

# Journal of Educational Psychology

## Deliberate Errors Promote Meaningful Learning

Sarah Shi Hui Wong and Stephen Wee Hun Lim

Online First Publication, December 30, 2021. <http://dx.doi.org/10.1037/edu0000720>

### CITATION

Wong, S. S. H., & Lim, S. W. H. (2021, December 30). Deliberate Errors Promote Meaningful Learning. *Journal of Educational Psychology*. Advance online publication. <http://dx.doi.org/10.1037/edu0000720>

# Deliberate Errors Promote Meaningful Learning

Sarah Shi Hui Wong and Stephen Wee Hun Lim

Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore

Our civilization recognizes that errors can be valuable learning opportunities, but for decades, they have widely been avoided or, at best, allowed to occur as serendipitous accidents. The present research tested whether greater learning success could paradoxically be achieved through making errors by intentional design, relative to traditional errorless learning methods. We show that deliberately committing and correcting errors even when one knows the correct answers enhances learning—a counterintuitive phenomenon that we termed the *derring effect*. Across two experiments ( $N = 160$ ), learners engaged in open-book study of scientific expository texts and were then tested on their retention and higher-order application of the material to analyze a novel news event. Deliberate error commission and correction during study produced not only better recall performance, but also superior knowledge application compared to two errorless study techniques that are popular among students and educational researchers: copying with underlining, and elaborative studying with concept mapping (Experiment 1). These learning benefits persisted even over generating alternative conceptually correct answers, revealing that the *derring effect* is not merely attributable to generation or elaboration alone, but is unique to producing incorrect responses (Experiment 2). Yet, learners were largely unaware of these advantages even after experiencing them. Our results suggest that avoiding errors in learning may not always be most optimal. Rather, deliberate erring is a powerful strategy to enhance meaningful learning. We discuss implications for educational practice in redesigning conventional approaches to errors: To err is human; to *deliberately* err is divine.

### **Educational Impact and Implications Statement**

Errors are ripe with learning opportunity, but have traditionally been avoided or simply allowed to occur spontaneously. Here, we provide evidence for the *derring effect*—deliberately committing and correcting errors even when we know the correct answers enhances learning. Relative to popular errorless techniques such as copying with underlining, concept mapping, and even generating alternative correct answers, deliberate erring improved not only knowledge retention, but also higher-order application of learning. Yet, students were largely unaware of these benefits. Deliberate erring is a powerful way of learning from failure for greater success.

**Keywords:** errors, higher-order learning, knowledge application, meaningful learning, metacognition

**Supplemental materials:** <https://doi.org/10.1037/edu0000720.supp>

Sometimes the joins  
are so exquisite

they say the potter  
may have broken the cup  
just so he could mend it.

—Chana Bloch, *The Joins*

In the traditional Japanese art of *kintsugi*, broken ceramics are given a new lease of life by mending them with lacquer dusted with gold—instead of being concealed, the cracks on the damaged artifacts are illuminated with seams of gold and valued as part of the objects' unique history. Indeed, errors are inevitable in life. However, while we may recognize the importance of learning from our errors, this is often easier said than done.

Over many decades, errors have traditionally been viewed as aversive events to be avoided at all costs in learning (Ausubel, 1968; Bandura, 1986; Skinner, 1958; Terrace, 1963)—an idea that has become deeply entrenched in educational systems (Metcalf, 2017) and organizations (Fresse & Keith, 2015). Yet, a growing body of research in cognitive and educational psychology has challenged the utility of not actively engaging with errors in low-stakes contexts since they can be valuable learning opportunities when accompanied by corrective feedback (for reviews, see Kapur, 2016; Metcalfe, 2017; Wong & Lim, 2019). If errors are avoided entirely, then we also miss out on their potential benefits. With such hurdles lining the road to learning from failure, how can errors be strategically positioned to overcome the odds?

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 premiated work was conferred the 2020 APS Student Research Award.

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, 117570, Singapore. Email: [psywshs@nus.edu.sg](mailto:psywshs@nus.edu.sg)

Previous research on errors in educational and workplace contexts has focused almost exclusively on “naturalistic” or induced errors that occur despite learners’ intentions to get things right, such as when learners inadvertently respond incorrectly and are unaware that they have erred until receiving feedback (Metcalfe & Xu, 2018) or when learners do not know the correct answers but make their best guess nevertheless (Kang et al., 2011; Kornell et al., 2009). However, waiting for errors to occur spontaneously does not allow for learning opportunities to be maximized. Rather, to take full advantage of errors, one solution lies in strategically redesigning them as systematic and intentional learning events.

The present research’s central premise is precisely this: *Deliberate* erring enhances learning, even when one already knows the correct answers. Whereas errors have typically been prevented or, at best, permitted in learning as serendipitous accidents, actively promoting errors by deliberately committing and correcting them may, in fact, produce greater learning gains. In the following sections, we consider how errors have traditionally been approached in education and how they can be optimally positioned to enhance learning through deliberate intention.

### Extant Approaches to Errors in Learning

Broadly, errors can be viewed as objectively incorrect responses that may deviate from the correct ones in various ways (e.g., misspellings, conceptual errors, procedural errors, etc.) and that may arise from different approaches and intentions. In a recent review, Wong and Lim (2019) proposed a Prevention–Permission–Promotion (3P) framework whereby errors in learning can be broadly approached in three ways: *prevention* (avoiding or observing errors), *permission* (allowing errors), and *promotion* (inducing or guiding errors). Under the error prevention approach, firsthand errors are avoided by preventing learners from being exposed to them (e.g., by providing learners only with correct information through worked examples and solutions; for a review, see Atkinson et al., 2000) or by having learners observe others’ errors instead. Avoiding errors has been shown to offer some benefits for novices (e.g., Sweller & Cooper, 1985) and in relatively less complex tasks (e.g., Terrace, 1963). However, the extent that error prevention benefits higher-order learning—a critical component of education (for a discussion, see Agarwal, 2019)—remains more tenuous. Indeed, extant research suggests that learners may miss out on valuable opportunities to fully benefit from errors if they do not actively commit them personally (Metcalfe & Xu, 2018), while learning how to effectively manage errors when they eventually do occur (Frese & Keith, 2015). Given that errors are bound to arise organically during the learning process, it may be unrealistic and even futile to attempt to prevent them entirely in educational contexts.

The approach of error permission resolves this problem by passively allowing errors to occur when learners engage in active exploration during “discovery learning” (e.g., Alfieri et al., 2011; Bruner, 1961). In other words, errors are neither prevented nor encouraged, but are simply permitted to arise in naturalistic ways. Allowing learners to commit errors firsthand has been found to produce better learning than merely exposing them to others’ incorrect responses (Metcalfe & Xu, 2018). However, because of the incidental and unpredictable nature of such errors, this approach poses challenges for providing systematic feedback, without which errors are unlikely to be spontaneously corrected

(Butler et al., 2008; for reviews on the importance of feedback for learning, see also Black & Wiliam, 1998; Hattie & Timperley, 2007).

In turn, actively promoting errors involves greater intentionality in eliciting learners’ errors. For instance, learners can be actively induced to err through purposefully adding challenge to the learning task or withholding information from them (Lorenzet et al., 2005), while forcing learners to generate guesses even when they do not know the correct answers (e.g., Cyr & Anderson, 2015; Kang et al., 2011; Kornell et al., 2009; Potts & Shanks, 2014). Notably, “productive failure” involves purposefully delaying instructional structure by having learners first engage in unscaffolded problem solving of complex and ill-structured problems such that errors inevitably occur, before receiving instruction (Kapur, 2008, 2016; Kapur & Bielaczyc, 2012). Likewise, error management training in workplace contexts involves active exploration with minimal guidance, while explicitly encouraging learners to make errors (e.g., Frese & Altmann, 1989; Keith & Frese, 2005). Relative to allowing errors to occur spontaneously, inducing errors may enable teachers to anticipate their students’ errors and prepare appropriate feedback. However, successful uptake of this feedback may be hampered when learners experience increased frustration and reduced self-efficacy (or “emotional backwash”; Pitt & Norton, 2017), particularly if they attribute their errors to internal causes such as poor ability (Lorenzet et al., 2005). Because failure hurts the ego, learners often disengage or “tune out” from the task upon receiving negative feedback, consequently learning less from their failures than successes (Eskreis-Winkler & Fishbach, 2019).

