuasive attacks can be developed by exposing individuals to weakened doses of counterarguments (McGuire, 1961). Likewise, some recent psychological interventions against misinformation involve exposing individuals to misinformation techniques by having them deliberately create fake news content in an online game setting (Ecker et al., 2022; Roozenbeek and van der Linden, 2019).

Embedded in all these instances across diverse domains is the fundamental idea that deliberate erring—intentionally engaging with and even producing "incorrect" responses—is useful in low-stakes contexts. This analogy has recently been fruitfully applied in educational contexts too. In a counterintuitive phenomenon known as the *derring effect*, deliberately committing and correcting errors has been shown to yield surprising benefits for learning (Wong and Lim, 2022a, 2022b), contrary to the traditional view that errors should be avoided at all costs during learning so that they are not ingrained and repeated in the future (e.g., Ausubel, 1968; Bandura, 1986; Skinner, 1958).

Learning From (Deliberate) Errors

Indeed, growing evidence has revealed that errors can facilitate new learning when accompanied by corrective feedback (for reviews, see Kapur, 2016; Mera et al., 2022; Metcalfe, 2017; Wong and Lim, 2019b). As delineated in Wong and Lim's (2019b) Prevention–Permission–Promotion (3P) framework, errors can be approached in different ways to optimize learning in various educational contexts, such as observing, allowing, inducing, or even guiding errors when not avoiding them entirely. For instance, errors are observed when studying others' incorrect solutions in worked examples or case studies (e.g., Adams et al., 2014; Barbieri and Booth, 2020; Booth et al., 2013; Durkin and Rittle-Johnson, 2012; Große and Renkl, 2007; Heemsoth and Heinze, 2014; Joung et al., 2006), whereas errors are passively allowed to occur naturalistically during exploration in discovery learning (e.g., Alfieri et al., 2011; Bruner, 1961; Kirschner et al., 2006; Klahr and Nigam, 2004). Alternatively, in errorful generation (e.g., Kang et al., 2011; Kornell et al., 2009; Potts and Shanks, 2014), productive failure (e.g., Kapur, 2008, 2016; Kapur and Bielaczyc, 2012), and error management training (e.g., Keith and Frese, 2008), errors are actively induced through purposefully adding challenge to the task or withholding instruction, and encouraging learners to guess or produce the correct answers. In such situations, errors inevitably arise because learners lack knowledge of the solutions despite their best intentions to derive them. Conversely, deliberate erring involves guiding learners to intentionally err, even when they are aware of the correct responses during open-book study (Wong and Lim, 2022a, 2022b).

In the first demonstration of the derring effect (Wong and Lim, 2022b), students were asked to learn scientific term-definition concepts (e.g., “Adaptation is a trait that increases the probability that an individual will leave offspring in subsequent generations.”) by deliberately generating conceptually incorrect definitions with or without correction, or copying and underlining them. On a cued-recall test (e.g., “What is adaptation?”), deliberate erring enhanced memory of the learned concepts more than errorless copying, with error correction conferring an additional benefit. Moreover, this advantage of deliberate error commission and correction persisted over more rigorous errorless learning techniques that involved actively generating alternative conceptually correct responses (i.e., conceptual synonyms) or even further elaborating on the concepts by generating examples.

Of note, the benefits of deliberate erring are not confined to memory retention, but have been extended to higher-order application of knowledge. In a subsequent study, Wong and Lim (2022a) pitted deliberate error commission and correction against a range of popular errorless learning techniques: copying with underlining, concept-mapping, and generating conceptual synonyms. Undergraduates studied scientific expository texts using these techniques, and were then tested on their ability to apply the material to analyze a novel news event on a related topic. For instance, after studying a text passage on “food allergies,” learners were presented with a news article that described an incident of a young boy who had suffered a life-threatening allergic reaction. They were then asked to apply their knowledge by developing predictions about the boy’s medical history, and proposing potential causes for his reported symptoms to formulate a diagnosis. The key finding was that deliberate erring produced superior application performance than all the errorless learning techniques.

Taken together, extant research reveals that deliberate erring is a promising technique that has improved a range of basic to more complex learning outcomes. To date, however, the effects of deliberate erring have been investigated with criterion tasks involving responses that are identical or similar to those that had been previously learned (e.g., recalling the definition of a scientific term or applying one’s learned knowledge about food allergies to analyze a case study in the same domain). A pertinent question for researchers and educators is whether deliberate erring can enhance the flexible application of knowledge to new contexts requiring responses that are dissimilar from those that had been learned. That is, to what extent does deliberate erring improve far transfer to different knowledge domains?

Transfer of Learning

The quest for transfer has inspired more than 100 years of research on this topic since Thorndike and Woodworth’s (1901) seminal studies. Transfer is a form of higher-order or meaningful learning (Anderson and Krathwohl, 2001; Bloom, 1956; Mayer, 2002) that involves applying one’s prior knowledge or skills in new contexts (Day and Goldstone, 2012; Detterman, 1993; Haskell, 2001; Perkins and Salomon, 1992, 2012). As outlined in Barnett and Ceci's (2022) taxonomy, transfer can be viewed as a continuum from "near" to "far" transfer along content (i.e., the nature of the learned skill, performance change, and memory demands of the transfer task) and context (i.e., knowledge domain, physical context, temporal context, functional context, social context, and modality) dimensions. Of particular interest here is transfer along the knowledge domain dimension. For instance, near transfer along knowledge domains could involve learning about early forms of inoculation against smallpox, then applying this knowledge to the closely related context of modern vaccines for COVID-19. In contrast, far transfer could involve adapting the concept of biomedical immunization to a markedly dissimilar domain such as digital media literacy interventions to combat misinformation.

Enhancing learners' ability to transfer their knowledge is widely regarded as an ultimate goal that most formal education aspires toward (Fisher et al., 2016; McDaniel, 2007; McKeough et al., 1995; Perkins and Salomon, 1992). In the OECD's (2019) Learning Compass 2030, for instance, transfer of learning across domains has been viewed as vital for developing interdisciplinary knowledge to understand and solve complex problems. Indeed, real-world problems are often "wicked" in that they are complex, ill-defined, and defy simple resolution (Rittel and Webber, 1973). Without engaging in cross-domain transfer, one would miss out on valuable opportunities to "connect the dots" and integrate knowledge across disciplines, beyond merely thinking within silos (Epstein, 2019).

Yet, far transfer is notoriously difficult to attain and has been argued to occur only rarely, whereas near transfer occurs more often (Detterman, 1993; Sala and Gobet, 2017). Successful transfer draws on an interplay of at least three processes: (a) recognizing that previously learned knowledge is relevant to the new context at hand, (b) correctly recalling that knowledge, and (c) effectively applying that knowledge in the new context (Barnett and Ceci, 2002; see also Nokes, 2009; Perkins and Salomon, 2012). Thus, transfer failures could stem from breakdowns in one or more of these processes. For instance, spontaneous transfer is rare because learners often fail to recognize that a new context is structurally similar to one that they had previously learned (e.g., Anolli et al., 2001; Gick and Holyoak, 1980, 1983). Accordingly, transfer can be facilitated by explicitly prompting learners to consider the previously learned information (e.g., Gick and Holyoak, 1980). Alternatively, when the prior and new contexts share a high level of surface similarity (i.e., common literal or superficial features) and structural similarity (i.e., common causal relations), correspondences between them are more likely to be noticed and mapped (Gentner et al., 1993; Holyoak and Koh, 1987; for a review, see Day and Goldstone, 2012). For instance, source problems or worked examples are more likely to facilitate transfer to new problems when their shared structural features or sub-goals are made salient (e.g., Atkinson et al., 2000; Catrambone and Holyoak, 1989, 1990; Quilici and Mayer, 1996; Renkl, 2005).

However, simply knowing that particular contexts are related is insufficient to produce transfer, which also hinges critically on learners' ability to successfully recall and apply their relevant prior knowledge. Indeed, even when learners are explicitly informed that previously solved problems should help them in solving later ones and are further allowed to consult those solutions, their transfer performance may remain limited (Reed et al., 1985). In particular, this may occur when initial learning of the source information is shallow (Chi and VanLehn, 2012; Salomon and Perkins, 1989). With deeper initial learning, the likelihood of positive transfer to a structurally similar task is increased (Kimball and Holyoak, 2000). For instance, when learners actively generate or invent their own solutions to source problems rather than simply read and practice those solutions, they are more likely to abstract and recall the problems' deep structure, in turn transferring better to semantically unrelated but structurally similar topics (Schwartz et al., 2011; see also Needham and Begg, 1991).

Of course, remembering relevant prior knowledge alone does not guarantee the flexible transfer of that knowledge (e.g., Agarwal, 2019; Anolli et al., 2001; Michael et al., 1993; Wong and Lim, 2019a), although it is likely helpful (Butler, 2010; Dunlosky et al., 2013; Schwartz et al., 2011). For instance, it would prove difficult—if not impossible—for learners who have forgotten how vaccines confer immunity to apply this concept to develop analogous interventions for countering misinformation, even if they recognized that both domains can potentially be linked in some way. Thus, to the extent learners are aware that their prior knowledge is relevant to a novel context and are able to recall that knowledge, they would be better poised for transfer success.

