AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2022, Vol. 151, No. 1, 25–40
https://doi.org/10.1037/xge0001072

The Derring Effect: Deliberate Errors Enhance Learning

Sarah Shi Hui Wong and Stephen Wee Hun Lim
Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore

How can we strategically and systematically learn from our errors? Over their long history, errors have traditionally been prevented entirely or, at best, permitted to occur spontaneously. Across three experiments, we tested and found evidence for a counterintuitive phenomenon that we termed the *derring effect*—deliberately committing errors even when one already knows the correct answers produces superior learning than avoiding them, particularly when one's errors are corrected. Learners engaged in an educationally relevant task of learning scientific term-definition concepts via open-book study by deliberately generating conceptually incorrect definitions with or without correction, or copying and underlining them (Experiment 1). On a cued recall test, deliberate erring outperformed errorless copying, with error correction yielding an additional benefit. This advantage of deliberate erring persisted over actively generating alternative conceptually correct answers (Experiment 2), which in turn surpassed copying. Even when errorless generation was given a further boost to involve a higher degree of elaboration by prompting learners to generate a specific real-world example that illustrated or applied each concept, deliberately committing and correcting errors still produced better learning (Experiment 3). Altogether, the derring effect is neither fully attributable to a generation nor an elaboration benefit, but stems at least in part from enhanced target processing specific to having first deliberately produced incorrect responses. Notwithstanding deliberate erring's prowess, learners were largely oblivious to its benefits, misjudging the strategy as less effective. Both theoretical and educational implications of positioning errors as active, systematic, and intentional events in learning are discussed.

Keywords: concept learning, elaboration, error correction, errors, generation

"If we learn from our mistakes, shouldn't I try to make as many mistakes as possible?," a student asks his teacher in a cartoon by American cartoonist Randy Glasbergen. This cartoon often draws laughter precisely because the notion of deliberate erring seems to defy common sense and rational logic. Understandably, we will be hesitant to err on important examinations, play the wrong notes in public musical performances, or incur financial losses from making bad investments. Errors can be costly and should reasonably be avoided when the stakes are high. In low-stakes learning contexts, however, taking Glasbergen's student's question at its most literal may transform what appears to be the tragedy of errors into powerful learning opportunities.

Errors have traditionally endured a poor reputation, with advocates of errorless learning suggesting that errors should be avoided at all costs so that learning is not inhibited when these errors are ingrained and reproduced in the future (e.g., Ausubel, 1968; Skinner, 1958). Contrary to such views, however, some empirical work has demonstrated that errors can be more beneficial for learning than previously assumed (for reviews, see Metcalfe, 2017; Wong & Lim, 2019). For instance, attempting and invariably failing to guess the correct responses to fictional trivia questions (i.e., questions with no real answers to begin with) or the targets in weak-associate word pairs (e.g., *frog-pond*) has been found to produce superior learning than errorless studying of the question-answer or cue-target (Kornell et al., 2009). These surprising learning benefits of errors have been replicated and extended in several studies (e.g., Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012; Kang et al., 2011), adding to our understanding of when errors are more likely to be helpful. For example, to reap the benefits of erring, it is important that learners personally commit errors, rather than merely observe or correct others' errors (Metcalfe & Xu, 2018; Sadler & Good, 2006). It is also crucial that feedback is provided after errors have been made (Metcalfe & Kornell, 2007; Pashler et al., 2005), because incorrect responses are unlikely to be spontaneously corrected (Butler et al., 2008).

Although a growing body of literature suggests that errors can enhance learning, this line of evidence overwhelmingly pertains to

This article was published Online First July 5, 2021.
Sarah Shi Hui Wong https://orcid.org/0000-0003-4243-212X
Stephen Wee Hun Lim https://orcid.org/0000-0003-3636-7587
This research was supported by a National University of Singapore Educational Research Grant (C-581-000-222-091) awarded to Sarah Shi Hui Wong and is based on part of her doctoral dissertation. We presented the idea of the derring effect and preliminary data at the inaugural Association for Psychological Science (APS) 2020 Virtual Poster Showcase; the premiered work was conferred the 2020 APS Student Research Award. All data reported in this article are original and have not appeared elsewhere previously.
Correspondence concerning this article should be addressed to Sarah Shi Hui Wong, Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4, 9 Arts Link, Singapore 117570. Email: psywshs@nus.edu.sg

a particular class of errors—mistakes that inevitably occur because learners have absent or incorrect knowledge of the appropriate responses (Reason, 1990). Such errors include those that are allowed to occur "naturalistically" or organically (Metcalfe & Xu, 2018) and those that are induced through forcing learners to guess even when they have no idea of the correct answers (Kang et al., 2011; Potts & Shanks, 2014). In such contexts, errors ensue despite learners' intentions to avoid them and get things right. These errors are then tolerated or accepted as serendipitous events after they have occurred against one's will. Yet, although students and educators may be assured in the knowledge that unintentional errors are typically not harmful for learning (Kang et al., 2011), leaving error occurrences to chance has limits when seeking an effective learning strategy that can be systematically implemented in educational contexts, such as students' self-regulated learning.

How can errors be strategically positioned to optimize learning opportunities? One counterintuitive approach is to actively—*deliberately*—engineer and invest in errors from the outset. Human error frameworks (e.g., Reason, 1990) have identified violations involving deliberate deviations from socially appropriate practices, beyond unintentional slips (incorrect responses that are not what were correctly intended) and mistakes (incorrect responses resulting from incorrect knowledge). Indeed, in the Prevention–Permission–Promotion (3P) framework of approaches to errors in learning (Wong & Lim, 2019), errors can be actively promoted through guiding learners to commit them in a structured way (see also Lorenzet et al., 2005), beyond simply allowing or inducing errors to occur.

Besides being committed first-hand, guided errors are advantageous in that they can be more systematically introduced and corrected than spontaneous errors. The approach of guiding errors also circumvents a key obstacle that hinders learning from failure: "emotional backwash" when learners react poorly to unfavorable feedback (Pitt & Norton, 2017). Discovering that one has failed is ego-threatening, and the tendency to tune out after receiving negative feedback undermines the learning that could otherwise have occurred by attending to one's errors (Eskreis-Winkler & Fishbach, 2019). Guiding learners to deliberately commit and correct errors in low-stakes contexts minimizes such negative emotional responses by allowing them to attribute their errors to the learning approach, rather than low ability on their part. Curiously, however, the consequences of guided deliberate errors in learning have not yet been illuminated—in the educational domain, learners have typically been set up to err inadvertently but do not necessarily set out to err intentionally.

The Derring Effect

Our central question relates to the extent that learning benefits from deliberately committing and correcting errors even when one knows the correct answers, as opposed to avoiding errors. We termed this the *derring effect*. Here, we broadly define errors as objectively incorrect responses, which may deviate from the correct targets in various ways. For example, misspelling "*cat*" as "*kat*" is an error that will most certainly be marked as incorrect on any elementary school test despite its letter-sound correspondence, as will the conceptual error of stating that cats are "*reptiles*" rather than "*mammals*."

On the surface, there are reasons to expect that deliberate erring may exert no effect on, or even impair, learning. In line with the traditional view of errorless learning (e.g., Ausubel, 1968; Skinner, 1958), deliberate erring may appear to be counterproductive practicing of incorrect responses that strengthens one's mental routes to these errors, increasing their likelihood of reoccurring in the future. Alternatively, generating errors when one already knows the correct answers may simply introduce redundant noise that does not harm, but also does not help, learning.

However, standing this intuition on its head, a deeper consideration unveils several potent theoretical reasons to expect that deliberate erring may, in fact, enhance learning. First, generating errors may improve subsequent memory by making the encoding of corrective feedback more effective than when the preceding error had not been committed (Hays et al., 2013; Kornell et al., 2009; Potts & Shanks, 2014). For instance, error commission itself may draw attention to the target during correction to foster a distinctive and memorable episodic trace (Metcalfe & Huelser, 2020). At the same time, intentionally exploring incorrect retrieval routes via deliberate erring may ironically weaken and cull those unproductive routes, thus increasing the relative retrieval strength of the correct association (Kornell et al., 2009; see also McClelland & Rumelhart, 1985). Alternatively, basing on the well-documented paradoxical rebound effect of thought suppression (Wegner et al., 1987): When learners already know the correct answer but must willfully suppress it during deliberate error commission, a desirable rebound may occur when the suppressed correct response continues to intrude and prosper in the mind.

Deliberate erring may also promote deeper elaborative encoding processes (Craik & Lockhart, 1972; Craik & Tulving, 1975) than error-avoidant learning. During error generation, the activation of various concepts that are related to the target may form a richer mental network that benefits subsequent retrieval (Kornell et al., 2009; Potts & Shanks, 2014; see Huelser & Metcalfe, 2012 for a discussion). In this way, deliberate errors may serve as useful cues or "stepping stones" to retrieve the correct responses (Cyr & Anderson, 2015).

Furthermore, the intentionality and purposeful design of deliberate erring may be helpful for learning. Consonant with this notion, the deliberate-practice view of expertise holds that prolonged engagement in effortful and structured tasks designed to improve performance, as opposed to mindless drill, is integral for expert skill acquisition (Ericsson & Charness, 1994; Ericsson et al., 1993). Taken together, these accounts provide a theoretical basis to expect that learning may benefit from deliberate erring.

To Correct or Not to Correct Deliberate Errors?