Accordingly, an alternative means of promoting errors is to guide learners into committing and correcting specific errors in a structured manner (Lorenzet et al., 2005; Wong & Lim, 2019, 2021). In this way, ego concerns and negative emotional responses can be minimized by allowing learners to veridically attribute their errors to the learning approach rather than dispositional factors, while enabling them to recognize incorrect responses and strategies that should be avoided in the future. Besides featuring the element of active, firsthand generation that has contributed to the success of allowing errors (Metcalfe & Xu, 2018), guided errors can be introduced and corrected more systematically in educational contexts such as students’ self-regulated learning, relative to “naturalistic” errors.

Yet, empirical work on guiding learners into committing and correcting errors remains curiously sparse, despite the potential of this approach to enhance learning. Moreover, whereas a myriad of causes may produce different types of errors such as accidental slips, mistakes that arise from incorrect knowledge, and even deliberate violations that intentionally deviate from socially appropriate practices (Reason, 1990), few studies have closely examined learners’ intentions when erring. In particular, guiding learners to *deliberately* err in low-stakes contexts may be a novel and systematic strategy that effectively harnesses the potential of errors to optimize learning opportunities (Wong & Lim, 2021).

### Learning From Deliberate Errors

Deliberately committing and correcting errors even when we already know the answers is a counterintuitive approach, especially when held up against the traditional view that errors should be

avoided since they may inhibit learning when they are ingrained and reproduced in the future (e.g., Ausubel, 1968; Skinner, 1958). Alternatively, one may expect that intentionally generating incorrect responses while being aware of the correct answers is superfluous, such that learning is neither impaired nor improved.

However, cognitive psychological principles suggest several compelling reasons to expect that learning may actually profit from deliberate erring. For one, deliberately generating errors may enhance memory by making the encoding of subsequent correction more effective (Hays et al., 2013; Kornell et al., 2009; Potts et al., 2019; Potts & Shanks, 2014). For instance, a more distinctive and memorable episodic trace may be fostered when the commission of an error itself draws attention to the correction, as opposed to having processed only correct information throughout (Metcalf & Huelser, 2020). It is also possible that exploring incorrect retrieval routes during deliberate erring may ironically weaken and cull those unproductive routes, thus boosting the relative retrieval strength of the correct response (Kornell et al., 2009). Moreover, the activation of associated concepts during error generation may foster deeper elaborative encoding processes (Craik & Lockhart, 1972; Craik & Tulving, 1975), thereby forming a richer mental network that facilitates subsequent retrieval (Huelser & Metcalfe, 2012; Kornell et al., 2009; Potts & Shanks, 2014). For instance, errors may serve as useful semantic mediators or “stepping stones” that cue or scaffold retrieval of the correct target (Cyr & Anderson, 2015). Taken together, these potential theoretical accounts motivate the expectation that deliberate erring may, counterintuitively, benefit learning more than traditional errorless techniques.

## The Present Study

The overarching goal of the present research was to test the hypothesis that guiding learners to deliberately commit and correct errors even when they know the correct answers produces superior learning than avoiding errors—a phenomenon that we termed the *derring effect* (Wong & Lim, 2021). Importantly, learning does not solely involve acquiring knowledge by remembering or understanding it, but also includes higher-order cognitive processes such as meaningfully using or applying one’s knowledge in novel situations (Bloom, 1956). For instance, after studying a lesson on food allergies, a retention test might ask students to recall the biological processes involved in food allergies. In contrast, a higher-order application test might present students with a novel case study on an allergic reaction, and ask students to apply their knowledge by developing predictions about the patient’s medical history and proposing potential causes for the reported symptoms to formulate a diagnosis. This latter test involves *meaningful learning*, which has often been viewed as a crucial educational goal (Mayer, 2002). As opposed to mere rote learning, meaningful learning involves the construction of organized, coherent, and integrated mental models that enable learners to successfully make inferences and apply their knowledge when faced with new problems or learning situations (Mayer, 1984, 2002). Accordingly, we examined the effects of deliberate erring on not only learners’ retention of knowledge, but also their ability to meaningfully apply that knowledge.

In two experiments, we investigated the derring effect when learners studied educationally relevant scientific expository texts either by deliberately committing and correcting conceptual errors

or by engaging in traditional errorless learning techniques. For generalizability of our findings, we adopted two scientific texts that described complex phenomena in two knowledge domains—geography (“volcanoes”) and medicine (“food allergies”). Each text contained interrelated elements and ideas, thus constituting relatively complex educational materials (for discussions, see Karpicke & Aue, 2015; Sweller, 2010). Learners studied the texts under open-book conditions, similar to how students are typically able to refer to their textbook or notes during their self-regulated learning. Subsequently, learners were tested on their basic retention of the material, as well as their higher-order learning in a relatively more complex educational task—applying the learned information to analyze a novel news event. We also assessed learners’ metacognitive knowledge about the effectiveness of deliberate erring versus errorless learning, in light of research suggesting that students are often unaware of the benefits of errors for their test performance (e.g., Huelser & Metcalfe, 2012; Pan et al., 2020; Yang et al., 2017), which has causal consequences for their metacognitive control in selecting study strategies that are most effective for their learning (Metcalf & Finn, 2008; Thiede et al., 2003).

## Experiment 1

Experiment 1 tested the derring effect in learners’ knowledge retention and higher-order application performance by comparing deliberate erring against traditional errorless learning techniques. Whereas previous error research has often pitted active errorful learning against passive errorless learning methods such as reading only (Huelser & Metcalfe, 2012; Kornell et al., 2009), we used two active errorless learning controls that are popular among students and educational researchers to dissociate the effects of deliberate erring from those of active learning (Freeman et al., 2014): copying with underlining, and elaborative studying with concept mapping.<sup>1</sup>

Underlining is a popular study strategy that students frequently report adopting, alongside the functionally similar technique of highlighting (Dunlosky et al., 2013), with a meta-analyzed self-reported frequency of use of 53% (Miyatsu et al., 2018). The technique of underlining is relatively easy to use and has even been spontaneously adopted by learners as young as those in the fifth grade (Brown & Smiley, 1978). Learner-generated underlining has been shown to benefit memory, particularly for marked compared to unmarked information, across a variety of assessments such as free-recall, short-answer, and fill-in-the-blank tests (Miyatsu et al., 2018), although its efficacy for higher-order learning outcomes such as inferencing remains questionable (Peterson, 1991). Presumably, underlining aids learning by encouraging elaborative processing through the active selection of important text content

<sup>1</sup> Other popular study techniques that students have routinely reported adopting include rereading and flash cards (i.e., self-testing; retrieval practice), with meta-analyzed frequencies of use of 78% and 55%, respectively (Miyatsu et al., 2018). However, both of these techniques were less suited as active errorless comparison methods in our study—rereading is largely passive in nature despite being error-free, whereas the use of flash cards inherently introduces naturalistic errors when learners inadvertently recall incorrect information or fail to recall it during study, although retrieval practice has been established as an effective learning technique for retention (Dunlosky et al., 2013).

(Dunlosky et al., 2013), such that learning benefits even without a subsequent opportunity to review one's underlined or marked text (e.g., Fass & Schumacher, 1978; Kulhavy et al., 1975).