Deliberate Erring and Far Transfer

How can learners be guided toward deep initial learning to increase their odds of transfer? In view that deliberately committing and correcting errors enhances recall and meaningful application of learned information within the same knowledge domain (i.e., near transfer; Wong and Lim, 2022a, 2022b), it holds promise for promoting deep learning that is crucial for far transfer across knowledge domains.

Although the precise mechanisms underlying the derring effect have yet to be fully specified, some theoretical accounts have been proposed based on the contributions of both deliberate error commission and correction. For one, exploring what a concept is *not* during error commission may more effectively cull those incorrect responses for better learning (e.g., Kornell et al., 2009), relative to considering what *else* the concept is when generating correct elaborations (see Gartmeier et al., 2008; Minsky, 1994; Oser and Spychiger, 2005). It may also be that generating errors enhances encoding of their subsequent correction (e.g., Hays et al., 2013; Kornell et al., 2009; Potts et al., 2019), such that correcting one's deliberate errors has been found to produce better learning than leaving them uncorrected (Wong and Lim, 2022b). For instance, deliberate erring may draw learners' attention to the target response during error correction and foster an episodically memorable event (e.g., Metcalfe and Huelser, 2020), whereas errorless learning may not afford such processing when only correct information is encountered during study. Relatedly, the juxtaposition of one's deliberate errors with their correction may highlight a concept's diagnostic features and build richer mental models for abstraction, which is crucial for concept acquisition and transfer (e.g., Corral and Carpenter, 2020; Gick and Holyoak, 1983). In tandem, these mechanisms suggest that deliberate erring could potentially foster deep learning that enables learners to effectively recall and apply relevant knowledge for better far transfer, relative to preventing learners from making errors.

The Present Study

In three experiments, the present study tested whether deliberately committing and correcting errors promotes far transfer of learned concepts to distant knowledge domains. To the extent learners recognize that previously learned concepts are related to a novel domain, the hypothesis was that deliberate erring would enhance their ability to recall and flexibly apply those concepts more than errorless learning. Using a set of scientific expository texts as educationally relevant materials, Experiment 1 aimed foremost to replicate the basic derring effect (Wong and Lim, 2022a, 2022b) in demonstrating that deliberate erring produces superior recall than errorless learning. As potential covariates that could influence their test performance, learners' English proficiency and need for cognition (Cacioppo et al., 1983, 1984) were measured. Experiment 2 assessed whether deliberate erring would further enhance far transfer to conceptual inferential questions from different knowledge domains that required responses dissimilar to those that had been previously studied (e.g., applying one's learned knowledge on vaccines to make inferences about related concepts in the distant domain of forest management techniques). Experiment 3 sought to replicate Experiment 2's findings, while further probing the mechanisms beneath the derring effect. Specifically, Experiment 3 investigated whether the learning benefits of deliberate erring are simply due to mere exposure to errors and/or juxtaposition of correct versus incorrect responses, even if learners had not committed those errors themselves.

A related question of interest was whether learners would be able to accurately predict and monitor the effects of deliberate erring on their test performance. This question is educationally vital because accurate metacognitive knowledge causally impacts students' self-regulated learning when they select study strategies that are most effective for them (i.e., metacognitive control; Metcalfe and Finn, 2008; Thiede et al., 2003). Yet, extant research has found that students are often unaware of the benefits of errors for their learning (e.g., Huelser and Metcalfe, 2012; Wong and Lim, 2022a, 2022b; Yang et al., 2017). Thus, in all three experiments here, a within-subjects design was used to assess the relative actual versus predicted effectiveness of the learning methods for each learner. After having the opportunity to personally experience both deliberate erring and the control learning method, learners made metacognitive judgments of their learning before and after being tested on their recall or far transfer performance.

Experiment 1

As successfully recalling previously learned knowledge is crucial for applying it in a structurally similar transfer task (Barnett and Ceci, 2002; Kimball and Holyoak, 2000), Experiment 1 tested and replicated the basic derring effect in memory retention (Wong and Lim, 2022a, 2022b) using the present set of educationally relevant scientific expository texts on “cyclones” and “vaccines”. Since deliberate erring involves generating novel incorrect responses, a suitably competitive errorless control should also induce active generation and elaboration, albeit of correct rather than incorrect responses. Indeed, it is well-established that information is remembered better when it has been actively generated than passively read (i.e., the *generation effect*; Bertsch et al., 2007; Jacoby, 1978; Slamecka and Graf, 1978). Moreover, when learners actively construct meaning from the material by elaborating on and integrating it with their prior knowledge, deeper learning ensues (Chi, 2009; Fiorella and Mayer, 2016; Levin, 1988; Pressley et al., 1987; Wittrock, 1974). Hence, Experiment 1 compared deliberate error commission and correction (*concept-error* method) against an errorless control in which learners generated alternative conceptually correct elaborations (*concept-synonym* method). In this way, both learning methods were exactly matched except for the production of incorrect versus correct responses, respectively.

A within-subjects design was used, whereby learners studied each expository text using one of the learning methods, respectively, and made a judgment of learning (JOL) to predict how much of the material they would remember. Then, learners were tested on their recall of both texts, and further rated the effectiveness of both learning methods after the test. As potential covariates, learners’ English proficiency and need for cognition (Cacioppo et al., 1984) were measured.

Method

Participants

The participants were 40 undergraduates (28 were female) between the ages of 19 and 24 ($M = 20.42$, $SD = 1.32$) from the National University of Singapore. Based on the effect size reported by Wong and Lim (2022a, Experiment 2) for the recall advantage of deliberate erring over generating conceptual synonyms for expository texts ($d = 0.51$), a power analysis (G*Power; Faul et al., 2007) indicated that at least 34 participants were required to detect a medium within-subjects effect ($d = 0.50$) in the present experiments at 80% power and $\alpha = .05$.

Across all experiments, all participants reported English as their first language and received either course credit or cash reimbursement for their participation. All experiments were conducted with ethics approval from the university’s institutional review board, and participants granted their written informed consent.

Design

The sole factor in this within-subjects design was learning method: *concept-error* (deliberate error generation and correction) versus *concept-synonym* (errorless generation and elaboration). Participants used both learning methods in a counterbalanced order and within the same studying duration for each method. The learning outcome of interest was participants' recall performance, as assessed by the number of idea units from the study texts that they correctly recalled at test.

Materials

Study Texts The study texts were two scientific expository passages adapted from Butler (2010) on the topics of "cyclones" and "vaccines". Each text was edited to contain exactly 320 words arranged in four paragraphs comprising 20 sentences. For scoring purposes, 45 idea units were identified in each study text. For instance, an idea unit in the "cyclones" text was "Cyclones are storm systems," whereas an idea unit in the "vaccines" text was "Vaccines are biological preparations". The study texts had Flesch-Kincaid grade levels of 10 and 13, respectively. In addition, two brief 29-word paragraphs each comprising three sentences served as the practice texts. Both practice texts on "muscle tissue" and "the human ear" (adapted from Karpicke and Blunt, 2011) did not relate to either of the critical study texts. All practice and study texts are available in the online supplementary materials.

Post-Learning Questionnaires A 4-item post-learning questionnaire adapted from Wong and Lim (2022a) was administered after participants had studied each study text. Specifically, participants (a) made a judgment of learning (JOL) to predict how much of the material from the study text they would remember later on an 11-point scale from 0 to 100% (i.e., 0%, 10%, 20%, … 100%), (b) rated how interesting the study text was (1 = *not at all*; 7 = *extremely*), (c) rated how understandable the study text was (1 = *not at all*; 7 = *extremely*), and (d) indicated how well they knew the subject matter covered in the study text prior to reading it (1 = *not very well*; 7 = *very well*). After participants had completed the test phase, they rated how effective they thought each learning method had been in helping them learn the study text (1 = *not at all*; 7 = *extremely*).

English Language Proficiency Test As in Wong and Lim (2022a), participants' English language proficiency was assessed as a potential covariate through 10 questions adapted from the verbal reasoning section of the Graduate Record Examinations (GRE). The maximum possible score was 10.

Need for Cognition Scale As another potential covariate, participants' need for cognition was measured using the 18-item short form of the Need for Cognition Scale, which assesses "an individual's tendency to engage in and enjoy effortful cognitive endeavors" (Cacioppo et al., 1984, p. 306). Need for cognition has been found to positively predict participants' recall of message content (e.g., Cacioppo et al., 1983). Sample scale items include "I would prefer complex to simple problems" and "Learning new ways to think doesn't excite me very much" (reverse-scored). All items were rated on a 7-point scale (1 = *strongly disagree*; 7 = *strongly agree*). Nine of the items were negatively worded, and were reverse-scored prior to analyses such that higher scores indicated higher need for cognition. A mean need for cognition score was then computed for each participant by averaging their ratings across all 18 items. The scale demonstrated high internal consistency in the present experiment, Cronbach's $\alpha = .91$.

Procedure

Before attending the experiment, all participants completed the English proficiency test and Need for Cognition Scale via an online questionnaire. Upon arriving at the laboratory, participants were informed that they would be studying scientific texts and that they would later be tested on the material. The exact nature of the tests was not disclosed. Participants then underwent three experimental phases: practice, studying, and test. Participants were run in small groups of up to six per session, and performed the task individually using pen and paper while seated separately. The total experimental duration was approximately 90 min.