To date, most error-related studies have presented learners with corrective feedback after they have erred (e.g., Kang et al., 2011; Kornell et al., 2009; Metcalfe & Xu, 2018; Potts & Shanks, 2014). This procedure has been necessary in extant work because of the nature of the mistakes commonly investigated—without feedback, learners would not know what the correct answers are. In contrast, deliberate erring presents an unusual choice: To correct or not to correct one's deliberate errors? Resolving this question involves considering various potential mechanisms for the derring effect that, crucially, offer opposing testable predictions. We outline each possible outcome here in turn.

If the derring effect derives solely from elaborative processing during deliberate error commission, then we would expect to see no additional benefit of correcting one's errors. Because learners already know the correct answers, error correction would be superfluous if all learning benefits have already been reaped from the preceding creation of an elaborate mental network during error generation. Accordingly, it follows that correcting one's deliberate errors versus leaving them uncorrected would produce comparable learning in this instance.

Conversely, if mental suppression of the correct answers during deliberate error commission is responsible for the derring effect, then *not* correcting one's deliberate errors should produce better memory for the targets than correcting one's errors. The rebound effect (i.e., increased accessibility of a thought) following suppression has been suggested to be fueled by a motivational need to use the suppressed construct, such that allowing people to express it satisfies this need and reduces the construct's accessibility (Liberman & Förster, 2000). From this perspective, being denied the opportunity to express a suppressed thought should make it particularly compelling (for a discussion, see Wenzlaff & Wegner, 2000). Indeed, interrupted and hence incomplete tasks or events are often remembered better than completed ones—the Zeigarnik effect (Zeigarnik, 1938; see also Baddeley, 1963). Thus, if suppressing the correct target during error commission underlies the derring effect, then expressing the target by writing it down during error correction would relax its suppression, thereby reducing or even eliminating the ironic rebound effect (e.g., Liberman & Förster, 2000). Consequently, the target would be less likely to be recalled on a later test than if learners had not corrected their deliberate errors during study.

However, if correction confers a recall advantage over leaving one's deliberate errors uncorrected, then suppression and elaborative processing alone will be rendered less plausible mechanisms for the derring effect. Rather, this pattern of results will be consistent with the idea that deliberate error commission potentiates subsequent encoding of the target when one's errors are corrected. Given that learners have full knowledge of the correct answers, this enhanced target encoding cannot be explained by curiosity (e.g., Potts et al., 2019) or surprise upon discovering that one has apparently responded wrongly (i.e., high-confidence errors; Butterfield & Metcalfe, 2001; see also Metcalfe, 2017 for a review).

The Present Study

In this research, we systematically tested the hypothesis that committing and correcting deliberate errors—violations that learners generate even while aware of the correct responses—produces superior learning over avoiding errors. Across three experiments, we investigated the derring effect when learners deliberately made conceptual (semantic) errors during an educationally relevant task of learning scientific term-definition concepts. Such concepts are commonly regarded as a core component of foundational knowledge that students must acquire to gain expertise in a particular domain, and concept learning features prominently in many academic courses across numerous disciplines (for discussions, see Pan & Rickard, 2017; Rawson & Dunlosky, 2016; Rawson et al., 2015). Similar to how students are often able to refer to their textbook or notes during their self-regulated learning of such concepts (Rawson & Dunlosky, 2016), participants in the present study learned the term-definition concepts under open-book conditions. Subsequently, participants completed a cued recall test in which they were prompted to recall the definition of each concept term. We also assessed learners' metacognitive knowledge about the effectiveness of deliberate erring versus errorless learning, given that poor metacognitive awareness may lead to the adoption of suboptimal learning techniques.

Experiment 1 tested the prediction that deliberate erring enhances learning more than avoiding errors (i.e., errorless learning), and examined the specific contributions of deliberate error commission versus correction. Experiment 2 replicated and extended these findings by ruling out a generation account of the derring effect, while providing support for the mechanism of enhanced target processing that was specific to having first deliberately produced an error rather than any other novel (correct) response. Finally, Experiment 3 further interrogated and foreclosed the possibility that the derring effect stemmed from an encoding elaboration advantage. To foreshadow, the derring effect reliably prevailed across all three experiments—deliberately committing and correcting errors produced better learning than copying with underlining (Experiment 1), generating an alternative conceptually correct answer (Experiment 2), and even when errorless generation had been further boosted with a higher degree of elaboration by prompting learners to generate their own specific example that illustrated or applied the concept in a real-world setting (Experiment 3). In addition, error correction consistently yielded a learning advantage over leaving one's errors uncorrected (Experiments 1 and 2), pointing toward the role of enhanced target processing during correction after having deliberately erred.

Experiment 1

In Experiment 1, we investigated the basic effect of guiding learners to make deliberate errors in enhancing their learning of scientific term-definition concepts, while addressing two gaps in current error research. First, active errorful learning approaches have often been pitted against passive errorless learning methods that only involve presenting learners with correct information, such that learners are prevented from responding incorrectly but also do not make any overt responses at all (e.g., Huelser & Metcalfe, 2012; Kornell et al., 2009). Hence, Experiment 1 used a copy condition as an active errorless learning control, in which participants copied the term-definition concepts and further underlined the main ideas contained in these concepts, rather than only passively reading them. Second, to evaluate the specific contributions of error commission versus correction, we included both an error-cancel condition in which learners only made deliberate errors, and an error-correction condition in which learners further corrected their deliberate errors. If one mechanism of the derring effect is the potentiation of subsequent corrective feedback, then correcting one's deliberate errors should produce better recall than leaving them uncorrected (see the Introduction).

Method

Participants

The participants were 50 undergraduate students (34 were female) between the ages of 18 and 25 (M = 20.20, SD = 1.49) from the National University of Singapore. The sample size was determined based on a power analysis (G*Power; Faul et al., 2007), which indicated that at least 41 participants would afford 80% power to detect medium within-subjects effects (d = 0.45) in the present experiment for two-tailed pairwise comparisons using an alpha (α) of .05. We estimated the effect size from preliminary studies with similar experimental conditions that we conducted on deliberate spelling (lexical) errors when studying word lists. Outcomes below are reported based on data from 45 participants; five participants who failed to conform to the experimental instructions were excluded from analyses.

In line with previous studies on concept learning (e.g., Rawson et al., 2015), our particular research interest was in the initial learning of unfamiliar concepts rather than on relearning already known concepts. As such, in all experiments reported here, only students who had not previously taken any biological psychology or neuroscience courses were recruited—this ensured that participants possessed minimal prior familiarity with the term-definition concepts in the experimental materials related to neuroscience. We further ascertained this by examining participants' self-reported prior knowledge of the concepts presented to them during the experiment.

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

Design

The single within-subjects factor of interest was learning method: *error-cancel* (deliberate erring) versus *error-correction* (deliberate erring with correction) versus *copy* (errorless learning; control condition). The dependent variable was the number of definitions that participants correctly recalled at test when cued with their corresponding concept terms.

Materials

We extracted 40 term-definition concepts from an introductory neuroscience textbook, *Behavioral Neuroscience* (Breedlove & Watson, 2018), versions of which have been widely used in both undergraduate curricula and educational psychology research (e.g., McDaniel et al., 2007; Pan & Rickard, 2017). To maximize content scope, we extracted one to three concepts from each of the textbook's 19 chapters. Each concept defined a biology or neuroscience term and averaged 15 words in length. The concepts were presented in a term-definition "A-is-B" sentence format, whereby each sentence began with the concept term, followed by the verb "is" or "are," and the definition of that term. Samples of the term-definition concepts are available in Table 1. The term in each concept was presented in bold font to closely resemble the format in which such concepts are typically presented in textbooks. Ten term-definition concepts were randomly designated as practice trial items (including one item for illustration purposes during training), whereas the remaining 30 concepts were sorted into three 10-item study lists. The concepts were closely matched for length across all three study lists, while ensuring that each textbook chapter was represented not more than once within each list for even content coverage. In each list, the order in which the concepts were presented was randomized.

Procedure

Upon their arrival at the laboratory, participants were informed that they would be learning scientific concepts and their definitions, and were asked to study the material to the best of their ability for a

Table 1
Sample Term-Definition Concepts and Responses Across Learning Conditions in Experiments 1–3

	Sample responses		
Term-definition concept	Error-correction method (Experiments 1–3)	Concept-synonym method (Experiments 2 and 3)	Concept-example method (Experiment 3)
Adaptation is a trait that increases the probability that an individual will leave offspring in subsequent generations.	Adaptation is a trait that ~~decreases~~ (increases) the probability that an individual will leave offspring in subsequent generations.	Adaptation is a trait that increases the probability that an individual will pass down their genes (leave offspring in subsequent generations).	Adaptation is a trait that increases the probability that an individual will leave offspring in subsequent generations. • The agility of the rabbit allows it to escape predators and survive to breed with other rabbits.
Astereognosis is the inability to recognize objects by touching and feeling them.	Astereognosis is the inability to recognize objects by ~~looking at~~ (touching and feeling) them.	Astereognosis is the inability to recognize objects through physical contact (by touching and feeling them).	Astereognosis is the inability to recognize objects by touching and feeling them. • When Tom is blindfolded, he cannot identify what he is holding in his hands.
James-Lange theory is the theory that our experiences of emotions are responses to the physiological changes that accompany them.	James-Lange theory is the theory that our experiences of emotions ~~produce~~ (are responses to) the physiological changes that accompany them.	James-Lange theory is the theory that our experiences of emotions are responses to the changes in our body (physiological changes) that accompany them.	James-Lange theory is the theory that our experiences of emotions are responses to the physiological changes that accompany them. • When I see a tiger, I tremble and my heart beats faster, hence I feel afraid.