Likewise, the construction of concept maps—a type of graphic organizer with nodes representing key concepts and links representing their relations—is a learning technique that has garnered significant interest among educational researchers over decades (Nesbit & Adesope, 2006), with students also reporting preferences for using graphic organizers in their learning (Wang et al., 2021). Concept mapping has been viewed as an effective generative learning strategy because it promotes elaborative processing when learners identify key conceptual ideas, organize these ideas' relations, and integrate the new information with their prior knowledge by translating and structuring it in a spatially arranged concept map (Fiorella & Mayer, 2016). Notably, concept mapping has yielded moderate learning benefits over a range of other instructional comparison conditions in increasing not only knowledge retention, but also higher-order transfer or application of learning (for meta-analyses, see Nesbit & Adesope, 2006; Schroeder et al., 2018). For instance, the use of concept mapping to study educational texts has been found to produce greater gains on a test with knowledge, comprehension, and application questions, relative to individual study plus discussion (Chularut & DeBacker, 2004).

As illustrated here, both copying with underlining and concept mapping have been viewed to involve one or more of the three primary cognitive processes that have been proposed to be vital for meaningful learning in Mayer's (1996) Select–Organize–Integrate (SOI) model: selecting relevant information, organizing the selected information into a coherent structure, and integrating the constructed knowledge with one's existing knowledge in long-term memory. Specifically, copying with underlining involves selecting specific information from the text, whereas concept mapping involves selecting, organizing, and integrating information when representing it visually and spatially. Hence, both errorless learning techniques served as potentially strong contenders in boosting meaningful learning in our study, although we expected that concept mapping would be relatively more effective than copying with underlining due to its generative—organizing and integrating—components, which we examined by scoring the structural complexity of learners' concept maps (Novak & Gowin, 1984).

Accordingly, Experiment 1 pitted the novel strategy of deliberate error commission and correction against both errorless learning methods. In addition, we assessed learners' metacognitive awareness about each method's effectiveness before and after being tested.

## Method

### Participants

The participants were 120 undergraduate students (87 were female) between the ages of 18 and 28 ( $M = 20.18$ ,  $SD = 1.64$ ) from the National University of Singapore. The target sample size was determined based on the average effect size of learning by constructing concept maps ( $g = 0.72$ ) reported in Schroeder et al.'s (2018) meta-analysis of 142 independent effect sizes. A power analysis (G\*Power; Faul et al., 2007) indicated that at least 32 participants per condition would afford 80% power in the present study to detect a concept-mapping effect for two-tailed between-subjects pairwise comparisons using an alpha of .05.

In both experiments, all participants reported English as their first language and received either course credit or cash reimbursement for their participation. Both experiments received and were conducted with ethics approval from the National University of Singapore's institutional review board, and all participants granted their written informed consent.

### Design

The primary between-subjects factor of interest was learning method: *copy* versus *concept-map* versus *concept-error* (deliberate error). For control purposes, we included study text (“volcanoes” vs. “food allergies”) as a second between-subjects factor to ensure that effects, if any, persisted across text topics. The dependent variables were participants' scores on an application test and free-recall test.

### Materials

**Educational Texts.** The study texts were two expository passages that described complex phenomena in geography (“volcanoes”) and medicine (“food allergies”). Both study texts were adapted from Griffin et al. (2019) and were trimmed to each contain exactly 310 words. Each study text was arranged in four paragraphs and comprised 20 sentences. For scoring purposes, we identified 40 idea units in each study text. The study texts had Flesch-Kincaid grade levels of 11 and 12, respectively, and Flesch reading ease scores of 42 and 34. Both study texts are available in the online supplemental materials.

In addition, a brief 42-word (four-sentence) paragraph served as the practice text. The practice text was on a topic (“muscle tissue”; adapted from Karpicke & Blunt, 2011b) that did not relate to either of the critical study texts.

**News Articles.** For the application test, we constructed two news article excerpts corresponding to the study texts on “volcanoes” and “food allergies,” respectively. Both news articles briefly described actual historical events—the 1980 eruption of Mount St. Helens volcano versus an incident of a young boy, Braxton Ong, who suffered a life-threatening allergic reaction—and were written based on the corresponding news reports from *The New York Times* and *The Straits Times*, respectively. The news articles were each 118 words in length and arranged in three paragraphs. They were designed to closely resemble actual news reports with a headline, a subheading, and an accompanying photo. Both news articles are available in the online supplemental materials.

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

### Procedure

Prior to attending the experiment, participants completed the English language proficiency test via an online questionnaire. Upon their arrival at the laboratory, participants sat at individual computer cubicles and were told that they would be studying a given text, with the expectation that they would later be tested on the material. The specific nature of the tests was not divulged. Participants were randomly assigned to the copy, concept-map, or concept-error condition. Within each learning condition, half of the participants were randomly allocated to study the “volcanoes”

text, whereas the other half studied the “food allergies” text. There were three experimental phases: practice, studying, and test.

**Practice Phase.** Participants were first introduced to the learning method that they had been randomly assigned to and practiced using the method to study the practice text. In the copy condition, participants wrote down the text exactly as it was presented. They were also instructed to identify and underline the key concepts in each sentence. As an illustration, for the training example “Bats are mammals that fly,” a sample response was “Bats are mammals that fly.” Underlining the key concepts was intended to be behaviorally comparable to the act of striking out—drawing a line across—one’s deliberate errors in the concept-error condition.

In the concept-map condition, participants were instructed to create a concept map of the concepts in the text. The characteristics of concept maps (e.g., representing concepts in nodes and showing their relations with links) were explained to participants, alongside examples of well-constructed concept maps (adapted from Novak, 2005). Participants were also told to ensure that the information from each sentence in the text had been represented in their concept map. This ensured that participants fully processed the text in its entirety, rather than studying a mere fraction of it if they so happened to select only a few concepts to be included in their map.

In the concept-error condition, participants deliberately erred by writing down each sentence in the text such that it contained a plausible conceptual error (i.e., an error in understanding or interpreting a concept), before striking out this error and correcting it by writing down the actual concept exactly as it was presented in the text. For instance, a sample response for the training example was “Bats are ~~birds~~ (mammals) that fly.” During the practice phase, participants were provided with examples of good responses that were considered conceptually wrong versus poor responses that were conceptually correct but merely erroneous in other ways (e.g., “Bats are mammals that fly” is conceptually correct but merely involves a spelling error). In line with previous research on incorrect guessing (Kang et al., 2011) and competitive incorrect responses (Little & Bjork, 2015), participants were encouraged to deliberately make plausible conceptual errors—responses that were objectively wrong but that were believable. For instance, “Bats are birds that fly” is a relatively more plausible conceptual error than “Bats are humans that fly.”

**Studying Phase.** After completing the practice phase, participants were given 1.5 min to read a printed handout of the study text (either “volcanoes” or “food allergies”). Pilot testing indicated that this time duration was sufficient for participants to comfortably read through the whole text. Following which, participants were provided with a blank sheet of paper and were given 25 min to study the text using the learning method that they had been randomly assigned—to copy the full text and underline its key concepts, construct a hand-drawn concept map of the text, or write the full text while deliberately erring. Taking the statement “Magmas that are low in silica tend to be very fluid” from the “volcanoes” text as an example, a sample response in the copy condition was “Magmas that are low in silica tend to be very fluid.” In the concept-map condition, a sample response was to represent “magma,” “silica,” and “fluid” in nodes and to hierarchically organize them via links to depict their relations, while further connecting these nodes to other related key concepts from the text. In the concept-error condition, a sample response was “Magmas that are low in silica tend to be very ~~viscous~~ (fluid).” Participants were also told

that if they finished before the time limit was up, they should spend the remaining time reviewing their response to ensure that they had included all sentences from the text. Thus, the total studying duration was identical across all three learning conditions.

Next, participants responded to a three-item questionnaire. They (a) made a judgment of learning (JOL) on an 11-point scale from 0% to 100% to predict how much of the material from the study text they would remember later, (b) rated the extent to which the study text was understandable (1 = *not at all*; 7 = *extremely*), and (c) indicated how well they knew the subject matter covered in the study text prior to reading it (1 = *not very well*; 7 = *very well*).