Practice Phase All participants were first instructed on the concept-synonym and concept-error methods (adapted from Wong and Lim, 2022a), and practiced using each method to study the practice texts. In the concept-synonym condition, participants wrote down each sentence in the text such that it contained a conceptual synonym (i.e., an alternative word or phrase that had the same meaning as the actual concept), underlined the synonym they had generated, then wrote down the actual concept exactly as it was presented in the text. For instance, given the training example "Bats are mammals that fly," a sample response was "Bats are <u>warm-blooded animals with fur</u> (mammals) that fly".

The concept-error method was identical to the concept-synonym method, except participants intentionally generated errors instead. Specifically, participants deliberately erred by writing down each sentence in the text such that it contained a plausible conceptual error (i.e., an error in understanding or interpreting a concept), struck out the error they had generated, then corrected it by writing down the actual concept exactly as it was presented in the text. For instance, given the training example "Bats are mammals that fly," a sample response was "Bats are ~~birds~~ (mammals) that fly". The act of striking out—drawing a line across—one's errors was behaviorally comparable to the act of underlining one's synonyms in the concept-synonym condition. To ensure that participants understood what was required of them during the practice phase, they were shown examples of conceptual versus non-conceptual errors (e.g., "Batz are mammals that fly" merely involves a spelling error, but is otherwise conceptually correct). As in Wong and Lim (2022a), participants were also encouraged to deliberately make plausible conceptual errors (i.e., responses that were objectively wrong but still believable), in line with previous research on competitive incorrect responses (e.g., Kang et al., 2011; Little and Bjork, 2015; Little et al., 2012). For instance, one would consider "Bats are birds that fly" as a more plausible conceptual error than "Bats are humans that fly".

Studying Phase After practicing both learning methods, participants began the studying phase. The order in which participants used both learning methods, as well as the pairing of study texts with learning methods, was counterbalanced. In both the concept-synonym and concept-error conditions, participants were given 1.5 min to read the given study text (either "cyclones" or "vaccines"), then used the specified learning method in the same way that they had done during the practice phase to study that text for 25 min (e.g., Wong and Lim, 2022a). Participants were also told that if they finished before the time was up, they should review their response to ensure they had included all 20 sentences from the text. Thus, the total studying duration was equated across both learning conditions. At the end of the study period, participants responded to the post-learning questionnaire—they made a JOL, rated the text's interestingness and understandability, and indicated their prior knowledge of the text content. The same procedure was repeated for the second study text.

Test Phase After studying both texts, participants were allowed to take a brief self-paced break before beginning the test phase, during which they were tested on their memory for both texts in the same order that they had been learned. Specifically, participants were asked to write down as much as they could remember from each study text. This procedure enabled a test of participants' memory for the texts' content in their entirety. No time limit was imposed. After completing the recall test, participants rated the effectiveness of each learning method. Finally, all participants were debriefed and thanked.

Results

Scoring

Participants' recall test performance was scored as the number of idea units from the study texts that they correctly recalled, out of a total possible score of 45 for each text. Two raters independently scored 10 of the 40 (25%) scripts, intraclass correlation (ICC) = .99, 95% CI [.97, .99], based on a two-way random-effects model. Discrepancies were reviewed and resolved through discussion to reach 100% agreement. Given the excellent interrater reliability, the remaining scripts were scored by one rater.

Preliminary Analyses

Participants' English language proficiency scores ($M = 3.18$, $SD = 1.43$) and need for cognition scores ($M = 3.93$, $SD = 0.95$) did not significantly correlate with their recall test performance in both the concept-synonym condition, $r(38) = .28$ and .28, both ps = .08, and concept-error condition, $r(38) = .25$ and .004, $p = .12$ and .98, respectively. Thus, neither English proficiency nor need for cognition was significantly associated with learners' knowledge retention.

Participants reported minimal familiarity with the study texts, as indicated by their relatively low prior knowledge ratings across the concept-synonym ($M = 2.45$, $SD = 1.54$) and concept-error ($M = 2.55$, $SD = 1.68$) conditions. There was no significant difference in participants' prior knowledge ratings across both conditions, $t(39) = -0.27$, $p = .79$, 95% $CI_{\text{mean difference}}$ [−0.85, 0.65]. In addition, participants' ratings of how interesting the study texts were did not differ across the concept-synonym ($M = 3.73$, $SD = 1.52$) and concept-error ($M = 3.63$, $SD = 1.69$) conditions, $t(39) = 0.40$, $p = .69$, 95% $CI_{\text{mean difference}}$ [−0.41, 0.61]. Neither was there a significant difference in participants' ratings of the study texts' understandability across the concept-synonym ($M = 4.95$, $SD = 1.45$) and concept-error ($M = 4.70$, $SD = 1.40$) conditions, $t(39) = 1.02$, $p = .31$, 95% $CI_{\text{mean difference}}$ [−0.25, 0.75].

Recall Test Performance

As predicted, participants recalled significantly more idea units from the study text in the concept-error condition ($M = 11.60$, $SD = 6.64$) than the concept-synonym condition ($M = 9.13$, $SD = 5.66$), $t(39) = -2.98$, $p = .005$, $d = 0.47$, 95% $CI_{\text{mean difference}}$ [−4.16, −0.79]. Thus, deliberately committing and correcting errors yielded superior memory performance than elaborating on the text by generating alternative correct responses.

Metacognitive Judgments of Learning

Yet, participants inaccurately predicted in their JOLs that they would remember just as much of the study material across the concept-synonym ($M = 41.00$, $SD = 22.85$) and concept-error ($M = 41.25$, $SD = 20.15$) conditions, $t(39) = -0.09$, $p = .93$, 95% $CI_{\text{mean difference}}$ [−5.65, 5.15]. This metacognitive illusion persisted even after the recall test—despite experiencing the benefits of deliberate erring for their memory performance, participants still believed that the concept-synonym method ($M = 4.03$, $SD = 1.49$) had been just as effective as the concept-error method ($M = 3.65$, $SD = 1.51$), $t(39) = 1.13$, $p = .27$, 95% $CI_{\text{mean difference}}$ [−0.30, 1.05].

Overall, 25 out of 40 learners (63%) benefited more from the concept-error than concept-synonym method at test. Conversely, 23 out of 40 learners (58%) predicted in their JOLs that the concept-error method would be just as effective as or even less effective than the concept-synonym method. Even after the recall test, the advantage of deliberate erring went largely unappreciated—24 out of 40 learners (60%) rated the concept-error method as just as effective as or even less effective than the concept-synonym method for their test performance. Taken together, participants' pre-test and post-test metacognitive judgments were at odds with how effective both learning methods had been. Table 1 shows the number of participants who actually performed better after deliberate erring than generating conceptual synonyms, the number who showed the opposite pattern, and the number who performed similarly across both conditions. For each of these three performance outcomes, Table 1 also shows the number of participants who made the corresponding pre-test metacognitive predictions (JOLs) and post-test metacognitive judgments (effectiveness ratings).

Table 1 Frequency count (and percentage) of participants showing different patterns of metacognitive ratings and actual recall test performance (Experiment 1)

Metacognitive ratings vs. actual performance	Performance outcome		
	Error > synonym	Error = synonym	Error < synonym
Metacognitive ratings			
Pre-test predictions (JOLs)	17 (43%)	6 (15%)	17 (43%)
Post-test judgments (effectiveness ratings)	16 (40%)	4 (10%)	20 (50%)
Actual recall test performance	25 (63%)	2 (5%)	13 (33%)

N = 40.

Discussion

Replicating the basic derring effect (Wong and Lim, 2022a, 2022b), Experiment 1 showed that deliberately committing and correcting errors improved learners' recall more than generating conceptual synonyms. Because both learning methods were experimentally identical except for the generation of incorrect versus correct elaborations, the present results suggest that generation and/or elaboration processes alone cannot fully explain the derring effect. In other words, to boost knowledge retention, it does not suffice to generate any novel elaboration. Rather, one must specifically generate *incorrect* elaborations to gain the mnemonic benefits observed here.

However, learners tended to be unaware of the benefits of deliberate erring and, on overall, incorrectly predicted the relative effectiveness of the concept-error and concept-synonym methods. Even after experiencing the effects of both methods for their recall performance, learners continued to underestimate the efficacy of deliberate erring. These metacognitive miscalibrations echo those observed in previous research (Wong and Lim, 2022a, 2022b), and align with broader findings that learners often fail to recognize the benefits of spontaneous errors or incorrect guesses too (e.g., Huelser and Metcalfe, 2012; Yang et al., 2017).

Experiment 2

Experiment 2 aimed to test whether deliberate erring further improves far transfer in applying one's learned conceptual knowledge to different domains, beyond enhancing recall. Although recalling relevant knowledge may facilitate transfer, it is not sufficient. For instance, one might be able to recall the biological processes that vaccines invoke in stimulating immunity, but fail to effectively apply this knowledge to the distant domain of developing forest management techniques to prevent wildfires. This would imply that such acquired knowledge is "inert," in that it is available but not used (Renkl et al., 1996). Conversely, if deliberate erring produces superior far transfer over errorless learning, then the derring effect is not confined to retention or verbatim reproduction of learned responses, but also includes meaningful application in loosely related domains.

To this end, Experiment 2 assessed learners' far transfer using inferential short-answer questions in different knowledge domains (e.g., Butler, 2010). Specifically, learners were required to use the concepts that they had studied using the concept-error or concept-synonym method to make inferences about related concepts in markedly dissimilar domains, thus serving as a measure of far transfer along the knowledge domain dimension in Barnett and Ceci's (2002) taxonomy. Concurrently, this procedure enabled an investigation of the effects of deliberate erring using a different test format—short-answer questions—that is frequently employed in the classroom.