Note. The term in each concept was presented in bold font to closely resemble the format in which such concepts are typically presented in textbooks. In the error-correction method, learners deliberately erred by writing down each concept such that it contained a conceptual error in its definition, before striking out the error they had made, and correcting it with the actual definition. The error-cancel method (Experiments 1 and 2) was identical to the error-correction method, except that learners did not correct their deliberate errors. The copy method (Experiment 1) was identical to the concept-example method, except that learners copied the underlined ideas again instead of generating an example.

later test. Participants then went through three experimental phases: practice, studying, and test.

Practice Phase. Participants first completed a series of practice trials to familiarize themselves with the three learning methods. In the *error-cancel* method, participants deliberately erred by writing down each concept such that it contained a conceptual error in its definition (i.e., an error in understanding or interpreting a concept's definition), before striking out the error they had made. As illustration, for the concept "*Cocktail party effect is the selective enhancement of attention to filter out distractions*," a sample response was: "*Cocktail party effect is the selective enhancement of attention ~~in making sense of~~ distractions*." To ensure that participants understood what was required of them, they were provided with sample responses that were considered conceptually wrong, as well as those that were not (e.g., "*Cocktail party effect is the focus of atention on a particular stimulus while filtering out distractions*" involves a spelling error but is otherwise conceptually correct). In line with previous error research on incorrect guessing (Kang et al., 2011) and competitive incorrect responses (Little & Bjork, 2015), participants were also encouraged to deliberately generate plausible conceptual errors (i.e., conceptually wrong but believable responses). For instance, "*Cocktail party effect is the selective enhancement of attention in making sense of distractions*" is a relatively more plausible conceptual error than "*Cocktail party effect is the selective enhancement of attention when drinking alcohol at parties*."

The *error-correction* condition was identical to the error-cancel condition, except that participants additionally corrected their deliberate error by writing down the actual definition (e.g., "*Cocktail party effect is the selective enhancement of attention ~~in making sense of~~ (to filter out) distractions*").

In the *copy* condition, participants wrote down the term-definition concept exactly as it was presented. They were also instructed to identify and underline a key idea in each concept, before writing it again (e.g., "*Cocktail party effect is the <u>selective enhancement of attention</u> to filter out distractions (selective enhancement of attention)*"). Underlining was intended to be behaviorally comparable to the act of striking out—drawing a line across—one's deliberate errors in the errorful conditions. Moreover, underlining is a popular learning technique that students frequently report adopting in their study routines (Dunlosky et al., 2013).

Studying Phase. After they had completed the practice trials, participants began the studying phase. In each learning condition, participants were presented with a handout of a study list containing 10 term-definition concepts. They were then given 1 min to read the list, before using the error-cancel, error-correction, or copy method to study it for 12 min. Participants studied the term-definition concepts under open-book conditions (i.e., referred to the study lists), similar to naturalistic self-regulated learning contexts in which students are often able to refer back to their textbook or notes. The order in which participants went through the three learning conditions was counterbalanced, as was the pairing of learning conditions and study lists. Participants were also told that if they finished before the 12-min period was up, they should spend the remaining time reviewing their response. Thus, total learning time was exactly matched across all conditions. All participants completed the task within the allocated time.

At the end of each learning condition, participants responded to a four-item questionnaire in which they made a *judgment of learning* (JOL) on a 11-point scale from 0% to 100% to predict how much of the material from the study list they would remember later, rated how *interesting* and *understandable* the study list was (1 = *not at all*; 7 = *extremely*), and indicated how well they knew the concepts in the list prior to studying it (i.e., *prior knowledge*; 1 = *not very well*; 7 = *very well*). Participants then took a short self-paced break before proceeding to the test phase.

Test Phase. Participants were tested on the definitions of the studied concepts via a cued recall test. On each test trial, a concept term was queried onscreen (e.g., "What is the Cocktail Party Effect?") and participants were asked to write down its full correct definition in as much detail as they could remember. The concepts were tested blocked by study lists corresponding to the order in which the lists had been experienced during the studying phase. Within each test block, the order of the queried concepts was systematically varied such that no two concepts appeared consecutively across the studying and test phases. After responding to all 30 concept terms in the cued recall test, participants rated how *effective* each of the three learning methods had been for them on a 7-point scale (1 = *not at all*; 7 = *extremely*).

Results

Scoring

Participants' test responses were scored by awarding one point for each definition that had been correctly recalled when cued with its corresponding concept term. Both verbatim restatements and paraphrases that preserved the meaning of the definitions were considered correct (e.g., Rawson & Dunlosky, 2016). For instance, for the concept of "cocktail party effect," an acceptable response that paraphrased the actual definition was: "*Cocktail party effect is the increased focus on a particular stimulus while tuning out distractions*." Conversely, a sample inadequate response was: "*Cocktail party effect is the enhancement of attention*." Each response was scored either as correct or incorrect, with a maximum possible score of 10 for each learning condition. To more closely examine any potential effects of the learning methods on participants' incorrect test responses (e.g., the extent that participants' deliberate errors during initial study were repeated at test), we also coded these errors according to four categories: (a) *commission errors* (i.e., inadequate or incorrect responses that were different from participants' initial deliberate errors), (b) *omission errors* (i.e., no response), (c) *confusion errors* (i.e., responses that gave the definition for another studied concept term instead), and (d) *intrusion errors* (i.e., in the errorful conditions only, responses that repeated the same deliberate errors that participants had committed during initial study).

Two raters independently scored 10 of the 45 scripts. Interrater reliability was high, intraclass correlation (ICC) = .97, 95% CI [.96, .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 Checks

Besides recruiting only learners who had not yet taken any biological psychology or neuroscience courses, we ascertained that learners possessed minimal familiarity with the term-definition concepts in the experimental materials. Indeed, learners reported low prior knowledge of the concepts on overall, with no significant differences across the copy (M = 1.89, SD = 1.09), error-cancel (M = 1.87, SD = 1.18), and error-correction (M = 1.89, SD = 0.91) conditions, $F(2, 88) = 0.02$, $p = .98$, $\eta_p^2 < .001$.

In addition, there were no significant differences across learning conditions in participants' ratings of how interesting the term-definition concepts were, $F(2, 88) = 1.07$, $p = .35$, $\eta_p^2 = .02$, as well as how understandable the concepts were, $F(2, 88) = 1.39$, $p = .25$, $\eta_p^2 = .03$. Means and standard deviations are presented in Table 2.

Cued Recall Performance

A one-way repeated-measures ANOVA revealed that participants' cued recall test performance significantly differed across learning conditions, $F(2, 88) = 42.43$, $p < .001$, $\eta_p^2 = .49$. As predicted, the error-cancel (M = 4.29, SD = 2.48) and error-correction (M = 5.13, SD = 2.50) methods produced superior learning over the copy method (M = 2.31, SD = 1.72), both ps $< .001$, d = 0.89 and 1.30, respectively. In addition, learning was significantly better in the error-correction than error-cancel condition, $p = .005$, d = 0.44. Thus, deliberate erring outperformed errorless copying, and correcting one's deliberate errors yielded an additional recall advantage (see Figure 1).

Error Type Analysis

Besides participants' correct test responses, we investigated the effect of learning method on their incorrect test responses by examining the proportion of commission, omission, confusion, and intrusion errors that occurred at test. Table 3 shows the means and standard deviations. Four participants attained perfect test scores (i.e., no incorrect responses) in at least one learning condition.

The majority of participants' incorrect test responses on overall were commission errors. Of particular interest, in both errorful conditions, only 1% of participants' incorrect test responses involved repeating the same deliberate errors that they had committed during initial study (i.e., intrusion errors), suggesting that deliberate erring conferred learning benefits while incurring little interference cost. The proportion of intrusion errors did not differ across both the error-cancel and error-correction conditions, $t(40) = 0.31$, $p = .76$, 95% CI [−0.02, 0.02].

The three learning conditions differed in the proportion of participants' incorrect test responses that were commission errors, $F(2, 80) = 7.74$, $p = .001$, $\eta_p^2 = .16$, and omission errors, $F(2, 80) = 8.44$, $p < .001$, $\eta_p^2 = .17$, but not confusion errors, $F(2, 80) = 0.44$, $p = .65$, $\eta_p^2 = .01$. Specifically, the proportion of commission errors was higher in the copy condition than the error-correction condition, $p < .001$, and the error-cancel condition, although the latter comparison fell just short of significance, $p = .051$. Both errorful conditions did not differ in their proportions of commission errors, $p = .069$. Conversely, participants made more omission errors in the error-correction condition than the copy and error-cancel conditions, $p < .001$ and $p = .016$, respectively. There was no difference in the proportion of omission errors across the copy and error-cancel conditions, $p = .24$.

Metacognitive Judgments

To analyze participants' metacognitive judgments of the three learning methods, we conducted one-way repeated-measures ANOVAs. Participants' JOLs and perceived effectiveness of each learning method (solicited before and after the test, respectively), differed across conditions, $F(2, 88) = 3.36$, $p = .039$, $\eta_p^2 = .07$, and $F(2, 88) = 11.71$, $p < .001$, $\eta_p^2 = .21$, respectively. Means and standard deviations are presented in Table 2. Interestingly, learners (mis)predicted that their performance would be significantly better in the copy condition than the error-cancel and error-correction conditions, $p = .038$ and .025, respectively; learners' predictions for both errorful conditions did not differ, $p = .83$. Even after experiencing the benefits of deliberate erring for their test performance, learners' metacognitive judgments remained largely inaccurate—although learners correctly perceived the error-correction method to be more effective than the error-cancel method ($p < .001$), they misjudged the copy method as more effective than the error-cancel method ($p = .001$) and equally effective as the error-correction method ($p = .92$).