**Test Phase.** Participants then completed a brief distractor task before undergoing an application test followed by a free-recall test.<sup>2</sup> Participants completed both tests without referring to the study text. In the application test, participants were presented with the news article on the same topic as the text that they had earlier studied and were asked to write down their response to the following prompt:

Analyze the event reported in the news article: Applying what you have learned from the prose passage that you studied earlier, (a) what might be [the formation history of Mount St. Helens volcano]/[the medical history of Braxton Ong], and (b) the causes of [its major eruption]/[his serious allergic reaction]? Please answer in as much detail as possible.

In the free-recall test, participants wrote down in any order as much as they could remember from the study text. No time limit was imposed for both tests. Finally, participants rated the effectiveness of their respective learning method in helping them learn the study text on a 7-point scale (1 = *not at all*; 7 = *extremely*). Participants were then debriefed and thanked.

## Results

### Scoring

Participants’ application test responses were scored by awarding 1 point for each idea unit from the study text that, crucially, had been correctly applied to analyze the specific news event. That is, participants’ responses had to effectively answer the specific application question within the context of the news event to be awarded points; idea units that were erroneously or ineffectively applied were not awarded points, even if they were correctly recalled. For example, for the “volcanoes” text, responses suggesting that Mount St. Helens volcano could have been formed by an oceanic-continental or oceanic-oceanic plate convergence would earn 1 point each, with further points awarded for appropriate explanations of the particular plate tectonics processes that could specifically have caused Mount St. Helens volcano’s formation and major eruption. However, a response that incorrectly suggested a continental-continental plate convergence would not earn any

<sup>2</sup>The application test was administered first because learners’ performance on this meaningful learning outcome was of relatively greater theoretical interest in the present study. It should be noted, however, that taking an initial test can serve as a learning event that produces downstream effects on later performance (e.g., Karpicke & Blunt, 2011b), although some evidence also suggests that completing an initial higher-order test may not necessarily benefit subsequent factual test performance, and vice versa (Agarwal, 2019).

**Table 1**  
*Mean Questionnaire Scores and Metacognitive Judgments in Experiment 1*

Variable	Copy		Concept-map		Concept-error	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
English proficiency (GRE)	3.55	1.47	3.93	1.69	4.20	1.84
Text understandability	5.45	1.06	5.15	1.51	5.38	1.19
Prior knowledge of text content	2.80	1.49	2.67	1.53	3.25	1.72
Judgment of learning	60.50	17.09	55.50	16.48	58.00	18.00
Method effectiveness	4.35	1.42	4.85	1.21	4.33	1.31

Note. *N* = 120. GRE = Graduate Record Examinations.

points since such convergences form mountains instead. Likewise, for the “food allergies” text, responses that simply narrated the biological processes involved in food allergies without clearly explaining how these processes could have led to the symptoms that Braxton Ong displayed, as well as why his allergic reaction manifested specifically after he had eaten buckwheat noodles and oranges, would not be successful in earning points.

For the free-recall test, participants’ responses were scored based on the total number of idea units from the study text that they correctly recalled. The maximum possible score was 40. Two raters independently scored both the application and free-recall tests for 30 of the 120 scripts. Interrater reliability was high, intraclass correlation (ICC) = .996, 95% CI [.992, .998], 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.

In addition, as a proxy for the extent that learners in the concept-mapping condition had engaged in generative processing (Fiorella & Mayer, 2016), we assessed their concept maps for structural complexity based on the scoring system proposed in Novak and Gowin’s (1984) first introduction of concept maps. This system has been considered a relatively objective measure that quantitatively analyzes concept maps for their level of knowledge structure complexity (for a review, see de Ries et al., 2021), with point values assigned to each structural characteristic based on Novak and Gowin’s (1984) assessment of its relative importance: (a) *propositions* (the number of meaningful, valid links between concepts; 1 point each), (b) *hierarchy* (the number of hierarchical levels in the concept map ranging from more to less general concepts; 5 points for each valid level of the hierarchy), and (c) *crosslinks* (the number of meaningful, valid connections across distinct segments of the concept hierarchy; 10 points each).<sup>3</sup> Summing the number of points across all three scoring criteria further yielded a composite structural complexity score for each concept map. Collectively, these criteria provided a measure of generative processing that has been viewed as essential for meaningful learning (Fiorella & Mayer, 2016; Mayer, 1996). For instance, when constructing propositional linkages and hierarchy levels in a concept map, learners must select and organize relevant concepts in the study text while differentiating among more general concepts versus their subordinate ones, whereas the creation of novel crosslinks between sets of concepts that might otherwise be viewed as independent suggests learners’ integrative reconciliation of concepts (Novak & Gowin, 1984).

### Preliminary Analyses

**Prior Knowledge, English Proficiency, and Text Understandability.** We ascertained that learners possessed minimal familiarity with the study texts. Indeed, participants reported low prior knowledge of the study material overall, with no significant differences across the copy ( $M = 2.80$ ,  $SD = 1.49$ ), concept-map ( $M = 2.67$ ,  $SD = 1.53$ ), and concept-error ( $M = 3.25$ ,  $SD = 1.72$ ) conditions,  $F(2, 117) = 1.46$ ,  $p = .24$ ,  $\eta_p^2 = .02$ . In addition, the three learning groups did not differ in their English language proficiency,  $F(2, 117) = 1.52$ ,  $p = .22$ ,  $\eta_p^2 = .03$ , and ratings of text understandability,  $F(2, 117) = 0.61$ ,  $p = .55$ ,  $\eta_p^2 = .01$ . Table 1 shows the means and standard deviations.

**Structural Complexity of Concept Maps.** Analyses of the structural complexity of learners’ concept maps revealed that they contained an average of 30.85 propositions ( $SD = 8.06$ ), 4.63 hierarchy levels ( $SD = 1.33$ ), and 0.50 crosslinks ( $SD = 0.68$ ). The composite structural complexity scores of learners’ concept maps ( $M = 58.97$ ,  $SD = 14.10$ ) positively and significantly correlated with their performance on both the recall test,  $r(38) = .53$ ,  $p < .001$ , and application test,  $r(38) = .32$ ,  $p = .041$ . Exploring the relations between each structural characteristic and learners’ test performance, we found that recall test performance in the concept-mapping condition significantly correlated with the number of propositions and hierarchy levels in learners’ maps,  $r(38) = .37$  and  $.45$ ,  $p = .02$  and  $.003$ , respectively, but not the number of crosslinks,  $r(38) = .23$ ,  $p = .16$ . Conversely, learners’ application test performance significantly correlated with the number of crosslinks in their maps,  $r(38) = .43$ ,  $p = .005$ , but not the number of propositions or hierarchy levels,  $r(38) = .06$  and  $.17$ ,  $p = .72$  and  $.29$ , respectively. These correlational patterns suggest that the recall test was more closely associated with organization processes in constructing propositional linkages and hierarchy levels in one’s concept map, whereas the application test was more closely related to higher-order integration processes in creating novel crosslinks among concepts.

### Main Analyses

**Application Test Performance.** To analyze participants’ application test performance across learning conditions, we conducted a

<sup>3</sup> Although Novak and Gowin (1984) proposed that the number of examples (i.e., specific events or objects that are valid instances of the concepts; scored as 1 point each) in learners’ concept maps can be included as a fourth scoring criterion, they also considered such examples as “less indicative of meaningful learning” (p. 107). Hence, we focused on the propositions, hierarchy, and crosslinks in learners’ concept maps in view of their particular pertinence for meaningful learning processes.

two-way between-subjects analysis of variance (ANOVA) with learning method and study text as the independent variables. As predicted, there was a significant main effect of learning method,  $F(2, 114) = 7.17, p = .001, \eta_p^2 = .11$ . Participants in the concept-error condition ( $M = 9.73, SD = 4.10$ ) outperformed those in the copy ( $M = 6.48, SD = 4.36$ ) and concept-map ( $M = 6.95, SD = 3.83$ ) conditions in applying the material they had learned from the study text to analyze the news event,  $p = .001$  and  $.003, d = 0.77$  and  $0.70$ , respectively. Participants in the copy and concept-map conditions did not differ in their application test performance,  $p = .61, d = 0.11$ . Mean application performance as a function of learning condition appears graphically in Figure 1A. There was neither a significant main effect of study text,  $F(1, 114) = 0.28, p = .60, \eta_p^2 = .002$ , nor a significant Learning Method  $\times$  Study Text interaction,  $F(2, 114) = 0.13, p = .88, \eta_p^2 = .002$ , indicating that the advantage of deliberate erring over copying and concept mapping persisted across both the “volcanoes” and “food allergies” study texts.