Method

Participants

The participants were 40 undergraduates (26 were female) between the ages of 19 and 25 ($M = 20.98$, $SD = 1.66$) from the National University of Singapore who did not take part in Experiment 1. A power analysis (G*Power; Faul et al., 2007) indicated that this sample size afforded sufficient sensitivity to detect medium within-subjects effects ($d \geq 0.45$) for two-tailed pairwise comparisons at 80% power and $\alpha = .05$.

Design

As in Experiment 1, the single within-subjects factor was learning method: *concept-error* (deliberate error generation and correction) versus *concept-synonym* (errorless generation and elaboration). The dependent variable of interest was learners' far transfer performance on a test with inferential questions in different knowledge domains.

Materials and Procedure

The materials and procedure were identical to those in Experiment 1, except learners completed a far transfer test instead of a recall test. The far transfer test comprised four inferential short-answer questions per study text, for a total of eight test questions (adapted from Butler, 2010; available in the online supplementary materials). Each question targeted a key concept that learners had to abstract across multiple sentences in the study text, as opposed to isolated facts that were presented within a single sentence only. Crucially, the far transfer test questions required learners to apply the concepts that they had acquired during study to make inferences about related concepts in different knowledge domains. A sample test question for the "cyclones" text was:

> Although cars are powered by a different energy source (gasoline/petrol) than cyclones, the process that drives a car's engine is essentially the same as the

process that powers the spinning vortex of a cyclone. What is the process that is responsible for spinning the engine components of a car?
Answer: A cyclone's primary energy source is the heat release of condensation from water vapor in the warm air that it sucks up from the ocean. In a car engine, gasoline/petrol is burned inside the cylinders, giving rise to a tremendous amount of heat, and this heat does the work of spinning the engine components.

Likewise, a sample test question for the "vaccines" text was:

Controlled burning of dry bush is a forest management technique that is used to prevent wildfires, and relies on the same principle as the practice of inoculation in vaccinating people. How does controlled burning work?
Answer: Inoculation is the practice of deliberate infection that produces a small, localized infection. Similarly, controlled burning involves setting small fires under controlled conditions that eliminate the dry bush that fuels wildfires and limits the risk of the fire spreading out of control.

As in Butler (2010), each test question briefly mentioned the relevant concept from the study text. This procedure precluded the need for learners to spontaneously recognize that the study text was relevant to the novel knowledge domain specified in the test question (e.g., Gick and Holyoak, 1980, 1983), thus enabling a direct investigation of whether learners could retrieve the learned concepts from memory and apply them effectively in an unfamiliar context, rather than their ability to spontaneously detect analogies between the seemingly disparate contexts.

Learners completed the far transfer test without reference to both study texts. The test questions were presented blocked by study text in the order that they had been learned during the studying phase. No time limit was imposed for the test.

Results

Scoring

Participants' far transfer performance was scored by awarding one point for each test question that they answered correctly. Following Butler's (2010) scoring procedure, each response was scored either as correct or incorrect, with a maximum possible score of 4 for each learning condition. To be considered correct, a response had to effectively apply the relevant concepts from the study text to the specific novel domain described in the question, and had to contain the key ideas from the idealized correct answer. For instance, for the sample "vaccines" test question presented earlier, a response would not be awarded any points if it simply recalled what inoculation involves without applying this concept to the domain of "forest management techniques" that the question had specified. Furthermore, a response would be considered inadequate if it merely stated that controlled burning involves setting small fires without clearly explaining that this eliminates the dry bush that fuels wildfires and limits the risk of the fire spreading out of control.

As in Experiment 1, two raters independently scored 10 of the 40 (25%) scripts. Interrater reliability was excellent, ICC = .99, 95% CI [.95, .99], based on a two-way random-effects model. Discrepancies were reviewed and resolved through discussion to reach 100% agreement. Given the high interrater reliability, one rater scored the remaining scripts.

Preliminary Analyses

Learners' English language proficiency scores (*M* = 3.58, *SD* = 1.92) and need for cognition scores (*M* = 4.22, *SD* = 1.01) did not significantly correlate with their far transfer performance in the concept-synonym condition, $r(38) = .03$ and .03, $p = .83$ and .88, respectively, and in the concept-error condition, $r(38) = .25$ and .13, $p = .12$ and .42, respectively. In view that both English proficiency and need for cognition consistently did not predict learners' test performance across Experiments 1 and 2, both variables were not examined any further in the subsequent experiment.

Learners reported relatively low prior knowledge of both study texts on overall, with no significant difference between the concept-synonym (*M* = 3.08, *SD* = 1.97) and concept-error (*M* = 2.68, *SD* = 1.58) conditions, $t(39) = 1.21$, $p = .24$, 95% $CI_{mean\ difference}$ [−0.27, 1.07]. Learners' ratings of the study texts' interestingness also did not differ across the concept-synonym (*M* = 4.23, *SD* = 1.56) and concept-error (*M* = 3.85, *SD* = 1.35) conditions, $t(39) = 1.29$, $p = .20$, 95% $CI_{mean\ difference}$ [−0.21, 0.96], although they rated the study text as more understandable in the concept-synonym (*M* = 5.25, *SD* = 1.37) than concept-error (*M* = 4.78, *SD* = 1.03) condition, $t(39) = 2.18$, $p = .035$, 95% $CI_{mean\ difference}$ [0.04, 0.92].

Far Transfer Test Performance

The key finding was that the concept-error method (*M* = 2.85, *SD* = 0.92) outperformed the concept-synonym method (*M* = 2.23, *SD* = 1.10) in enhancing learners' far transfer test performance, $t(39) = -3.10$, $p = .004$, $d = 0.49$, 95% $CI_{mean\ difference}$ [−1.03, −0.22]. Thus, deliberate erring produced superior far transfer than generating correct conceptual synonyms.

Metacognitive Judgments of Learning

However, learners failed to accurately predict the relative effectiveness of both learning methods. In contrast to their actual performance, learners erroneously predicted in their JOLs that the concept-synonym (*M* = 49.75, *SD* = 21.30) and concept-error (*M* = 47.25, *SD* = 19.08) methods would help their learning to similar extents, $t(39) = 0.74$, $p = .47$, 95% $CI_{mean\ difference}$ [−4.36, 9.36]. Curiously, after the test, learners even believed that the concept-synonym method (*M* = 4.75, *SD* = 1.45) had been more effective than the concept-error method (*M* = 3.73, *SD* = 1.50), $t(39) = 4.15$, $p < .001$, $d = 0.65$, 95% $CI_{mean\ difference}$ [0.53, 1.52].

Table 2 shows the predicted versus actual effectiveness of both learning methods for participants' far transfer performance. Overall, 22 out of 40 (55%) learners performed better in the concept-error than concept-synonym condition. Yet, 28 out

Table 2 Frequency count (and percentage) of participants showing different patterns of metacognitive ratings and actual transfer test performance (Experiment 2)

Metacognitive ratings vs. actual performance	Performance outcome		
	Error > synonym	Error = synonym	Error < synonym
Metacognitive ratings			
Pre-test predictions (JOLs)	12 (30%)	10 (25%)	18 (45%)
Post-test judgments (effectiveness ratings)	6 (15%)	12 (30%)	22 (55%)
Actual transfer test performance	22 (55%)	12 (30%)	6 (15%)

$N = 40$.

of 40 (70%) learners predicted that the concept-synonym method would be just as effective as or even more effective than the concept-error method. Strikingly, this divergence from learners' actual test performance was amplified in their post-test ratings of each method's effectiveness, whereby 34 out of 40 (85%) learners rated the concept-synonym method as just as effective as or even more effective than the concept-error method.

Discussion

Experiment 2 marked the first demonstration of the derring effect with an inferential short-answer test, which is an assessment format that is often used in classrooms. Notably, Experiment 2 also provided first evidence for the advantage of deliberate erring over errorless learning in enhancing far transfer of learned concepts to knowledge domains that are different from those in the initial learning task. This crucial finding extends the mnemonic benefits of deliberate erring to the flexible application of knowledge in new, remote contexts, illuminating an effective strategy to attain the central but challenging educational goal of transfer. Nonetheless, echoing Experiment 1's results, learners lacked metacognitive awareness of deliberate erring's effectiveness. In fact, after the test, the majority of learners even believed that deliberate erring had been less effective than generating conceptual synonyms.

Experiment 3

Experiment 3 aimed to replicate the principal finding in Experiment 2 that deliberate erring promoted far transfer to different knowledge domains, while examining the locus of this effect. As revealed through the concept-synonym control in the previous experiments, the benefits of deliberate erring for recall and transfer are not fully attributable to generation or elaboration alone. Rather, these benefits are specific to deliberately generating *incorrect* elaborations; generating correct elaborations does not yield the same advantage.