Discussion

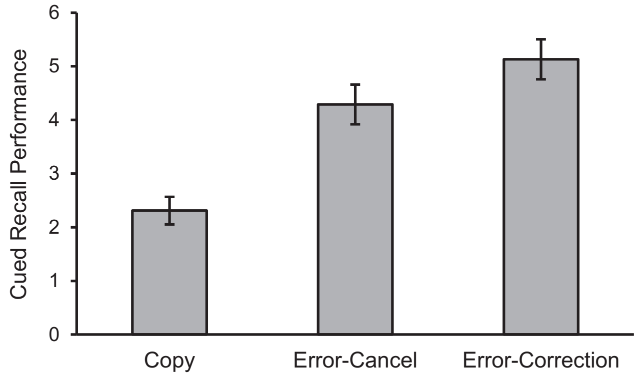
Experiment 1 provided initial evidence of the derring effect: Deliberately generating an incorrect definition of a scientific concept term produced better learning than errorless copying with underlining. Of note, this advantage emerged even when learners never once wrote the full correct definitions at all in the error-cancel condition, whereas they repeatedly wrote the target definitions more times in the copy condition. Furthermore, deliberate erring enhanced learning while introducing little interference at test, during which learners rarely repeated the errors that they had intentionally committed during initial study. Yet, learners were largely unaware of the benefits of deliberate erring, even after having just experienced better test performance for the concepts in the errorful conditions.

Table 2
Mean Questionnaire Scores and Metacognitive Judgments (Experiment 1)

Variable	Copy		Error-cancel		Error-correction	
	M	*SD*	*M*	*SD*	*M*	*SD*
Prior knowledge of concepts	1.89	1.09	1.87	1.18	1.89	0.91
Concept interestingness	4.04	1.40	3.96	1.52	3.82	1.40
Concept understandability	4.33	1.41	4.02	1.42	4.11	1.53
Judgment of learning (JOL)	42.00	18.04	37.56	20.58	37.11	18.04
Method effectiveness	4.24	1.21	3.22	1.41	4.22	1.52

Note. $N = 45$.

Figure 1
Mean Cued Recall Test Performance (Experiment 1)



Note. The maximum score was 10. Error bars indicate standard errors.

Another key finding in Experiment 1 was that learners profited further from correcting their errors than leaving them uncorrected, indicating that both deliberate error commission and correction contribute to the efficacy of this learning approach. Crucially, the benefit of error correction suggests that the derring effect is not simply an ironic rebound effect (Wegner et al., 1987) from having suppressed the correct answers when deliberately generating incorrect ones. If so, then *not* correcting one's errors would have produced better recall by encouraging mental preoccupations with the suppressed correct answers. Rather, a more promising notion that aligns with Experiment 1's data is that deliberate erring enhances subsequent encoding of the correct response during error correction. To further interrogate this account, we conducted Experiment 2.

Experiment 2

Experiment 2 aimed to replicate Experiment 1's findings and extend them in two ways. First, we sought to rule out generation as an alternative explanation of the derring effect. In both errorful conditions in Experiment 1, learners generated incorrect definitions of the concept terms, whereas the errorless copy condition did not induce such generation processes. However, information is often remembered better when it has been actively produced than passively read (Jacoby, 1978; Slamecka & Graf, 1978). To foreclose this competing account, we pitted the error-cancel and error-correction conditions against an even more rigorous errorless *concept-synonym* control condition that involved actively generating novel correct responses, rather than only copying and underlining the target definitions. Specifically, in the concept-synonym condition, learners elaborated on each concept by generating an alternative conceptually correct definition before writing the actual one. In prior research on elaborative learning strategies, the active generation of semantically related elaborations has been found to produce robust mnemonic benefits for the stimulus information that prompted those elaborations (Pressley et al., 1987). By this line of reasoning, actively generating alternative correct definitions should enhance memory for their corresponding concept terms more than simply copying the actual definitions without engaging in generation. Thus, we expected the concept-synonym condition to promote better learning than Experiment 1's copy condition, and tested this prediction via cross-experiment analyses.

Second, because the error-correction benefit observed in Experiment 1 suggested that a key mechanism underlying the derring effect is that deliberate errors potentiate subsequent encoding of their correction, Experiment 2 examined this idea more directly. We investigated whether this enhanced target processing was specific to having first deliberately produced an error (i.e., an *incorrect* elaboration) or whether it was a generalized benefit that followed the production of any other novel response, including correct elaborations in the concept-synonym condition.

Method

Participants

Forty-five undergraduate students (22 were female) between the ages of 18 and 26 (M = 20.31, SD = 1.47) from the National University of Singapore participated in this study. Outcomes reported below are based on data from 42 participants; three participants who failed to follow the experimental instructions were excluded from analyses. A power analysis (G*Power; Faul et al., 2007) indicated that this sample size afforded sufficient sensitivity to detect medium within-subjects effects ($d \geq 0.44$) for two-tailed pairwise comparisons at 80% power and $\alpha = .05$.

Design

The single within-subjects factor of interest was learning method: *error-cancel* (deliberate erring) versus *error-correction* (deliberate erring with correction) versus *concept-synonym* (generating an alternative correct response then writing the actual one).

Table 3
Proportion of Commission, Omission, Confusion, and Intrusion Errors in Learners' Incorrect Test Responses Across Learning Conditions (Experiment 1)

	Copy		Error-cancel		Error-correction	
Error category	*M*	*SD*	*M*	*SD*	*M*	*SD*
Commission	.66	.28	.58	.32	.50	.35
Omission	.28	.27	.32	.30	.43	.37
Confusion	.06	.09	.08	.19	.06	.12
Intrusion	—	—	.01	.04	.01	.04

Note. Intrusion errors—incorrect test responses that repeated learners' errors generated during initial study—were not applicable to the errorless copy condition.

As in Experiment 1, the dependent variable was participants' performance on a cued recall definition test.

Materials and Procedure

Experiment 2 used the same materials and procedure as Experiment 1, except that the copy control condition was replaced with the *concept-synonym* condition. The concept-synonym method was identical to the error-correction method, but instead of generating a conceptually incorrect definition, participants generated an alternative (correct) word or phrase with the same conceptual meaning as the actual definition (i.e., a conceptual synonym). That is, participants wrote down each term-definition concept such that it contained a conceptual synonym in its definition, before underlining this synonym, and writing down the actual definition exactly as it was presented. For instance, a sample response for the concept of "cocktail party effect" was: "*Cocktail party effect is the increased focus on a particular object (selective enhancement of attention) to filter out distractions*." Table 1 illustrates more sample responses for the concept-synonym method.

Results

Scoring

As in Experiment 1, participants' responses were scored as correct if they contained either verbatim restatements or paraphrases that preserved the meaning of the definitions. We note that this scoring procedure favored the concept-synonym condition, because the conceptual synonyms that participants generated during initial study would be considered correct if these paraphrases were later recalled at test, whereas participants' deliberate errors would be considered incorrect if these were recalled at test. Similar to Experiment 1, we also coded and classified participants' incorrect test responses into four error categories: commission, omission, confusion, versus intrusion errors. Two raters independently scored 10 of the 42 scripts, ICC = .99, 95% CI [.98, .99], based on a two-way random-effects model. Given the high interrater reliability, the remaining scripts were scored by one rater.

Preliminary Checks

We ascertained that learners reported low prior knowledge of the term-definition concepts on overall, with no significant differences across the concept-synonym (M = 1.81, SD = 1.04), error-cancel (M = 1.83, SD = 1.03), and error-correction (M = 1.83, SD = 1.10) conditions, $F(2, 82) = 0.02$, $p = .98$, $\eta_p^2 < .001$. In addition, learners perceived the concepts across all three learning conditions to be equally interesting, $F(2, 82) = 0.79$, $p = .46$, $\eta_p^2 = .02$, and understandable, $F(2, 82) = 0.85$, $p = .43$, $\eta_p^2 = .02$. Means and standard deviations are presented in Table 4.

Cued Recall Performance

Participants' cued recall test performance differed significantly across learning conditions, $F(2, 82) = 18.28$, $p < .001$, $\eta_p^2 = .31$. As predicted, learners displayed superior memory for term-definition concepts that had been studied via the error-cancel (M = 4.10, SD = 2.31) and error-correction (M = 4.74, SD = 2.18) methods than the concept-synonym method (M = 3.19, SD = 2.10), $p = .003$ and $p < .001$, d = 0.49 and 0.97, respectively. Thus, the derring effect persisted even over active generation of alternative conceptually correct responses. Replicating our findings in Experiment 1, correcting one's deliberate errors was also more potent than error commission alone, $p = .009$, $d = 0.42$. Figure 2 shows participants' test performance across learning conditions.