**Recall Test Performance.** Likewise, participants’ recall test performance significantly differed across the three learning conditions,  $F(2, 114) = 6.64, p = .002, \eta_p^2 = .10$ . Deliberate erring ( $M = 18.58, SD = 6.48$ ) led to better knowledge retention than the copy ( $M = 13.13, SD = 7.96$ ) and concept-map ( $M = 14.95, SD = 7.01$ ) conditions,  $p = .001$  and  $.02, d = 0.75$  and  $0.54$ , respectively. The copy and concept-map conditions did not significantly differ,  $p = .23, d = 0.24$ . Mean recall performance as a function of learning condition is presented in Figure 1B. There was a significant main effect of study text whereby participants recalled more idea units overall from the “volcanoes” ( $M = 17.98, SD = 7.47$ ) than “food allergies” ( $M = 13.12, SD = 6.69$ ) text,  $F(1, 114) = 15.33, p < .001, \eta_p^2 = .12$ . Importantly, however, there was no Learning Method  $\times$  Study Text interaction,  $F(2, 114) = 0.32, p = .73, \eta_p^2 = .006$ , indicating that the recall advantage of deliberate erring over the errorless conditions held reliably across both study texts.

**Metacognitive Judgments.** In contrast to their test performance, participants’ JOLs and perceived effectiveness ratings of the learning methods (solicited before and after the tests, respectively) did not differ across learning conditions,  $F(2, 117) = 0.85, p = .43, \eta_p^2 = .01$ , and  $F(2, 117) = 2.02, p = .14, \eta_p^2 = .03$ , respectively. Table 1 shows the means and standard deviations of participants’ metacognitive judgments. This suggests that participants had little metacognitive knowledge that deliberate erring had been helpful for their learning at the end of the studying phase and even after its benefits on the tests were obtained. Figure 1C displays participants’ JOLs across learning conditions.

## Discussion

Relative to both errorless learning techniques of copying with underlining and elaborative studying with concept mapping, deliberately committing and correcting errors produced superior recall performance and, importantly, also enhanced students’ meaningful learning in applying their knowledge to analyze a novel news event. Taken together, our findings in Experiment 1 provide evidence for the derring effect in both knowledge retention and higher-order application, although learners’ metacognitive judgments revealed that they were largely not cognizant of these benefits.

Interestingly, concept mapping did not offer a learning advantage over copying with underlining (e.g., Karpicke & Blunt, 2011b). Although concept mapping has been proposed to be a relatively

intuitive technique that can be learned with minimal training within a duration as brief as 10–20 min (e.g., Hay et al., 2008; for a discussion, see Karpicke & Blunt, 2011a), it is possible that its benefits may be more robust when learners are provided with more extensive training to construct higher-quality maps (Fiorella & Mayer, 2016). For instance, in the case of Chularut and DeBacker’s (2004) concept-mapping group that showed greater learning gains than an individual-study-plus-discussion group, learners underwent a 60-min initial session in which they received instruction and practice on concept mapping, and were further given feedback on their concept maps in four subsequent 60-min sessions. However, it should be noted that evidence from randomized controlled experiments for the need for extensive training on concept mapping currently appears to be lacking; improvements in students’ concept maps over time do not necessarily mean that training is a prerequisite for concept mapping to be effective (Karpicke & Blunt, 2011a). Moreover, from a practical standpoint, tedious and time-consuming training or map-building processes may lead students to lose interest in this technique (Fiorella & Mayer, 2016). Crucially, this simultaneously implies that deliberate erring is a relatively efficient learning strategy, given that students in our study did not require extensive training to benefit from it on both measures of knowledge retention and application.

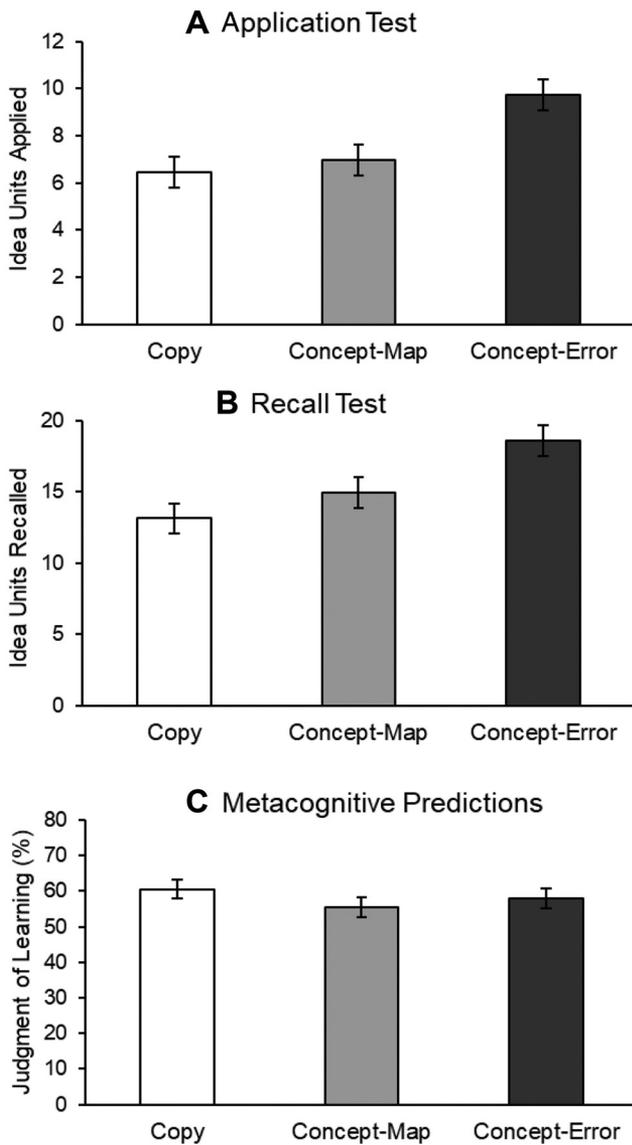
## Experiment 2

Experiment 2 aimed to replicate Experiment 1’s findings and to further explicate the derring effect. Specifically, we sought to show that neither generation nor elaboration explained the learning benefits of deliberate erring. Cognitive psychology research has demonstrated that information is often remembered better when it has been actively generated rather than passively read (i.e., the *generation effect*; Bertsch et al., 2007; Jacoby, 1978; Slamecka & Graf, 1978) and when it has been elaborated on through the addition of meaning (Craik & Tulving, 1975; Levin, 1988; Pressley et al., 1987, 1992). In Experiment 1, deliberate erring involved generating incorrect elaborations of what the concepts in the text were not, whereas both copying with underlining and concept mapping did not induce such processes. Accordingly, it is possible that the superior performance in Experiment 1’s concept-error condition arose from a generation or an elaboration benefit, rather than from deliberate erring per se.

To foreclose these accounts, Experiment 2 employed an even more rigorous errorless *concept-synonym* control that was experimentally identical to the concept-error condition except that learners actively generated conceptual synonyms—correct elaborations of what else the concepts in the text meant. As a proxy for the elaboration that learners engaged in, we assessed the text length of the study responses that they generated (e.g., Daley & Rawson, 2019) across both the concept-synonym and concept-error conditions.