One might then question whether the derring effect stems from simply exposing learners to incorrect responses, particularly when their deliberate errors are juxtaposed with the actual concepts during error correction. For instance, relative to studying only correct solutions, some studies have found that comparing a mixture of correct and incorrect solutions produces better transfer in domains such as experimental design concept learning (Corral and Carpenter, 2020), mathematical problem-solving (Große and Renkl, 2007; Loibl and Leuders, 2019), computer programming (Beege et al., 2021), and behavior modeling training (Baldwin, 1992; Taylor et al., 2005). Presumably, comparing correct versus incorrect responses may highlight the diagnostic features that define the to-be-learned concept, and in turn foster richer mental models and deeper learning that facilitate transfer. Conversely, studying only correct information may not trigger such beneficial processes (see Corral and Carpenter, 2020 for a discussion). This account bears close parallels to the discriminative-contrast hypothesis in inductive learning, which proposes that juxtaposing examples of different concepts highlights the critical differences between them, thereby promoting better abstraction of those concepts (Kang and Pashler, 2012; Kornell and Bjork, 2008). In the present study, the implication is that learners' superior performance in the concept-error condition could have arisen from mere exposure to errors and/or comparing them with the correct answers during error correction, even if learners had not made these errors themselves.

Does the derring effect require that learners deliberately commit and correct *their own* errors? To pursue this potential account, Experiment 3 pitted the concept-error condition against a more stringent *spot-and-fix* control condition, in which learners received both the study text and an incorrect version of the text that contained a conceptual error in each sentence. Specifically, each learner was presented with the same errors that a peer had deliberately generated in Experiment 2's concept-error condition.[1] This yoked procedure ensured that the quality of errors to which learners were exposed was controlled for across the spot-and-fix versus concept-error conditions. Using the spot-and-fix method, learners compared both texts to identify and correct their peers' errors in the incorrect version, and further generated a conceptually correct elaboration (i.e., synonym) of their own. Thus, Experiment 3 enabled a comparison of deliberate error commission and correction against errorless elaboration supplemented with spotting and correcting others' errors. If the concept-error condition still prevails over the spot-and-fix condition, then mere exposure to others' errors and/or juxtaposing incorrect versus correct responses is inadequate to explain the derring effect. Rather, deliberately committing and correcting one's own errors is vital.

[1] It was not possible to present learners with the deliberate errors generated by a peer in Experiment 3's concept-error condition while still preserving a within-subjects design similar to that in the previous experiments. On balance, the deliberate errors from Experiment 2 were used—given that the learning materials and procedure were identical and that participants were sampled from the same population, the concept-error condition was expected to be comparable across Experiments 2 and 3. To ascertain this, cross-experiment analyses were conducted (see the "Results" subsection of "Experiment 3").

Method

Participants

The participants were 40 undergraduates (34 were female) between the ages of 19 and 24 ($M = 20.80$, $SD = 1.36$) from the National University of Singapore who did not take part in Experiments 1 and 2. Based on the medium-sized effect observed in Experiment 2 ($d = 0.49$), a power analysis (G*Power; Faul et al., 2007) indicated that at least 35 participants were required for 80% power at $\alpha = .05$.

Design

The single within-subjects factor was learning method: *concept-error* (deliberate error generation and correction) versus *spot-and-fix* (spotting and correcting others' errors, and generating an alternative correct elaboration). As in Experiment 2, the dependent variable was participants' far transfer performance on inferential short-answer test questions in different knowledge domains.

Materials and Procedure

The materials and procedure were identical to those in Experiment 2, with two exceptions. First, the concept-synonym condition was replaced with an even more stringent *spot-and-fix* condition that involved not only generating conceptual synonyms, but also spotting and correcting others' errors. A yoked design was used, whereby the errors that were presented in the spot-and-fix condition were the same ones that had been generated in the concept-error condition in Experiment 2. Specifically, in the spot-and-fix condition, each participant was presented with the actual study text and an incorrect version of the text that contained a conceptual error generated by a peer in each sentence. Thus, each participant was exposed to a different set of errors that a yoked peer from Experiment 2 had deliberately made. Participants were asked to compare and contrast both text versions, and to spot, strike out, and correct each error in the incorrect version. Then, participants wrote each sentence from the actual study text such that it contained a conceptual synonym of their own, and underlined this synonym. For instance, given the training example "Bats are mammals that fly" (correct version) and "Bats are birds that fly" (incorrect version), participants were to spot and correct the error in the incorrect version (e.g., cancel "~~birds~~" and write "mammals"), then generate and underline their own conceptual synonym while writing the full sentence from the correct version (e.g., "Bats are <u>warm-blooded animals with fur</u> that fly."). Thus, whereas the concept-error condition involved generating incorrect responses and correcting them, the spot-and-fix condition involved generating correct responses and correcting others' errors. In both conditions, participants applied the respective learning method to write each of the 20 sentences from the study text once.

Second, as learners' English language proficiency and need for cognition scores consistently did not correlate with their test performance in Experiments 1 and 2, both the English proficiency test and Need for Cognition Scale were dropped from Experiment 3.

Results

Scoring

Two raters independently scored 10 of the 40 (25%) scripts in a similar way as in Experiment 2, ICC = .94, 95% CI [.87, .98], based on a two-way random-effects model. Discrepancies were reviewed and resolved through discussion to reach 100% agreement. Given the high interrater reliability, the remaining scripts were scored by one rater.

Preliminary Analyses

Learners reported low prior knowledge of both study texts on overall, with no significant difference across the spot-and-fix ($M = 2.85$, $SD = 1.76$) and concept-error ($M = 2.80$, $SD = 1.73$) conditions, $t(39) = 0.12$, $p = .91$, 95% $\text{CI}_{\text{mean difference}}$ [− 0.82, 0.92]. There was also no significant difference in learners' ratings of how interesting the text was across the spot-and-fix ($M = 4.20$, $SD = 1.32$) and concept-error ($M = 4.28$, $SD = 1.43$) conditions, $t(39) = -0.23$, $p = .82$, 95% $\text{CI}_{\text{mean difference}}$ [− 0.73, 0.58], as well as how understandable the text was across the spot-and-fix ($M = 4.80$, $SD = 1.32$) and concept-error ($M = 5.03$, $SD = 1.46$) conditions, $t(39) = -0.68$, $p = .50$, 95% $\text{CI}_{\text{mean difference}}$ [− 0.90, 0.45].

Far Transfer Test Performance

Replicating Experiment 2's finding, the concept-error method ($M = 2.80$, $SD = 1.04$) produced superior far transfer performance than the spot-and-fix method ($M = 2.03$, $SD = 1.05$), $t(39) = -3.86$, $p < .001$, $d = 0.61$, 95% $\text{CI}_{\text{mean difference}}$ [− 1.18, − 0.37]. Thus, deliberately committing and correcting errors was more potent in enhancing far transfer, as compared to spotting and correcting others' errors coupled with generating correct conceptual synonyms.

Cross-Experiment Comparisons of Learning Methods

As the spot-and-fix condition involved exposure to the same errors generated in Experiment 2's concept-error condition, both yoked groups were compared. Indeed, learners in Experiment 2's concept-error condition outperformed those in Experiment 3's spot-and-fix condition, $t(78) = 3.74$, $p < .001$, $d = 0.83$, 95% $\text{CI}_{\text{mean difference}}$ [0.39, 1.27]. For completeness, it was further ascertained that learners' far transfer test performance in the concept-error condition did not differ across both experiments, $t(78) = 0.23$, $p = .82$, 95% $\text{CI}_{\text{mean difference}}$ [− 0.39, 0.49]. The concept-synonym

Table 3 Frequency count (and percentage) of participants showing different patterns of metacognitive ratings and actual transfer test performance (Experiment 3)

Metacognitive ratings vs. actual performance	Performance outcome		
	Error > spot-and-fix	Error = spot-and-fix	Error < spot-and-fix
Metacognitive ratings			
Pre-test predictions (JOLs)	19 (48%)	3 (8%)	18 (45%)
Post-test judgments (effectiveness ratings)	18 (45%)	13 (33%)	9 (23%)
Actual transfer test performance	25 (63%)	9 (23%)	6 (15%)

$N = 40$.

and spot-and-fix conditions also did not significantly differ, $t(78) = 0.83$, $p = .41$, 95% $CI_{\text{mean difference}}$ [−0.28, 0.68].

Metacognitive Judgments of Learning

In contrast to their actual test performance, learners incorrectly predicted that both learning methods would be just as helpful for them, with no significant difference in their JOLs across the spot-and-fix ($M = 45.25$, $SD = 21.84$) and concept-error ($M = 45.75$, $SD = 22.18$) conditions, $t(39) = -0.12$, $p = .90$, 95% $CI_{\text{mean difference}}$ [−8.85, 7.85]. Despite experiencing the benefits of deliberate erring for their test performance, learners still rated the spot-and-fix ($M = 3.93$, $SD = 1.47$) and concept-error ($M = 4.40$, $SD = 1.60$) methods as just as effective immediately after the test, $t(39) = -1.52$, $p = .14$, 95% $CI_{\text{mean difference}}$ [−1.11, 0.16].

Table 3 shows the predicted versus actual effectiveness of both learning methods for participants' far transfer performance. Overall, 25 out of 40 (63%) learners performed better in the concept-error than spot-and-fix condition. Yet, 21 out of 40 (53%) learners predicted that the spot-and-fix method would be just as effective as or even more effective than the concept-error method. This metacognitive illusion persisted even after the far transfer test, whereby 22 out of 40 (55%) learners rated the spot-and-fix method as just as effective as or even more effective than the concept-error method.