Cross-Experiment Comparison of Errorless Learning Methods

The concept-synonym condition in Experiment 2 was specifically designed to be an even more rigorous errorless learning method that involved actively generating novel conceptually correct responses, relative to the copy condition in Experiment 1. Indeed, comparing both errorless control conditions across Experiments 1 and 2,[1] an independent-samples t test revealed that the concept-synonym method outperformed the copy method, $t(85) = -2.15$, $p = .035$, 95% CI [−1.69, −0.07]. Hence, generating alternative conceptually correct elaborations boosted learning more than copying, although both errorless methods still paled in comparison to deliberate erring with or without correction. Across both experiments, participants' test performance was similar in the error-cancel condition, $t(85) = 0.38$, $p = .71$, 95% CI [−0.83, 1.22], as well as the error-correction condition, $t(85) = 0.78$, $p = .44$, 95% CI [−0.61, 1.40].

Error Type Analysis

To examine any potential effects of learning method on participants' incorrect test responses, we analyzed the proportion of commission, omission, confusion, and intrusion errors that occurred at test. Table 5 shows the means and standard deviations. Two participants attained perfect test scores (i.e., no incorrect responses) in at least one learning condition.

As in Experiment 1, the majority of participants' incorrect test responses were commission errors. Learners' deliberate errors during initial study rarely intruded at test—intrusion errors accounted for only 1–2% of participants' incorrect test responses and occurred at similar rates across the error-cancel and error-correction conditions, $t(39) = 0.54$, $p = .59$, 95% CI [−0.02, 0.03]. In addition, there were no significant differences in the proportion of all other error types across the three learning conditions: commission errors, $F(2, 78) = 0.15$, $p = .86$, $\eta_p^2 = .004$; omission errors, $F(2, 78) = 0.15$, $p = .87$, $\eta_p^2 = .004$; confusion errors, $F(2, 78) = 1.07$, $p = .35$, $\eta_p^2 = .03$.

Metacognitive Judgments

Echoing Experiment 1's findings, participants were largely unaware of the advantages of deliberate erring. In contrast to their actual test performance, participants' JOLs revealed that they erroneously predicted no difference in their learning across all three conditions, $F(2, 82) = 0.37$, $p = .69$, $\eta_p^2 = .01$. Even after personally experiencing the benefits of deliberate erring on the cued recall test, participants' metacognitive judgments of each learning method's effectiveness remained inaccurate, with differences across

[1] These analyses were made possible by the fact that the study and test materials were identical across Experiments 1 and 2, which was planned in advance to facilitate cross-experiment comparisons. In addition, participants were sampled from the same population without a break in data collection between experiments.

Table 4
Mean Questionnaire Scores and Metacognitive Judgments (Experiment 2)

Variable	Concept-synonym		Error-cancel		Error-correction	
	M	*SD*	*M*	*SD*	*M*	*SD*
Prior knowledge of concepts	1.81	1.04	1.83	1.03	1.83	1.10
Concept interestingness	3.62	1.56	3.86	1.54	3.74	1.52
Concept understandability	4.24	1.30	4.31	1.18	4.10	1.23
Judgment of learning (JOL)	36.90	23.11	34.76	21.78	35.00	20.75
Method effectiveness	3.95	1.55	3.36	1.43	4.02	1.49

Note. $N = 42$.

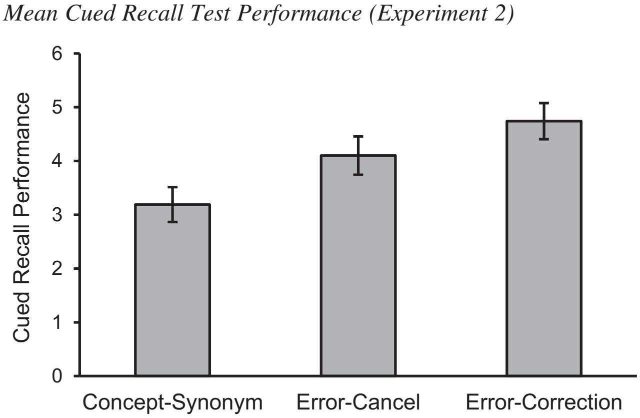
conditions that bordered on significance, $F(2, 82) = 3.07$, $p = .052$, $\eta_p^2 = .07$. Specifically, learners misjudged the concept-synonym method to be just as effective as the error-cancel and error-correction methods, $p = .065$ and .81, respectively, although the error-correction method was correctly rated as more effective than the error-cancel method, $p = .017$. Table 4 shows the means and standard deviations.

Discussion

Experiment 2 extended the derring effect observed in Experiment 1 by showing that it prevailed even over an errorless control that had been bolstered with active generation. Although generating an alternative conceptually correct definition was more effective for concept learning than copying, both of these errorless methods remained less potent than deliberately generating a conceptually incorrect definition. Notwithstanding the benefits of deliberate error commission alone, the error-correction condition reliably yielded greater learning gains than the error-cancel condition across both Experiments 1 and 2, suggesting that deliberate errors must be corrected to reap their full benefits.

Considered in tandem, these findings shed light on the theoretical underpinnings of the derring effect by demonstrating that it cannot be fully attributed to the mere act of engaging in generation per se. That learners benefited more from generating conceptual errors than conceptual synonyms suggests that their superior performance after deliberate erring is not simply a general consequence of producing any other novel response. Rather, our data are consistent with the notion that deliberate erring potentiates encoding of subsequent correction more than when the preceding error had not been made. That is, learners' enhanced target processing is specific to having first deliberately generated an *incorrect* elaboration; generating a correct elaboration does not offer the same benefit.

Figure 2
Mean Cued Recall Test Performance (Experiment 2)



Note. The maximum score was 10. Error bars indicate standard errors.

An alternative view, however, is that the derring effect may have arisen in part from potential qualitative differences in the extent of elaborative processing across the errorful versus errorless conditions. Indeed, information is often remembered better when it has been elaborated on through the addition of meaning (Craik & Lockhart, 1972; Craik & Tulving, 1975; Levin, 1988). Although learners similarly generated elaborations across all conditions in Experiment 2, it may be that generating plausible deliberate errors (i.e., what the target definitions were *not*) involved a greater degree of encoding elaboration than generating conceptual synonyms that paraphrased the definitions (i.e., what *else* the target definitions meant). If so, then such differences in elaborative encoding may have contributed to the superior performance in the errorful conditions than the concept-synonym condition by facilitating learners' construction of coherent and well-integrated mental representations of the material (e.g., Fiorella & Mayer, 2016; Kintsch, 2004). To more conclusively lay this possibility to rest, we conducted Experiment 3 in which we directly manipulated the level of elaboration that learners engaged in.

Experiment 3

The key question that we pursued in Experiment 3 pertained to the extent that the derring effect stemmed from an elaboration advantage. Specifically, we interrogated whether the benefits of deliberate erring would persist even over errorless learning that had been further supplemented with a higher degree of elaboration.

Elaboration techniques draw on generative learning principles (Fiorella & Mayer, 2016; Wittrock, 1974), which suggest that effective learning involves actively constructing meaning from the to-be-learned information by organizing and integrating it with one's own knowledge and experiences. For instance, one extensively researched elaboration technique is *elaborative interrogation* (e.g., Pressley et al., 1987), in which learners answer "why" questions by generating explanations for explicitly stated facts (e.g., "Why is this fact true?"). Elaborative interrogation has been found to promote learning, presumably because it prompts learners to form associations between the material and their background knowledge (for a review, see Dunlosky et al., 2013). However, its advantages appear to be limited for learners with low levels of

Table 5

Proportion of Commission, Omission, Confusion, and Intrusion Errors in Learners' Incorrect Test Responses Across Learning Conditions (Experiment 2)

Error category	Concept-synonym *M*	Concept-synonym *SD*	Error-cancel *M*	Error-cancel *SD*	Error-correction *M*	Error-correction *SD*
Commission	.56	.30	.53	.32	.53	.30
Omission	.39	.30	.41	.34	.39	.32
Confusion	.06	.11	.04	.09	.06	.11
Intrusion	—	—	.02	.06	.01	.04

Note. Intrusion errors—incorrect test responses that repeated learners' errors generated during initial study—were not applicable to the errorless concept-synonym condition.

prior domain knowledge (Dunlosky et al., 2013; Woloshyn et al., 1992). In contrast, learners profit from elaborative studying via *example generation* (i.e., generating personal examples of target concepts), even—or especially—when they possess low prior knowledge or low-quality mental representations of the to-be-learned material (Rawson & Dunlosky, 2016; Roelle & Nückles, 2019). Although example generation has received relatively less attention than elaborative interrogation in extant educational psychology research, a direct comparison of both techniques' effects on students' text comprehension has offered some evidence that they may be equally effective (Dornisch et al., 2011).

Examples are commonly used as a pedagogical tool in textbooks and classrooms to illustrate concepts (Rawson et al., 2015). Of interest, having learners generate their own examples of concepts produces better cued recall of the definitions than restudy, with example quality enhanced under open-book study conditions when the concepts' definitions are provided during example generation (Rawson & Dunlosky, 2016). Notably, example generation appears to be particularly effective when studying material with low cohesion and elaboration (e.g., text that contains few explicated interrelations between ideas and no concrete examples that illustrate the learning content). In such contexts, learners with low prior knowledge have been found to benefit from elaborating on and organizing the material by generating illustrative examples and identifying main ideas during open-book study, whereas engaging in the well-established technique of retrieval practice[2] did not enhance their comprehension and transfer performance (Roelle & Nückles, 2019; see Karpicke, 2017 for a review of retrieval-based learning). Taken together, these characteristics of example generation made it a particularly suitable errorless elaboration control in the present study, in which learners had minimal prior knowledge of the term-definition concepts that were not directly related to one another and did not contain illustrative examples, thus constituting educational material that was low in cohesion and elaboration.