From an applied perspective, the concept-synonym condition resembles the learning technique of paraphrasing during note taking, which is a popular study strategy among students with a meta-analyzed frequency of use of 30% (Miyatsu et al., 2018). Drawing on one’s prior knowledge to paraphrase to-be-learned material in one’s own words encourages greater depth of elaborative processing that is crucial for effective note taking, and is a generative technique that has been widely viewed to produce higher-quality notes than verbatim copying or transcribing (Jansen et al., 2017; Kiewra, 1985, 1989; Miyatsu et al., 2018; Wittrock, 1989). Indeed, paraphrasing

**Figure 1**  
Results of Experiment 1



*Note.* Panels A and B show the mean application test and recall test scores, respectively. The maximum possible score for the recall test was 40. Panel C shows the proportion of information that participants predicted they would remember on the final test (i.e., their metacognitive judgments of learning [JOLs]). Error bars indicate standard errors.

textual information has been shown to enhance retention, relative to reading only, reading for recognition of particular types of statements in the text, or selecting and listing keywords (Glover et al., 1981). Besides boosting retention, the learning advantages of paraphrasing extend to test questions that require integrating a combination of facts from the studied text, for which paraphrasing has been found to produce better performance than reading only, searching for capital letters, or extracting and copying main ideas verbatim, while being as effective as writing a summary (Bretzing & Kulhavy, 1979).

Accordingly, by directly pitting the concept-error method against the concept-synonym method, Experiment 2 enabled a head-to-head empirical comparison of generating incorrect versus correct elaborations of the study material. If deliberately committing and correcting errors prevails even over generating alternative correct answers, then this result would provide further compelling evidence for the derring effect, while suggesting that it cannot be merely attributed to a generation or an elaboration benefit alone. That is, simply generating any novel (correct) elaboration is insufficient—rather, one must specifically produce an *incorrect* response to reap meaningful learning gains.

Concurrently, to examine the relative effectiveness of both learning methods for each individual learner, Experiment 2 used a within-subjects design. This enabled us to determine the number of students who benefited more from deliberate erring than generating conceptual synonyms, versus the number of students who showed the opposite pattern, versus the number of students who showed no difference across both learning methods. In addition, this design allowed us to examine whether learners would exhibit greater metacognitive awareness of the benefits of deliberate erring after having the opportunity to personally experience the effects of both the concept-synonym and concept-error methods for their test performance.

## Method

### Participants

The participants were 40 undergraduate students (32 were female) between the ages of 19 and 28 ( $M = 20.90$ ,  $SD = 1.89$ ) from the National University of Singapore. A power analysis (G\*Power; Faul et al., 2007) indicated that at least 34 participants would afford 80% power at  $\alpha = .05$  to detect a moderate within-subjects effect ( $d = 0.50$ ) similar to the most conservative effect size that we observed in Experiment 1 between the concept-error and concept-map conditions on the recall test.

### Design

The single within-subjects factor of interest was learning method: *concept-synonym* (errorless generation and elaboration) versus *concept-error* (deliberate erring). As in Experiment 1, the dependent variables were participants' application and recall test scores.

### Materials

The materials were identical to those in Experiment 1, with one exception. Specifically, two practice texts were used, each comprising of 29 words (three sentences). Adapted from Karpicke and Blunt (2011b), the practice texts were on “muscle tissue” (as in Experiment 1) and “the human ear,” both of which were not related to the critical study texts on “volcanoes” and “food allergies.”

### Procedure

As in Experiment 1, prior to attending the experiment, participants completed an online questionnaire that assessed their English language proficiency. Upon arriving at the laboratory, participants underwent three experimental phases: practice, studying, and test.

**Practice Phase.** Participants were first introduced to the concept-synonym and concept-error methods, and practiced using each method to study the practice texts. In the concept-synonym

condition, participants elaborated on the text by writing down each sentence such that it contained a conceptual synonym (i.e., an alternative word or phrase with the same meaning as the actual concept), before underlining this synonym and writing down the actual concept exactly as it was presented in the text. As an illustration, for the training example “Bats are mammals that fly,” a sample response was “Bats are warm-blooded animals with fur (mammals) that fly.”

The concept-error condition was the same as that in Experiment 1. Participants deliberately erred by writing down each sentence in the text such that it contained a plausible conceptual error, before striking out this error and correcting it by writing down the actual concept exactly as it was presented in the text. For instance, a sample response for the training example was “Bats are ~~birds~~ (mammals) that fly”.

**Studying and Test Phases.** After completing the practice trials, participants began the studying phase. Taking the statement “Magmas that are low in silica tend to be very fluid” from the “volcanoes” text as an example, a sample response in the concept-synonym condition was “Magmas that are low in silica tend to flow very easily like liquid (be very fluid),” whereas a sample response in the concept-error condition was “Magmas that are low in silica tend to be very ~~viscous~~ (fluid).” The order in which participants went through the concept-synonym and concept-error conditions was counterbalanced, as was the pairing of learning conditions and study texts. In both conditions, participants were first given 1.5 min to read the given text before using the specified learning method to study the text for 25 min. Thus, the total studying duration was exactly matched across both conditions. At the end of the 25-min period, participants completed the same three-item questionnaire used in Experiment 1—they made a JOL, rated the understandability of the study text, and indicated their prior knowledge of the text content. The same procedure was then repeated for the second learning condition.

After studying both texts, participants were allowed to take a brief, self-paced break before undergoing the test phase, which was identical to that in Experiment 1 except that participants now completed four tests—an application test and free-recall test for each of the two study texts. The four tests were blocked by learning condition and were administered in the same order that participants had studied both texts, with the application test preceding the free-recall test for each condition. Finally, after completing all tests, participants rated the effectiveness of each learning method on a 7-point scale (1 = *not at all*; 7 = *extremely*).

## Results

### Scoring

The application and free-recall tests were scored in the same manner as in Experiment 1. Two raters independently scored 12 of the 40 scripts. Interrater reliability was high, ICC = .994, 95% CI [.988, .997], based on a two-way random-effects model. Discrepancies were reviewed and resolved through discussion to yield 100% agreement. Given the high interrater reliability, the remaining scripts were scored by one rater.

### Preliminary Analyses

**Prior Knowledge, English Proficiency, and Text Understandability.** Participants’ mean GRE score was 3.38

( $SD = 1.96$ ). As in Experiment 1, participants reported low prior knowledge of the study texts on overall, with no significant difference between the concept-synonym ( $M = 2.45$ ,  $SD = 1.52$ ) and concept-error ( $M = 3.03$ ,  $SD = 1.56$ ) conditions,  $t(39) = -1.71$ ,  $p = .10$ , 95% CI [-1.25, 0.10]. In addition, participants’ ratings of the study texts’ understandability did not significantly differ across the concept-synonym ( $M = 4.65$ ,  $SD = 1.46$ ) and concept-error ( $M = 5.00$ ,  $SD = 1.28$ ) conditions,  $t(39) = -1.36$ ,  $p = .18$ , 95% CI [-0.87, 0.17].

**Length of Elaborations.** As a proxy for the elaboration that participants engaged in, we scored and analyzed the word count of their study responses across both learning conditions (e.g., Daley & Rawson, 2019). A paired-samples  $t$  test revealed that participants wrote significantly longer elaborations when generating conceptual synonyms ( $M = 346.45$ ,  $SD = 20.04$ ) than conceptual errors ( $M = 337.95$ ,  $SD = 15.14$ ),  $t(39) = 2.65$ ,  $p = .01$ ,  $d = 0.42$ , 95% CI [2.01, 14.99].