Interestingly, though, 18 out of 40 (45%) learners predicted before the test that the concept-error method would be less effective than the spot-and-fix method, but only 9 out of 40 (23%) learners made this judgment after the test. Instead, more learners (13 out of 40; 33%) judged the concept-error and spot-and-fix methods to be equally effective after the test, as compared to before the test (3 out of 40; 8%). Whereas they were not asked to explain their judgments, it is possible that some learners did realize after the test that the concept-error method had been helpful, but failed to accurately calibrate that the benefit it conferred was greater than that by the spot-and-fix method.

Discussion

The critical finding in Experiment 3 was that exposure to errors during study—and even correcting them after comparing them with the target responses—is less advantageous for far transfer if these errors have not been intentionally made by oneself. Despite involving the generation of correct elaborations plus spotting and correcting others' errors, the spot-and-fix condition still fell short of deliberately committing and correcting one's own errors. Moreover, cross-experiment comparisons of the concept-synonym versus spot-and-fix conditions revealed that no transfer advantage was accrued from additionally spotting and correcting others' errors, beyond generating conceptual synonyms alone. Thus, observing others' errors did not harm learning relative to errorless generation only, but also did not help learning as much as deliberate erring did. Altogether, the derring effect cannot be viably explained by mere error exposure and/or juxtaposition of correct versus incorrect responses, but hinges crucially on personally committing and correcting one's own deliberate errors. These findings will be taken up further in the "General Discussion" section.

Yet, Experiment 3 showed that, once again, learners failed to accurately predict or recognize the learning benefits of deliberate erring. As in the previous experiments, such faulty metacognitive awareness lingered even after learners had personally experienced the effects of both learning methods for their test performance.

General Discussion

Improving learners' far transfer of knowledge to contexts that are well removed from those in initial learning is an important pedagogical goal that has been considered the "holy grail" of education (Haskell, 2001; McDaniel, 2007), but to achieve it is a challenging feat (Barnett and Ceci, 2002; Detterman, 1993). Focusing on enhancing far transfer along the knowledge domain dimension of Barnett and Ceci's (2002) taxonomy, the present set of three experiments demonstrated that deliberate error commission and correction is an effective learning technique for bringing this goal into fruition.

Experiment 1 reliably obtained the basic derring effect, whereby deliberately committing and correcting errors improved knowledge retention more than generating conceptually correct responses. Crucially, Experiment 2 unveiled that this knowledge did not remain "inert" (Renkl et al., 1996), but could be flexibly used in new contexts. Indeed, deliberate erring enhanced far transfer on short-answer inferential questions that required applying learned concepts to make inferences about related concepts in distant knowledge domains, relative to errorless generation and elaboration. Moreover, in Experiment 3, deliberate erring still produced better far transfer even over a more stringent control where learners not only generated correct responses, but also spotted and corrected the very same errors that their peers had deliberately made. Yet, across all three experiments, learners' metacognitive predictions and judgments were at odds with their actual test performance. Notwithstanding deliberate erring's actual prowess, learners tended to believe that it was just as effective as—or even inferior to—avoiding first-hand errors.

In all, these results yield several key theoretical and practical insights on the critical role of deliberate errors in learning. Whereas the derring effect had previously been observed in recall and near transfer within the same knowledge domain (Wong and Lim, 2022a, 2022b), the present research has extended it to far transfer of learned concepts to different knowledge domains. This result is remarkable, given that experimental demonstrations of far transfer are scarce (Barnett and Ceci, 2002), with some exceptions (e.g., Butler, 2010; Chen and Klahr, 1999; Gick and Holyoak, 1980). Moreover, with meaningful learning as a primary goal of formal education, it is imperative that the efficacy of a learning technique is not limited merely to rote retention or verbatim reproduction (Mayer, 2002).

Besides expanding the generality of the derring effect, the present study showcases the utility of guiding errors in learning—an approach that has been largely neglected in cognitive and educational research but that complements our repertoire of errorful learning techniques (Wong and Lim, 2019b). Notably, Experiment 3 revealed that guided deliberate errors improved transfer more than observed errors. This finding aligns with those in some studies suggesting that it is crucial for learners to generate their own errors for the benefits of errorful learning to be maximized. For instance, Metcalfe and Xu (2018) found that learners' memory for the correct responses was enhanced when they had inadvertently erred themselves and received corrective feedback, rather than witnessed a peer making errors and being corrected. Likewise, Kapur (2014) found that learners displayed better conceptual understanding and transfer when they had been induced to inadvertently generate incorrect solutions to a complex problem before receiving instruction, relative to evaluating their peers' incorrect solutions then receiving the same instruction. Extending these results, the present study showed that the advantage of deliberately generating errors holds even when the control condition further involves generating conceptually correct responses, in addition to spotting and correcting others' errors. Altogether, these findings advance a more nuanced understanding of how different errorful approaches fare against each other, beyond the binary presence versus absence of errors in learning.

Why Does Deliberate Erring Improve Far Transfer?

Here, learners had to recall and apply their learned knowledge in novel contexts after being told that the inferential test questions were related to the material that they had studied. Thus, the memory demands of the task tapped on the recall and execution components of transfer (Barnett and Ceci, 2002), whereas learners were not required to recognize that the previously encountered versus new contexts were related. Given that deliberate erring produced better recall than generating conceptual synonyms (Experiment 1), this mnemonic benefit could have contributed to learners' superior transfer in applying the learned concepts to different knowledge domains. It is also possible that deliberate erring enhanced learners' ability to execute their remembered knowledge in new contexts, although it should be noted that the present study was not designed to tease apart the recall and execution components of transfer. As Chi and VanLehn (2012) proposed, failure to transfer often reflects a lack of deep initial learning—when learners do learn the source information deeply, they are more likely to successfully transfer that information. Thus, explaining the advantage of deliberate erring for the recall and/or execution components of transfer can be reframed as a question of how this technique fosters deep initial learning.

As Experiments 1 and 2 demonstrated when comparing the concept-error versus concept-synonym conditions, the derring effect is unique to generating incorrect, but not correct, elaborations. This suggests that engaging in generation and elaboration per se cannot fully explain the deeper learning that deliberate erring affords (Wong and Lim, 2022b). Then, is this learning benefit driven by mere exposure to incorrect information and/or its juxtaposition with the actual concepts during error correction? This conjecture is unlikely given Experiment 3's findings. Even when learners spotted and corrected others' errors besides generating conceptual synonyms, their transfer performance was still worse than when they had deliberately committed and corrected their own errors. Moreover, because the spot-and-fix condition involved presenting learners with the same deliberate errors that their peers had generated in the concept-error condition, the differential test performance across these conditions cannot be due to differences in the quality of errors that learners were exposed to. Put together, it does not suffice to generate any novel response or to study others' incorrect responses. Rather, it is vital that learners personally commit—invest in—their own deliberate errors to reap the corresponding learning benefits (see also Kapur, 2014; Metcalfe and Xu, 2018).

It is conceivable that deliberate erring induces mental processes that are not inherently associated with the study material but that are beneficial for learning it, whereas errorless study may not evoke such processing or does so to a lesser degree (Wong and Lim, 2022a; see also McDaniel and Einstein, 1989). Although the present data do not speak directly to these processes, some mechanisms underlying other errorful approaches can potentially apply to deliberate erring. For instance, theories on inducing "naturalistic" errors during initial problem-solving before instruction (i.e., PS-I; Loibl et al., 2017) have proposed an interaction among three cognitive mechanisms that facilitate learning and transfer: (a) prior knowledge activation, (b) awareness of knowledge gaps, and (c) recognition of deep features. By extension, generating deliberate errors (or conceptual synonyms) may activate learners' prior knowledge, such that subsequent organization or integration of their mental models can occur. In particular, deliberate erring may raise learners' awareness of their knowledge gaps more effectively than errorless elaboration. When intentionally generating plausible conceptual errors, learners must search their prior knowledge to derive responses that would definitively challenge or subvert it. For example, when generating the statement "Bats are mammals that swim" and considering whether it is conceptually wrong, one must search one's knowledge about bats' swimming abilities. Just because one does not know or is unsure does not necessarily mean that bats do not swim (in fact, they do). In this way, deliberate erring may more acutely highlight gaps in one's understanding while culling unproductive responses for better learning (Kornell et al., 2009), relative to building on one's existing knowledge to generate correct elaborations in the concept-synonym condition. Moreover, because knowledge gaps are better identified when they are personally experienced (Glogger-Frey et al., 2015; Loibl et al., 2017; Needham and Begg, 1991), it follows that deliberately committing and correcting one's own errors was more effective than receiving those same errors without generating them oneself in the spot-and-fix condition.

When knowledge gaps have been detected, learners may also be more likely to actively process subsequent corrective information to remedy their mental models (Loibl and Rummel, 2014). Through this process, deliberate error correction could enhance encoding of the to-be-learned concepts, in line with previous findings that correcting one's deliberate errors yields superior learning than leaving them uncorrected (Wong and Lim, 2022b). Furthermore, comparing one's deliberate errors with the actual concepts during error correction may promote recognition and abstraction of the concepts' diagnostic features, relative to learning from correct responses only (see Corral and Carpenter, 2020 for a discussion). Thus, these cascading processes may lead to deeper learning that facilitates transfer (Loibl et al., 2017). It is worthwhile to directly test these potential mechanisms in future research.