Hence, in Experiment 3, we introduced the *concept-example* condition as an effective errorless elaborative learning method, in which participants actively generated real-world examples that illustrated or applied the to-be-learned concepts. We expected that this method would involve a greater degree of elaboration than generating conceptual synonyms, and assessed the text length of participants' study responses as a proxy for the elaboration that they engaged in (e.g., Daley & Rawson, 2019). We then pitted the errorless concept-synonym and concept-example methods against the error-correction method, which had consistently produced better concept learning performance than leaving one's errors uncorrected in the previous experiments. To the extent that the concept-example method demands more elaboration than the concept-synonym and error-correction methods, it should produce correspondingly greater learning gains if the benefit of deliberate erring arises predominantly from encoding elaboration. However, if deliberate erring still outperforms errorless example generation, then this would suggest that an elaboration account alone is inadequate to explain the derring effect.

Method

Participants

The participants were 45 undergraduates (35 were female) between the ages of 18 and 22 (M = 19.38, SD = 0.98) from the National University of Singapore. Outcomes reported below are based on data from 40 participants; five participants who did not conform to the experimental instructions were excluded from analyses. A power analysis (G*Power; Faul et al., 2007) indicated that this sample size afforded sufficient sensitivity to detect medium within-subjects effects ($d \geq .45$) for two-tailed pairwise comparisons at 80% power and $\alpha = .05$.

Design

The single within-subjects factor of interest was learning method: *error-correction* (deliberate erring with correction) versus *concept-synonym* (generating an alternative correct response then writing the actual one) versus *concept-example* (writing and underlining the correct definition, then generating a real-world example of the concept). As in the previous experiments, participants' concept learning was assessed via a cued recall definition test.

Materials and Procedure

Experiment 3 employed identical materials and procedures as Experiment 2, except that the error-cancel condition was replaced

[2] Although a wealth of research has demonstrated that retrieval practice is a potent technique that fosters learning in many educational settings (e.g., Dunlosky et al., 2013; Karpicke, 2017), it was less suitable as a comparison method in the present study for two key reasons. First, as detailed above, retrieval practice has been found to be less beneficial than generative learning (e.g., example generation) for studying materials with low cohesion and elaboration (Roelle & Nückles, 2019), which are features of the term-definition concepts that served as the educational materials in our study. Second, whereas the present research specifically aimed to test the effects of deliberate erring against traditional errorless learning methods, retrieval practice inherently introduces spontaneous errors during the study process (e.g., when learners unknowingly recall incorrect information or fail to recall it during study).

with the *concept-example* condition. In this errorless control, learners wrote down each term-definition concept exactly as it was presented, then identified and underlined a key idea contained in the concept, and generated a one-sentence example that illustrated or applied the concept in a real-world setting (e.g., Roelle & Nückles, 2019). To encourage elaboration processes, learners were also instructed to generate examples that were as specific as possible. For instance, a sample response for the concept of "cocktail party effect" was: "*Cocktail party effect is the selective enhancement of attention to filter out distractions. At a noisy party, Kate was able to focus on what her partner was saying while ignoring other people's conversations.*" More sample responses for the concept-example method are available in Table 1.

Results

Scoring

The scoring procedure was identical to that in the previous experiments—test responses that contained either verbatim restatements or paraphrases that preserved the meaning of the definitions were scored as correct. We note that this scoring procedure favored the errorless concept-synonym and concept-example methods. We also coded the proportion of participants' incorrect test responses that were commission, omission, confusion, versus intrusion errors. Two raters independently scored 10 of the 40 scripts. Because interrater reliability was high, ICC = .99, 95% CI [.98, .99], based on a two-way random-effects model, the remaining scripts were scored by one rater.

Preliminary Checks

Similar to the previous experiments, learners reported low prior knowledge of the term-definition concepts on overall, with no significant differences across the concept-synonym (M = 1.95, SD = 1.09), concept-example (M = 2.00, SD = 1.06), and error-correction (M = 2.05, SD = 1.15) conditions, $F(2, 78) = 0.17$, $p = .85$, $\eta_p^2 = .004$. Across all three learning conditions, participants also perceived the concepts to be similarly interesting, $F(2, 78) = 2.31$, $p = .11$, $\eta_p^2 = .06$, and understandable, $F(2, 78) = 1.75$, $p = .18$, $\eta_p^2 = .04$. Means and standard deviations are presented in Table 6.

Because the concept-example method was specifically intended to involve a higher degree of elaboration, we ascertained this by analyzing the text length of participants' study responses as a proxy for elaboration (e.g., Daley & Rawson, 2019). Indeed, the extent of elaboration that participants engaged in significantly differed across learning conditions, $F(2, 78) = 267.38$, $p < .001$, $\eta_p^2 = .87$. As expected, learners wrote significantly longer elaborations when generating examples (M = 289.88, SD = 41.38) than conceptual synonyms (M = 172.93, SD = 13.58) or deliberate errors (M = 172.65, SD = 15.26), both $ps < .001$; the latter two conditions did not differ, $p = .93$. On average, the elaboration in the concept-example condition represented more than a 60% increase in response length over the concept-synonym and error-correction conditions.

Cued Recall Performance

Learners' cued recall test performance differed significantly as a function of learning method, $F(2, 78) = 14.84$, $p < .001$, $\eta_p^2 = .28$. Consistent with our predictions, learners performed better in the concept-example (M = 4.75, SD = 2.30) than concept-synonym (M = 3.87, SD = 2.58) condition, $p = .006$, $d = 0.46$. Crucially, however, the error-correction method (M = 5.43, SD = 2.46) produced superior learning than not only the concept-synonym method, $p < .001$, $d = 0.81$, but also the concept-example method, $p = .011$, $d = 0.43$. Figure 3 displays learners' test performance across learning conditions.

Cross-Experiment Comparison of Learning Methods

For completeness, we compared Experiment 3's concept-example condition against the error-cancel condition in Experiments 1 and 2, which had employed identical study and test materials. Interestingly, participants' test performance was similar in the concept-example condition and the error-cancel condition in both Experiment 1, $t(83) = -0.89$, $p = .38$, 95% CI [−1.50, 0.58], and Experiment 2, $t(80) = -1.29$, $p = .20$, 95% CI [−1.67, 0.36]. In other words, deliberately committing errors alone does not harm learning relative to generating correct examples, although both methods were edged out by deliberate erring with correction. For good measure, we also established that participants' test performance in the error-correction condition was comparable across all three experiments, $F(2, 124) = 0.86$, $p = .43$, $\eta_p^2 = .01$.

Error Type Analysis

Besides participants' correct test responses, we probed the types of errors that occurred at test. Table 7 shows the means and standard deviations of the proportion of commission, omission, confusion, and intrusion errors in participants' incorrect test responses as a function of learning method. Four participants attained perfect test scores (i.e., no incorrect responses) in at least one learning condition.

Echoing our findings in Experiments 1 and 2, the majority of incorrect test responses were commission errors, which occurred at differing rates across learning conditions, $F(2, 70) = 6.32$, $p =$

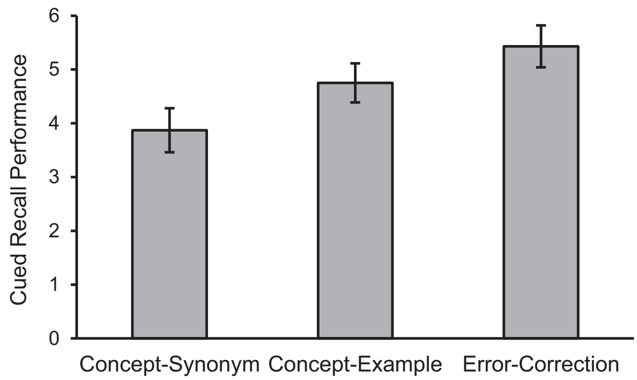
Table 6
Mean Questionnaire Scores and Metacognitive Judgments (Experiment 3)

Variable	Concept-synonym		Concept-example		Error-correction	
	M	*SD*	*M*	*SD*	*M*	*SD*
Prior knowledge of concepts	1.95	1.09	2.00	1.06	2.05	1.15
Concept interestingness	3.93	1.40	4.12	1.34	4.23	1.27
Concept understandability	4.00	1.24	3.85	1.41	4.17	1.30
Judgment of learning (JOL)	37.00	17.86	33.50	20.07	40.25	17.02
Method effectiveness	3.95	1.34	3.67	1.86	3.77	1.53

Note. N = 40.

Figure 3

Mean Cued Recall Test Performance (Experiment 3)



Note. The maximum score was 10. Error bars indicate standard errors.

.003, η_p^2 = .15. Specifically, the error-correction condition saw a lower proportion of commission errors than the concept-synonym and concept-example conditions, p = .001 and .013, respectively, with no significant difference between both errorless conditions, p = .60. The proportion of omission errors also differed across learning conditions, $F(2, 70) = 3.45$, p = .037, η_p^2 = .09, with a higher rate of omission errors in the error-correction than concept-synonym condition, p = .002, but no differences between the concept-example and concept-synonym or error-correction conditions, p = .27 and .21, respectively. In addition, the proportion of confusion errors did not differ across learning conditions, $F(2, 70) = 0.82$, p = .44, η_p^2 = .02. Importantly, intrusion errors accounted for only 4% of incorrect test responses in the error-correction condition, indicating that learners' deliberate errors were seldom repeated at test.