### Main Analyses

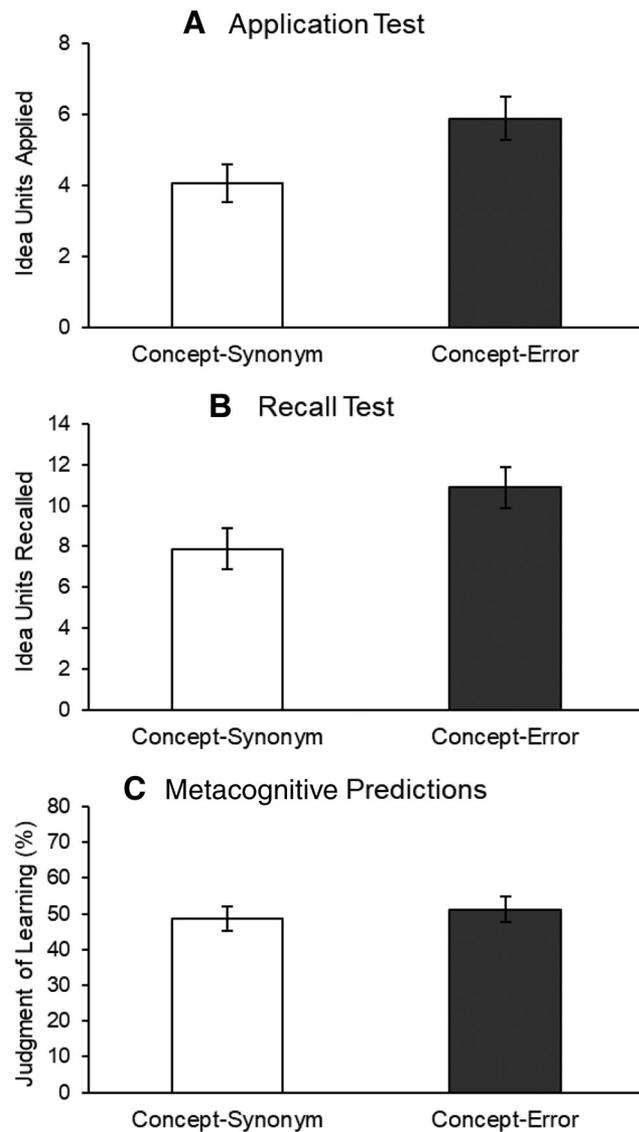
**Application Test Performance.** As predicted, the concept-error method produced superior application test performance than the concept-synonym method,  $t(39) = -2.89$ ,  $p = .006$ ,  $d = 0.46$ , 95% CI [-3.10, -0.55]. Relative to elaborating on the text by generating correct conceptual synonyms ( $M = 4.05$ ,  $SD = 3.28$ ), deliberate erring ( $M = 5.88$ ,  $SD = 3.91$ ) enhanced learners’ performance in applying the material to analyze a novel news event (Figure 2A).

**Recall Test Performance.** Participants recalled significantly more idea units from the study text in the concept-error ( $M = 10.88$ ,  $SD = 6.47$ ) condition than the concept-synonym ( $M = 7.88$ ,  $SD = 6.42$ ) condition,  $t(39) = -3.24$ ,  $p = .002$ ,  $d = 0.51$ , 95% CI [-4.87, -1.13]. Thus, deliberate erring yielded a recall advantage over errorless generation of conceptual synonyms (Figure 2B).

**Metacognitive Judgments.** Yet, in contrast to their actual test performance, participants’ JOLs revealed that they inaccurately predicted no difference in their learning across the concept-synonym ( $M = 48.50$ ,  $SD = 21.79$ ) and concept-error ( $M = 51.25$ ,  $SD = 21.86$ ) conditions,  $t(39) = -0.89$ ,  $p = .38$ , 95% CI [-9.02, 3.52]. Figure 2C shows participants’ JOLs across learning conditions. Even after experiencing the benefits of deliberate erring for their test performance, learners rated both the concept-synonym ( $M = 4.25$ ,  $SD = 1.55$ ) and concept-error ( $M = 4.25$ ,  $SD = 1.41$ ) methods as similarly effective,  $t(39) = 0.00$ ,  $p = 1.00$ , 95% CI [-0.62, 0.62].

We further examined the predicted and actual effectiveness of the concept-error versus concept-synonym methods for each individual learner. Table 2 shows the number of participants who actually performed better after deliberate erring than generating correct conceptual synonyms, the number who showed the opposite pattern, and the number who performed equivalently across both learning conditions. For each of these three outcomes, Table 2 also displays the number of participants who made the corresponding pretest metacognitive predictions (JOLs) and posttest metacognitive judgments (effectiveness ratings). Overall, 26 out of 40 learners (65%) benefited more from the concept-error than concept-synonym method on the application and recall tests. Yet, 22 out of 40 learners (55%) predicted that the concept-synonym method would be just as

**Figure 2**  
Results of Experiment 2



*Note.* Panels A and B show the mean application test and recall test scores, respectively. The maximum possible score for the recall test was 40. Panel C shows the proportion of information that participants predicted they would remember on the final test (i.e., their metacognitive judgments of learning [JOLs]). Error bars indicate standard errors.

effective or even more effective than the concept-error method, as indicated by their JOLs. This metacognitive illusion persisted even after the tests, whereby 25 out of 40 learners (63%) rated the concept-synonym method as just as effective or even more effective than the concept-error method for their test performance. In sum, students tended to fail to predict or realize that deliberate erring would benefit their learning more than generating conceptual synonyms, when it actually did.

## Discussion

The derring effect reliably emerged in Experiment 2. Extending our findings in Experiment 1, deliberately committing and correcting errors produced superior retention and higher-order application performance even over generating conceptual synonyms. In view that both learning techniques similarly involved generating elaborations, and that learners in fact wrote significantly longer elaborations when producing conceptual synonyms than errors, this result suggests that the derring effect cannot be parsimoniously explained by generation or elaboration processes alone. Rather, the learning benefits observed are specific to having first deliberately produced an incorrect, but not a correct, response.

Yet, replicating Experiment 1's findings, there was a disconnect between learners' metacognitive judgments and actual test performance. Despite having personally experienced both learning methods for themselves in Experiment 2, learners remained largely oblivious to the benefits of deliberate erring. Instead, participants tended to inaccurately predict that the concept-synonym method would be just as effective or even more effective than the concept-error method, and continued to believe that this was the case even after their actual test performance revealed otherwise.

## General Discussion

Our findings across two experiments challenge error-avoidant learning while lending support to the counterintuitive approach of actively promoting errors by deliberately committing and correcting them. In Experiment 1, deliberate erring produced superior memory for complex, educationally relevant material, as compared to errorless copying with underlining and elaborative studying with concept mapping. Crucially, deliberate erring did not merely induce rote learning, but promoted meaningful learning in enhancing students' higher-order application of their knowledge to analyze a novel news event. In view that both copying with underlining and concept mapping fell short despite being active or even generative learning methods in the latter's case (Fiorella & Mayer, 2016), our data suggest that active learning alone is insufficient to explain the benefits of deliberate erring.

Moreover, in Experiment 2, deliberate erring reliably outperformed generating alternative correct answers—a learning technique that encourages semantic elaboration and greater depth of processing when learners integrate the to-be-learned material with their prior knowledge during paraphrasing (e.g., Bretzing & Kulhavy, 1979; Jansen et al., 2017). That students' knowledge retention and application benefited more from deliberately generating errors than correct conceptual synonyms—even when the latter induced longer elaborations—suggests that the derring effect is not attributable to generation or elaboration alone. Instead, the learning advantages accrued are specific to having first deliberately produced an error, rather than any other novel (correct) response.