Implications for Education

This work suggests that students would gain more from incorporating deliberate errors in conceptual learning, relative to generating correct elaborations or spotting and correcting their peers' errors within the same study duration. For instance, when studying scientific texts and concepts, intentionally formulating plausible incorrect responses (e.g., considering what a concept is *not*) and correcting them may yield deeper learning that boosts retention and far transfer. Because students often do not recognize the learning benefits of deliberate erring, though, they may profit from being introduced to and guided on effectively using this technique during their study routines. Whereas learners in the present study generated and wrote down a deliberate conceptual error for each sentence in the study text to ensure that they fully processed the text in its entirety, the technique of deliberate erring can plausibly be more efficiently implemented while still preserving its effectiveness. For instance, during class discussions, students could be guided to verbally propose deliberate errors then correct and explain them (see also McDaniel et al., 2009 for a portable read-recite-review strategy when learning from educational texts).

From a practical perspective, deliberate erring in low-stakes contexts allows for errors to be systematically harnessed in classrooms and self-regulated study to maximize learning opportunities, rather than waiting for errors to occur spontaneously—especially in high-stakes tests when they are costly. Moreover, guiding learners to deliberately err could potentially mitigate some negative side effects of other errorful approaches (Wong and Lim, 2019b). For instance, errors that arise incidentally or are induced during learning may unwittingly erode motivation and incur emotional costs such as frustration and shame, especially when learners attribute these errors to low ability on their part (Brodbeck et al., 1993; Pekrun, 2006; Weiner, 1985). In turn, such negative emotions may impair learning by diverting attention away from the task at hand (Eskreis-Winkler and Fishbach, 2019; Frese and Keith, 2015; Pekrun et al., 2002). Furthermore, negative affect toward failure has been associated with learners' reduced willingness to take academic risks in tackling more challenging tasks for their greater learning (Abercrombie et al., 2022; see also Clifford, 1991). Conversely, because deliberate errors are framed as part of the intentional learning design, they may offer a means to mute or offset potential ego threats that would otherwise interfere with learning (Wong and Lim, 2022a, 2022b).

Of course, creating a "safe distance" from errors does not itself necessitate better learning. Whereas spotting and correcting others' errors also spares learners from any negative emotions associated with first-hand erring, it produced poorer transfer than deliberate erring in the present study, although teachers and students may be assured that observing others' errors does not appear to harm learning relative to generating correct responses only. However, when errors are seen as less aversive or threatening, learners may be more receptive—or, at least, more tolerant—toward actively engaging with them and capitalizing on the learning opportunities that they offer (Ivancic and Hesketh, 1995/1996; Keith and Frese, 2005).

Limitations and Future Directions

The present study raises new questions and prospects for future research on the content and context of transfer that deliberate erring promotes. In terms of the content of transfer, the derring effect has been observed with conceptual knowledge but not yet with procedural knowledge, such as that in mathematical problem-solving (e.g., Rittle-Johnson and Alibali, 1999; Rittle-Johnson et al., 2001), statistical hypothesis testing (e.g., Wong et al., 2019), and learning to execute motor sequences (e.g., Palmer and Meyer, 2000; Shea and Morgan, 1979). Some evidence in the domain of mathematical problem-solving suggests that conceptual and procedural knowledge develop iteratively, whereby gains in conceptual understanding facilitate the generation and transfer of correct procedures, and gains in procedural knowledge in turn increase conceptual understanding (Rittle-Johnson and Alibali, 1999; Rittle-Johnson et al., 2001, 2015). As such, one may be cautiously optimistic that the learning benefits of deliberate erring are not limited to conceptual knowledge only. In terms of the context of transfer, much more work is needed to test the derring effect in transfer across time (e.g., on delayed tests) and space (e.g., within and outside the classroom), beyond the knowledge domain dimension.

In addition, this research focused on learners' ability to recall and apply learned concepts when they had been informed that their prior learning was relevant to the new, remote context at hand. Thus, an unresolved issue is whether deliberate erring enhances the rarer and more demanding outcome of spontaneous transfer (Gick and Holyoak, 1980, 1983)—the ability to spontaneously "see" or notice the deep structure between learned versus new problems, and to determine whether and how studied content can be fruitfully applied in a new context, without relying on explicit hints to retrieve or use one's relevant prior knowledge (for discussions, see Barnett and Ceci, 2002; Chi and VanLehn, 2012).

Whereas Experiments 1 and 2 provided a head-to-head comparison of the concept-error versus concept-synonym methods, it remains an open question whether the advantage of deliberate erring for far transfer would prevail over errorless methods that involve greater or more sophisticated elaboration. To date, such comparisons have been investigated in the context of recall performance. For instance, Wong and Lim (2022b, Experiment 3) showed that deliberate erring produced better recall than an errorless control that involved generating examples of to-be-learned concepts. This benefit occurred even when example generation involved significantly greater elaboration in inducing responses that were more than 1.6 times the length of those produced when generating deliberate conceptual errors or synonyms. Thus, in this "David versus Goliath" battle, deliberate erring and correction in its basic form was more potent for recall than errorless learning that had been asymmetrically boosted. It would be interesting to test whether this advantage sustains for the more complex learning outcome of transfer.

Relatedly, it is possible that any effects of learning from others' errors in the spot-and-fix condition may be more robust when, for example, learners are further prompted to explicitly self-explain those errors, as in some studies on incorrect worked examples in mathematical problem-solving (e.g., Booth et al., 2013; Durkin and Rittle-Johnson, 2012; Loibl and Leuders, 2019). Although self-explanation does not seem to be a prerequisite for benefitting from studying others' errors so long as learners are made aware of the target responses (e.g., Corral and Carpenter, 2020), it may encourage learners to organize and integrate new information with their prior knowledge for better learning (Chi et al., 1989; Fiorella and Mayer, 2016), particularly with scaffolding to support high-quality self-explanations (for reviews, see Atkinson et al., 2000; Bisra et al., 2018; Dunlosky et al., 2013; Rittle-Johnson et al., 2017).

Another avenue for future research is whether and how learning from deliberate errors is affected by one's domain-specific prior knowledge, which has high variability in its predictive power for knowledge gains (Simonsmeier et al., 2022). Here, the advantage of deliberate erring was established among students who reported relatively low prior knowledge of the learning material. In contrast, when observing errors in worked examples, learners with low prior knowledge have been found to benefit more from studying correct solutions only, whereas learners with high prior knowledge benefit from studying a mixture of correct and incorrect solutions (Große and Renkl, 2007). To reconcile these findings and inform predictions about when errorful learning is more or less likely to be helpful, future work could examine the nature of interactions between learner characteristics and the learning task.

In particular, introducing errors during instruction has been closely associated with "desirable difficulties" that encourage deep processing for better learning (Bjork, 1994). However, such difficulties may become "undesirable" when learners are not well-equipped to overcome them (McDaniel and Butler, 2011), such as when the task demands impose cognitive overload (Sweller et al., 1998, 2019). Consequently, the benefits of errorful learning may be stunted when learners are unable to productively self-explain or remedy errors due to a lack of knowledge (e.g., Beege et al., 2021; Große, 2018; Große and Renkl, 2007; Heemsoth and Heinze, 2014). But when more scaffolding is provided, then even learners with low prior knowledge can profit from errors (e.g., Durkin and Rittle-Johnson, 2012; Toh and Kapur, 2017). In the present study, guiding learners to generate plausible conceptual errors during the practice phase, as well as implementing deliberate erring during open-book study, could have enabled productive engagement with errors without imposing undesirable load. Examining how deliberate erring can be enriched with strategic instructional design that is tailored to learners' prior knowledge and the learning task may buoy or even amplify the derring effect.

Besides the design of deliberate erring, there is merit in considering how its implementation can be supported at both the individual and social levels. As all three experiments revealed, one individual-level barrier is that learners tend to lack metacognitive awareness of deliberate erring's benefits. To produce sustainable change in learners' actual study practices, pursuing the sources of their metacognitive illusions and how best to mend them is a challenge worth addressing more specifically in future research. For instance, mnemonic-based debiasing (e.g., soliciting delayed JOLs instead of immediately after study) coupled with theory-based debiasing (e.g., explaining why deliberate erring outperforms errorless learning) can be explored as a route to actively counter learners' misconceptions (Koriat and Bjork, 2006; see also Biwer et al., 2020; McCabe, 2011; Yang et al., 2017).

More broadly, future research should identify and engineer features of the social context that afford the active promotion of deliberate errors in learning (i.e., *psychological affordances*; Walton and Yeager, 2020). For instance, the implementation of deliberate erring may be supported when teachers build a positive "error climate" in the classroom that encourages errors as integral parts of the learning process, and that is psychologically safe for exploration and risk-taking (e.g., Edmondson and Lei, 2014; Steuer et al., 2013), such as when learners try the new strategy of generating deliberate errors during class discussions. Combined with metacognitive interventions, fostering constructive learning environments in which deliberate erring can be fruitfully applied would buttress the effective use of this strategy.

Conclusion

We began with the analogy of how vaccines confer immunity through deliberate but safe exposure to pathogens. Consistent with this principle, the present research has shown that deliberately committing and correcting errors in low-stakes contexts improves retention and, crucially, far transfer. As compared to generating correct elaborations or spotting and correcting others' errors, deliberate erring enhanced students' ability to apply learned concepts to distant knowledge domains. Actively guiding deliberate errors in learning, rather than avoiding or observing them, is a potent strategy to gain generalizable knowledge.