Metacognitive Judgments

Similar to the previous experiments, learners were largely unaware that deliberate erring had been helpful for their concept learning. Participants' JOLs revealed that they incorrectly predicted that all three learning methods would yield comparable test performance, $F(2, 78) = 2.83$, p = .065, η_p^2 = .07. Even after experiencing the benefits of deliberate erring for their actual test performance, participants still inaccurately perceived all three learning methods to be equally effective, $F(2, 78) = 0.33$, p = .72, η_p^2 = .008. Means and standard deviations are presented in Table 6.

Discussion

As expected, learners engaged in a higher degree of elaboration when generating examples that illustrated or applied the to-be-learned concepts in real-world settings. Indeed, we found that example generation induced elaborations that were more than 160% the response length of those produced when generating conceptual synonyms or deliberate errors. The increased elaboration in the concept-example condition was, to some extent, beneficial in improving learners' test performance relative to generating conceptual synonyms of the definitions within the same study duration. This result aligns with the mnemonic benefits of example generation over restudy documented in previous studies (Rawson & Dunlosky, 2016; Roelle & Nückles, 2019) and extends them by providing evidence for the relative efficacy of generating examples over alternative correct answers.

Yet, errorless example generation still led to poorer concept learning than deliberately committing and correcting errors. Affirming the robustness of the derring effect, the error-correction method outperformed not only the concept-synonym method in replication of Experiment 2's findings, but also the concept-example method that had been bolstered with a higher degree of elaboration. The key implication is that the benefit of deliberate erring cannot be parsimoniously explained by encoding elaboration alone. This and other theoretical accounts of the derring effect will be taken up in the General Discussion.

Although the efficacy of errorless learning can potentially be increased even further by supplementing example generation with additional generative learning techniques (e.g., drawing, mapping, enacting; Fiorella & Mayer, 2016) over a longer study duration, this would also imply that such an errorless learning approach requires marshalling disproportionately more aids to gain an asymmetric advantage over—or match—deliberate erring and correction in its basic form. From a practical perspective, deliberate erring can then be considered a more efficient learning strategy. Notably, deliberate errors enhanced learners' performance while rarely interfering at test—similar to the previous experiments, intrusion errors accounted for only up to 4% of incorrect test responses in the error-correction condition. However, learners had poor metacognitive awareness of the benefits of deliberate erring and misjudged all three learning methods to be equally effective.

General Discussion

For decades, errors have traditionally been avoided in learning (Ausubel, 1968; Skinner, 1958). Yet, errors are natural, ubiquitous,

Table 7

Proportion of Commission, Omission, Confusion, and Intrusion Errors in Learners' Incorrect Test Responses Across Learning Conditions (Experiment 3)

	Concept-synonym		Concept-example		Error-correction	
Error category	*M*	*SD*	*M*	*SD*	*M*	*SD*
Commission	.59	.28	.56	.30	.41	.31
Omission	.30	.28	.37	.33	.45	.33
Confusion	.11	.16	.07	.12	.10	.16
Intrusion	—	—	—	—	.04	.10

Note. Intrusion errors—incorrect test responses that repeated learners' errors generated during initial study—were not applicable to the errorless concept-synonym and concept-example conditions.

and inevitable events in life, and they may even be desirable in low-stakes educational contexts (e.g., Kornell et al., 2009; Metcalfe, 2017; Potts & Shanks, 2014). Whereas extant research has largely focused on errors that are allowed or induced, we tested the benefits of guiding learners to deliberately commit and correct errors even when they know the correct answers, as a systematic strategy to optimize learning opportunities.

Across three experiments, we established a converging chain of evidence for the derring effect in an educationally relevant task of learning scientific term-definition concepts, while probing the locus of this effect. In Experiment 1, deliberately generating conceptually incorrect definitions produced superior cued recall test performance than errorless copying, even when the latter involved underlining and writing the correct targets more times. Furthermore, correcting one's deliberate errors yielded an additional recall advantage over leaving them uncorrected, providing the first indication that deliberate errors potentiate encoding of their subsequent correction. Experiment 2 replicated this finding and pursued this account further by showing that the enhanced processing of the target definitions was specific to having first deliberately produced an incorrect, but not alternative correct, response. Although generating correct conceptual synonyms outperformed copying (i.e., no generation), deliberate erring with or without correction surpassed both of these errorless conditions. This result simultaneously suggested that the derring effect is not merely a generation benefit that follows the production of any other novel (correct) response. Experiment 3 additionally showed that elaboration alone is insufficient to fully explain the derring effect, which reliably persisted even when errorless learning had been bolstered with a higher degree of elaboration. Specifically, generating examples of the concepts induced greater elaboration and better learning performance than generating conceptual synonyms, but both errorless methods still fell short of generating deliberate errors and correcting them.

Overall, the present research findings demonstrate the effectiveness of deliberate erring in promoting concept learning, particularly when it is accompanied by correction. Attesting to the strength of this counterintuitive learning strategy, the benefits of deliberate error commission and correction over errorless learning were robust across all three experiments, with medium to large effect sizes (Cohen's *d*) ranging from 0.43 when compared against example generation (Experiment 3), an average of 0.89 when compared against synonym generation (Experiment 2 = 0.97; Experiment 3 = 0.81), to 1.30 when compared against copying (Experiment 1).

Despite the utility of deliberate erring, participants in the present study often displayed limited metacognitive knowledge of its benefits and even inaccurately predicted that this strategy would be *less* effective for their learning than copying (Experiment 1), when their test performance in fact revealed the exact opposite. This result echoes the metacognitive illusions observed in previous error research (Huelser & Metcalfe, 2012), in which learners are often unaware that error generation has helped their learning even after having just experienced its benefits for their performance. Unsurprisingly, lacking awareness of the positive effects of erring may contribute to error aversion, whereby people dislike making errors or being seen making them in school and at work (see Frese & Keith, 2015 for a review). As our findings suggest, however, avoiding errors may be counterproductive in low-stakes learning contexts. Ultimately, failing to capitalize on our errors and learn from them may be the greater error.

Theoretical Explanations for the Derring Effect

Why does deliberate error commission and correction enhance learning? Whereas there have been several prominent theoretical accounts that are amenable to explaining other commonly investigated classes of errors (e.g., "naturalistic" errors, incorrect guesses), they do not lend themselves readily to the derring effect. As outlined in the Introduction, one potential explanation for the benefits of error generation is that the error itself serves as a semantic mediator or an alternative retrieval route to the correct target when an elaborate mental network is created during the process of responding incorrectly (Cyr & Anderson, 2015; Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012). However, this account is incompatible with two of our findings. First, correcting one's deliberate errors produced a mnemonic advantage over not correcting them (Experiments 1 and 2). If the derring effect were solely attributable to the semantic mediator hypothesis, then error correction should have been superfluous after a rich mental network had presumably already been formed during error generation. Coupled with the fact that learners already knew the targets when deliberately erring, the correction would hardly appear to be useful, contrary to the benefit that we observed. Second, deliberate erring outperformed elaborating on the targets by generating alternative correct definitions (conceptual synonyms; Experiments 2 and 3) or even illustrative examples (Experiment 3), even though both errorless techniques arguably produced responses that were more strongly related to the targets and thus more likely to serve as useful semantic mediating information.

It is also difficult to reconcile our findings with the explanation that learners' enhanced recall of the targets reflected an ironic rebound effect (Wegner et al., 1987) from having mentally suppressed them during deliberate error commission. If so, then preventing learners from expressing the suppressed targets during error correction should have fueled the rebound effect (Wenzlaff & Wegner, 2000; see also Baddeley, 1963; Zeigarnik, 1938), such that uncorrected targets would have been recalled better at test than corrected targets. Yet, we found the reverse.

One potential theoretical account is that the act of deliberately committing errors enhances subsequent encoding of the targets. By this view, when an error has been deliberately made, a window of opportunity may be introduced for more effective processing of the target when it is reencountered during correction. When learners do not correct their deliberate errors, this opportunity is missed. Consequently, although both errorful and errorless generation improve learning more than copying (i.e., no generation), errorful generation alone does not offer additional gains over errorless example generation that has been boosted with greater elaboration, as revealed by our cross-experiment analyses. However, if the opportunity is seized by correcting one's deliberate errors, then a learning advantage may be accrued. This account resonates with reconsolidation theory (Nader & Hardt, 2009; Nader et al., 2000; for a discussion, see Metcalfe, 2017), which postulates that existing memory traces are destabilized and placed in a labile state when they are retrieved, such that these active traces are transiently sensitive to modification and can be reconsolidated with subsequent learning events. Likewise, the coupling of deliberate error commission and correction may form a potent learning technique by strengthening encoding of the target response during correction after an error has been made.

Why does deliberate error generation enhance subsequent target processing, and why is this effect unique to generating incorrect but not alternative correct responses? Our data do not speak directly to these questions, but some explanations are more plausible than others given the paradigm and evidence at hand. For one, whereas curiosity to learn the answers may account for better processing of corrective feedback after generating incorrect guesses to novel stimuli such as translations for foreign vocabulary (Potts et al., 2019; Potts & Shanks, 2014), the same cannot be said for deliberate errors—because learners already know the targets during open-book study, there is no information gap to be filled after they err. Thus, it is also unlikely that enhanced target processing in the derring effect stems from any element of surprise at the apparent discrepancy between one's deliberate errors and the correct answers (e.g., Butterfield & Metcalfe, 2001). Rather, one possibility is that this benefit draws on episodic recollection (Tulving, 1972). In other words, deliberately committing errors may direct learners' attentional focus to the correction, thereby enhancing its encoding by creating an episodically memorable event (Metcalfe & Huelser, 2020). After having intentionally generated an error, the correct target may appear more distinctive by comparison, thus affording a more memorable episodic trace than if the preceding error had not been made or if the target had followed another similarly correct alternative. Further research is needed to illuminate these potential processes.