Materials and Data Availability

The materials for the experiments reported in this study are available in the online supplementary materials. The data that support the findings of this study are available from the author upon reasonable request.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10648-023-09739-z.

Acknowledgements The author thanks Kagen Y. L. Lim and Huiyi Liu for their assistance with data scoring.

Funding This research was supported by a National University of Singapore Educational Research grant (C-581–000-222–091) awarded to Sarah Shi Hui Wong.

Declarations

Ethics Approval This research was conducted with the appropriate ethics approval from the National University of Singapore's institutional review board.

Consent to Participate Informed consent was obtained from all participants included in the study.

Competing Interests The author declares no competing interests.

References

Abercrombie, S., Carbonneau, K. J., & Hushman, C. J. (2022). (Re)examining academic risk taking: Conceptual structure, antecedents, and relationship to productive failure. *Contemporary Educational Psychology*, *68*, Article 102029. https://doi.org/10.1016/j.cedpsych.2021.102029

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

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology, 111*(2), 189–209. https://doi.org/10.1037/edu0000282

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. https://doi.org/10.1037/a0021017

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Anolli, L., Antonietti, A., Crisafulli, L., & Cantoia, M. (2001). Accessing source information in analogical problem-solving. *The Quarterly Journal of Experimental Psychology, 54A*(1), 237–261. https://doi.org/10.1080/02724980042000093

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.

Baldwin, T. T. (1992). Effects of alternative modeling strategies on outcomes of interpersonal-skills training. *Journal of Applied Psychology, 77*(2), 147–154. https://doi.org/10.1037/0021-9010.77.2.147

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

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

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637. https://doi.org/10.1037//0033-2909.128.4.612

Beege, M., Schneider, S., Nebel, S., Zimm, J., Windisch, S., & Rey, G. D. (2021). Learning programming from erroneous worked-examples. Which type of error is beneficial for learning? *Learning and Instruction*, *75*, Article 101497. https://doi.org/10.1016/j.learninstruc.2021.101497

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201–210. https://doi.org/10.3758/BF03193441

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

Biwer, F., de Bruin, A. B. H., Schreurs, S., & oude Egbrink, M. G. A. (2020). Future steps in teaching desirably difficult learning strategies: Reflections from the Study Smart Program. *Journal of Applied Research in Memory and Cognition, 9*(4), 439–446. https://doi.org/10.1016/j.jarmac.2020.07.006

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.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. McKay.

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

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

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21–32.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45*(4), 805–818. https://doi.org/10.1037/0022-3514.45.4.805

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(6), 1147–1156. https://doi.org/10.1037/0278-7393.15.6.1147

Catrambone, R., & Holyoak, K. J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition, 18*(6), 593–603. https://doi.org/10.3758/BF03197102

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098–1120. https://doi.org/10.1111/1467-8624.00081

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

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

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

Clifford, M. M. (1991). Risk taking: Theoretical, empirical, and educational considerations. *Educational Psychologist, 26*(3–4), 263–297. https://doi.org/10.1080/00461520.1991.9653135

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

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., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

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

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its

resistance to correction. *Nature Reviews Psychology, 1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y
Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 23–43. https://doi.org/10.1146/annurev-orgpsych-031413-091305
Epstein, D. (2019). *Range: Why generalists triumph in a specialized world*. Riverhead Books.
Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure—the greatest failure of all. *Psychological Science, 30*(12), 1733–1744. https://doi.org/10.1177/0956797619881133
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
Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9
Fisher, D., Frey, N., & Hattie, J. (2016). *Visible learning for literacy, grades K-12: Implementing the practices that work best to accelerate student learning grades*. Corwin.
Frese, M., & Keith, N. (2015). Action errors, error management, and learning in organizations. *Annual Review of Psychology, 66*, 661–687. https://doi.org/10.1146/annurev-psych-010814-015205
Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2008). Negative knowledge: Understanding professional learning and expertise. *Vocations and Learning, 1*(2), 87–103. https://doi.org/10.1007/s12186-008-9006-1
Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25*(4), 524–575. https://doi.org/10.1006/cogp.1993.1013
Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306–355. https://doi.org/10.1016/0010-0285(80)90013-4
Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6
Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001
Große, C. S. (2018). "Copying allowed – but be careful, errors included!" – Effects of copying correct and incorrect solutions on learning outcomes. *Learning and Instruction, 58*, 173–181. https://doi.org/10.1016/j.learninstruc.2018.06.004
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
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
Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*(4), 332–340. https://doi.org/10.3758/BF03197035
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.
Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*(6), 649–667. https://doi.org/10.1016/S0022-5371(78)90393-6
Joung, W., Hesketh, B., & Neal, A. (2006). Using "war stories" to train for adaptive performance: Is it better to learn from error or success? *Applied Psychology: An International Review, 55*(2), 282–302. https://doi.org/10.1111/j.1464-0597.2006.00244.x

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*(1), 97–103. https://doi.org/10.1002/acp.1801

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. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M. (2014). 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. (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., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775. https://doi.org/10.1126/science.1199327

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

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*(1), 59–69. https://doi.org/10.1037/0021-9010.93.1.59

Kimball, D. R., & Holyoak, K. J. (2000). Transfer and expertise. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 109–122). Oxford University Press.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661–667. https://doi.org/10.1111/j.0956-7976.2004.00737.x

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1133–1145. https://doi.org/10.1037/0278-7393.32.5.1133

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction?" *Psychological Science, 19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

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

Levin, J. R. (1988). Elaboration-based learning strategies: Powerful theory = powerful application. *Contemporary Educational Psychology, 13*(3), 195–205. https://doi.org/10.1016/0361-476X(88)90020-3

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., & 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

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

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

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition, 39*(3), 462–476. https://doi.org/10.3758/s13421-010-0035-2
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., & 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., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20*(4), 516–522. https://doi.org/10.1111/j.1467-9280.2009.02325.x
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.
McGuire, W. J. (1961). Resistance to persuasion conferred by active and passive prior refutation of the same and alternative counterarguments. *The Journal of Abnormal and Social Psychology, 63*(2), 326–332. https://doi.org/10.1037/h0048344
McKeough, A., Lupart, J., & Marini, A. (Eds.). (1995). *Teaching for transfer: Fostering generalization in learning*. Lawrence Erlbaum. https://doi.org/10.4324/9781315044736
Mera, Y., Rodríguez, G., & Marin-Garcia, E. (2022). Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychonomic Bulletin & Review, 29*(3), 753–765. https://doi.org/10.3758/s13423-021-02022-8
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
Michael, A. L., Klee, T., Bransford, J. D., & Warren, S. F. (1993). The transition from theory to therapy: Test of two instructional methods. *Applied Cognitive Psychology, 7*(2), 139–153. https://doi.org/10.1002/acp.2350070206
Minsky, M. (1994). Negative expertise. *International Journal of Expert Systems, 7*(1), 13–19.
Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition, 19*(6), 543–557. https://doi.org/10.3758/BF03197150
Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning, 15*(1), 1–36. https://doi.org/10.1080/13546780802490186
OECD. (2019). *OECD learning compass 2030: A series of concept notes*. OECD. https://www.oecd.org/education/2030-project/contact/OECD_Learning_Compass_2030_Concept_Note_Series.pdf
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
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.
Perkins, D. N., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist, 47*(3), 248–258. https://doi.org/10.1080/00461520.2012.693354
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

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

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(2), 291–300. https://doi.org/10.1037/0278-7393.13.2.291

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*(1), 144–161. https://doi.org/10.1037/0022-0663.88.1.144

Reed, S. K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(1), 106–125. https://doi.org/10.1037/0278-7393.11.1.106

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

Renkl, A. (2005). The worked-out examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 229–245). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.016

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169. https://doi.org/10.1007/BF01405730

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., 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

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., 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

Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, *5*, Article 65. https://doi.org/10.1057/s41599-019-0279-9

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science, 26*(6), 515–520. https://doi.org/10.1177/0963721417712760

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*(2), 113–142. https://doi.org/10.1207/s15326985ep2402_1

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

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*(2), 179–187. https://doi.org/10.1037/0278-7393.5.2.179

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist, 57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Skinner, B. F. (1958). Teaching machines. *Science, 128*(3330), 969–977. https://doi.org/10.1126/science.128.3330.969

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

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

Sweller, J., van Merrienboer, 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

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

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*(4), 692–709. https://doi.org/10.1037/0021-9010.90.4.692

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions: (I). *Psychological Review, 8*(3), 247–261. https://doi.org/10.1037/h0074898

Toh, P. L. L., & Kapur, M. (2017). Is having more prerequisite knowledge better for learning from productive failure? *Instructional Science, 45*(3), 377–394. https://doi.org/10.1007/s11251-016-9402-0

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological Science, 29*(3), 219–226. https://doi.org/10.1177/0963721420904453

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

Wittrock, M. C. (1974). Learning as a generative process. *Educational Psychologist, 11*(2), 87–95. https://doi.org/10.1080/00461527409529129

Wong, S. S. H., & Lim, S. W. H. (2019a). From JOLs to JOLs+: Directing learners' attention in retrieval practice to boost integrative argumentation. *Journal of Experimental Psychology: Applied, 25*(4), 543–557. https://doi.org/10.1037/xap0000225

Wong, S. S. H., & Lim, S. W. H. (2019b). 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., 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

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