Future Directions

In the present study, we found that deliberately committing and correcting errors was an effective strategy in promoting learners' knowledge retention of educationally relevant scientific term-definition concepts. From both educational and theoretical perspectives, validating the derring effect across diverse knowledge domains, learning materials, assessments, error types, and learner characteristics will yield farther insight on when and how best to apply deliberate erring to enhance learning across a wide range of educational settings.

In particular, it will be valuable for further investigations to generalize the derring effect to complex, higher order learning outcomes. In Bloom's (1956) classic taxonomy of educational objectives, for instance, such higher order learning outcomes include applying, analyzing, evaluating, and creating information, beyond remembering and understanding it. To the degree that better knowledge retention facilitates transfer of that knowledge to novel contents and contexts (Barnett & Ceci, 2002; Butler, 2010; cf. Agarwal, 2019), it is possible that the benefits of deliberate erring may extend beyond the retention of a specific response.

Exploring the effects of different kinds of deliberate errors across various educational materials will also be useful in determining the contexts in which deliberate erring is likely to be more effective, beyond the present study's focus on conceptual errors when learning scientific term-definition concepts. For instance, procedural errors or "bugs" (Brown & Burton, 1978) are common in mathematical problem-solving, whereby errors arise from learners' use of inappropriate or faulty procedures (e.g., substituting incorrect values into a formula), rather than absent or incorrect concepts and principles (Engelhardt, 1982; O'Connell, 1999). Accordingly, one potential avenue for future work is to examine the extent that deliberately committing and correcting procedural errors enables learners to solve problems more effectively and efficiently, while improving their ability to apply learned procedures to solve novel problems. Notably, in the domain of probability problem-solving, learning with worked examples containing incorrect solution steps rather than only correct ones has been found to enhance far transfer to novel test problems with dissimilar surface and structural features, particularly for learners with high prior knowledge (Große & Renkl, 2007). It will thus be worthwhile to consider how deliberate erring can be integrated with such methods of instruction, in conjunction with examining potential interactions with learner characteristics. For instance, whereas the university students in our study were able to effectively commit and correct deliberate conceptual errors in a knowledge domain that they possessed minimal prior familiarity with, younger students may require additional guidance or scaffolding to successfully implement deliberate errors in their learning.

Practical Implications for Education

Our findings suggest that preventing errors in low-stakes learning contexts may, in itself, be an error. Rather, to improve learning, teachers can actively promote errors in the classroom by guiding students into deliberately committing and correcting them, particularly conceptually plausible but inaccurate responses that can also stimulate class discussion. In addition, teachers can design homework assignments that systematically incorporate deliberate erring and correction. Through these experiences, students may be propelled to consistently stay a step ahead in their learning by uncovering potential gaps in their understanding, intentionally embracing these pitfalls, and transforming them into positive educational opportunities. At the same time, by muting any ego threats, deliberate errors allow for errors to be reframed as meaningful teachable moments rather than as debilitating events (Eskreis-Winkler & Fishbach, 2019). This may build an error-friendly learning climate in classrooms, toward developing a culture of investing in our errors and harnessing the potential to learn from them (Wong & Lim, 2019). Indeed, it may be that "those who are afraid to take risks or make mistakes restrict their capacity to learn" (Fisher & Lipson, 1986, p. 788).

Given that students often appear to be unaware of the benefits of deliberate erring, it may be crucial for teachers to explicitly address such metacognitive illusions. By fostering awareness of what techniques are effective for learning and how to implement them effectively, teachers can ensure that their students do not persist in using suboptimal methods during their self-regulated study routines but instead take advantage of strategies informed by the cognitive science of learning to experience meaningful gains.

Conclusion

In their long and turbulent history, errors have often been avoided in learning, up till contemporary approaches that have adopted relatively more tolerant attitudes toward errors by allowing or inducing them to occur. Preventing errors or leaving them to chance, however, hinders us from systematically learning from them and fully optimizing the benefits of the teachable opportunities that they can be. Here, we suggest that one novel and counterintuitive solution to unleash the potential of errors is to purposefully embrace them in low-stakes educational contexts. Building on the evolving tradition of errors, the present research has demonstrated that deliberate erring can be a powerful strategy to enhance learning, particularly when we correct our errors. In response to the student in Glasbergen's cartoon who asked, "If we learn from our mistakes, shouldn't I try to make as many mistakes as possible?," it appears that the teacher should answer, "Indeed, perhaps we should—*deliberately*."

References

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

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

Baddeley, A. D. (1963). A Zeigarnik-like effect in the recall of anagram solutions. *The Quarterly Journal of Experimental Psychology*, *15*(1), 63–64. https://doi.org/10.1080/17470216308416553

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

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

Breedlove, S. M., & Watson, N. V. (2018). *Behavioral neuroscience* (8th ed.). Oxford University Press.

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, *2*(2), 155–192. https://doi.org/10.1016/S0364-0213(78)80004-4

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

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918–928. https://doi.org/10.1037/0278-7393.34.4.918

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491–1494. https://doi.org/10.1037/0278-7393.27.6.1491

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Cyr, A.-A., & Anderson, N. D. (2015). Mistakes as stepping stones: Effects of errors on episodic memory among younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 841–850. https://doi.org/10.1037/xlm0000073

Daley, N., & Rawson, K. A. (2019). Elaborations in expository text impose a substantial time cost but do not enhance learning. *Educational Psychology Review*, *31*(1), 197–222. https://doi.org/10.1007/s10648-018-9451-9

Dornisch, M., Sperling, R. A., & Zeruth, J. A. (2011). The effects of levels of elaboration on learners' strategic processing of text. *Instructional Science*, *39*(1), 1–26. https://doi.org/10.1007/s11251-009-9111-z

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

Engelhardt, J. M. (1982). Using computational errors in diagnostic teaching. *The Arithmetic Teacher*, *29*(8), 16–19. https://doi.org/10.5951/AT.29.8.0016

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725–747. https://doi.org/10.1037/0003-066X.49.8.725

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

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, K. M., & Lipson, J. I. (1986). Twenty questions about student errors. *Journal of Research in Science Teaching*, *23*(9), 783–803. https://doi.org/10.1002/tea.3660230904

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

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513. https://doi.org/10.3758/s13421-011-0174-0

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

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

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

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

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

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (J. H. Byrne, Series Ed., pp. 487–514). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 1270–1328). International Reading Association.

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), 191–205. https://doi.org/10.1016/0361-476X(88)90020-3

Liberman, N., & Förster, J. (2000). Expression after suppression: A motivational explanation of postsuppressional rebound. *Journal of Personality and Social Psychology*, *79*(2), 190–203. https://doi.org/10.1037/0022-3514.79.2.190

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

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

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*(2), 159–188. https://doi.org/10.1037/0096-3445.114.2.159

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4-5), 494–513. https://doi.org/10.1080/09541440701326154

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

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

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225–229. https://doi.org/10.3758/BF03194056

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

Nader, K., & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews Neuroscience*, *10*(3), 224–234. https://doi.org/10.1038/nrn2590

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726. https://doi.org/10.1038/35021052

O'Connell, A. A. (1999). Understanding the nature of errors in probability problem-solving. *Educational Research and Evaluation*, *5*(1), 1–21. https://doi.org/10.1076/edre.5.1.1.3887

Pan, S. C., & Rickard, T. C. (2017). Does retrieval practice enhance learning and transfer relative to restudy for term-definition facts? *Journal of Experimental Psychology: Applied*, *23*(3), 278–293. https://doi.org/10.1037/xap0000124

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Pitt, E., & Norton, L. (2017). 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, *42*(4), 499–516. https://doi.org/10.1080/02602938.2016.1142500

Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1023–1041. https://doi.org/10.1037/xlm0000637

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644–667. https://doi.org/10.1037/a0033194

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

Rawson, K. A., & Dunlosky, J. (2016). How effective is example generation for learning declarative concepts? *Educational Psychology Review*, *28*(3), 649–672. https://doi.org/10.1007/s10648-016-9377-z

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*(3), 483–504. https://doi.org/10.1007/s10648-014-9273-3

Reason, J. (1990). *Human error*. Cambridge University Press.

Roelle, J., & Nückles, M. (2019). Generative learning versus retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology*, *111*(8), 1341–1361. https://doi.org/10.1037/edu0000345

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1–31. https://doi.org/10.1207/s15326977ea1101_1

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

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). Academic Press.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, *53*(1), 5–13. https://doi.org/10.1037/0022-3514.53.1.5

Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, *51*, 59–91. https://doi.org/10.1146/annurev.psych.51.1.59

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

Woloshyn, V. E., Pressley, M., & Schneider, W. (1992). Elaborative-interrogation and prior-knowledge effects on learning of facts. *Journal of Educational Psychology*, *84*(1), 115–124. https://doi.org/10.1037/0022-0663.84.1.115

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

Zeigarnik, B. (1938). On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 300–314). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-025

Received March 20, 2020
Revision received February 18, 2021
Accepted February 26, 2021 ■