## Theoretical Explanations for the Derring Effect

Our data speak against a number of theoretical explanations for the benefits of deliberate erring. For one, unlike incorrect guessing (Potts et al., 2019; Potts & Shanks, 2014), the derring effect cannot be explained by learners' aroused curiosity to learn the correct

**Table 2**  
*Number of Participants Showing Different Patterns of Metacognitive Ratings and Actual Test Performance (Experiment 2)*

Metacognitive ratings vs. actual performance	Performance outcome		
	Error > synonym	Error = synonym	Error < synonym
Metacognitive ratings			
Pretest predictions (JOLs)	18	7	15
Posttest judgments (effectiveness ratings)	15	10	15
Actual performance			
Application test	26	2	12
Recall test	26	3	11

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

responses, since they already know them. Likewise, neither does this effect stem from learners' surprise at having apparently responded wrongly, in contrast to errors committed with high confidence (Butterfield & Metcalfe, 2001). Yet another conjecture is that generating deliberate errors may create alternative cues or routes that aid retrieval of the correct responses (e.g., Cyr & Anderson, 2015; Kornell et al., 2009). However, we found that generating additional retrieval cues in the form of conceptual synonyms produced poorer learning than deliberate erring (Experiment 2), even though the former arguably paved more appropriate routes to the correct responses. For instance, whereas the conceptual synonym "warm-blooded animals with fur" may reasonably lead participants to retrieve the semantically related correct response that bats are "mammals," it is not immediately obvious how the conceptual error "birds" may more effectively cue "mammals" instead of introducing interference.

Rather, a viable interpretation is that encoding of the correct response is enhanced after deliberate erring (Hays et al., 2013; Potts et al., 2019). This account aligns with reconsolidation theory, which proposes that existing memory traces enter a labile and malleable state when they are reactivated, during which a transient window of opportunity opens for them to be modified and reconsolidated with new learning events (Nader & Hardt, 2009; Nader et al., 2000; see also Metcalfe, 2017). Likewise, deliberate error commission may introduce an opportunity for heightened processing of the target response during error correction. For instance, after having intentionally generated erroneous answers, learners' attention may be drawn to the target response during correction, thereby fostering an episodically memorable event that catalyzes learning (Metcalfe & Huelser, 2020). Conversely, errorless learning via copying with underlining, concept mapping, or generating conceptual synonyms may not offer such enrichment because less distinctive memory traces may be produced when the target response is preceded by similarly correct information, thus failing to provide opportunities to reap its subsequent potentiation.

Alternatively, deliberately committing and correcting errors may strengthen retrieval routes to the correct concepts by inducing complementary or additional kinds of processing that are not invited by the study material but that are relevant and useful for learning it (for a discussion of material-appropriate processing, see McDaniel & Einstein, 1989). For instance, in the present learning context, considering what something is *not* (i.e., generating deliberate conceptual errors) may, paradoxically, enhance learning of

the correct concept more than considering what *else* it is (i.e., generating conceptual synonyms). In turn, exploring such incorrect responses during deliberate erring may more effectively cull those inappropriate responses, thus facilitating future test performance (Kornell et al., 2009; see also Gartmeier et al., 2008). Examining these potential mechanisms in future research would illuminate the theoretical underpinnings of the derring effect.

### Metacognitive Illusions of the Derring Effect

Another key finding in our study was that learners were largely unaware that deliberate erring was helpful for their test performance. This metacognitive illusion echoes those in other studies on "naturalistic" errors (e.g., Huelser & Metcalfe, 2012; Potts & Shanks, 2014; Yang et al., 2017) and is often observed when learners make their JOLs immediately after study. Indeed, JOLs tend to be more accurate when they are solicited following a short delay after study, such that learners' metamemory monitoring is improved and their judgments are better calibrated to be more closely attuned to their actual learning (i.e., the *delayed-JOL effect*; Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991; for a meta-analytic review, see Rhodes & Tauber, 2011). Thus, it was particularly surprising that learners' inaccurate metacognitive judgments persisted even after they were given the opportunity to personally experience the effects of both errorful and errorless learning for their test performance, which in fact provided diagnostic information to inform their judgments (Experiment 2). Despite benefiting more from deliberate erring, most participants continued to believe that this strategy was as effective as, if not actually less so than, errorless learning just moments after completing the application and recall tests.

One source of learners' inaccurate metacognitive judgments may be their intrinsic naïve beliefs about the relative efficacy of errorful versus errorless learning (Huelser & Metcalfe, 2012). For instance, learners may assume that generating errors interferes with their memory for the correct responses on a later test, which may partially contribute to a bias against believing that errors are beneficial (Yang et al., 2017; cf. Pan et al., 2020).

Another possibility stems from learners' reliance on processing fluency—or an "easily learned, easily remembered" heuristic—when making their metacognitive judgments (Begg et al., 1989; Koriat, 2008; Koriat & Ma'ayan, 2005). Specifically, learners tend to judge information as being better learned when they experience greater subjective ease of processing during study (e.g., Koriat, 2008; Rawson &

Dunlosky, 2002). Typically, this heuristic results in errorful generation being mistakenly judged as inferior to errorless learning techniques such as reading (e.g., Huelser & Metcalfe, 2012; Potts & Shanks, 2014). Presumably, the production of incorrect responses may induce greater processing disfluency than merely reading correct answers, such that learners underestimate the efficacy of erring. Across both experiments, though, learners' metacognitive predictions did not significantly differ across the deliberate erring versus errorless learning conditions. Thus, it is possible that learners experienced similar degrees of processing fluency across the various study methods, particularly when the concept-synonym method was closely matched against the concept-error method in Experiment 2, such that they failed to appreciate the actual benefits of deliberate erring for their test performance.

### Educational Implications

The present results offer a new perspective on how errors can be strategically repositioned in educational contexts through intentional design, while illustrating how greater learning success can paradoxically be achieved through first seeking out and embracing failure. One practical implication of this research is that deliberately incorporating errors in learning can be more potent than avoiding them entirely. Importantly, unlike chance and induced errors, learners can systematically and independently apply deliberate errors even during their self-regulated study to boost both knowledge retention and higher-order application performance. For instance, rather than considering only correct responses when studying to-be-learned material, our data suggest that learners would gain more from deliberately formulating conceptually plausible but erroneous responses and then correcting them.

One foreseeable challenge, though, lies in our finding that students often fail to appreciate that deliberate erring is helpful for their learning, even after benefiting from this strategy. Such metacognitive illusions are often remarkably difficult to dispel due to, for instance, erroneous a priori theories and naïve intuitions that learners hold about what they believe works best for them (Yan et al., 2016). Lacking metacognitive awareness of the advantages of errors may contribute to people's aversion against making them or being seen making them in school or at work (Frese & Keith, 2015), and lead students to select suboptimal study strategies (Metcalfe & Finn, 2008; Thiede et al., 2003). For instance, a recent survey of 1,052 undergraduate students across three large public universities in North America revealed that although the vast majority (90%) of students reported sometimes or often spending time to study or analyze the errors that they made, 81% of students believed that it was moderately or very important for them to avoid making errors when learning an academic subject (Pan et al., 2020). These findings suggest that students are often disinclined toward deliberately generating errors during learning, even while they may acknowledge the pedagogical value of errors and make efforts to learn from them when they do occur. Thus, overcoming this hurdle of aversion toward error commission is a vital step toward building a "translational educational science", whereby recommendations based on research-informed principles are translated into action by educational systems (Roediger, 2013).

In particular, some evidence suggests that learners' metacognitive illusions may be partially dispelled when they are informed prior to study about the benefit of generating errors (Yang et al.,

2017). Indeed, receiving targeted instruction on applied learning and memory topics has been found to improve students' metacognitive judgments and endorsement of learning strategies that are actually effective for academic success (McCabe, 2011). As such, one promising pathway forward may be for students to be introduced to the learning advantages of deliberate erring so that they may make more informed choices on applying this technique during their study routines.

### Future Directions

To formulate more precise recommendations for educational practice, validating the derring effect in authentic classrooms and representative learning contexts presents a meaningful endeavor for future research. For instance, whereas the learning material in the present research comprised relatively brief scientific expository texts, it would be beneficial for future work to replicate the derring effect using lengthier texts and across other kinds of complex educational materials, while exploring how the implementation of deliberate erring can be optimized. Indeed, although the studying phase in the current experiments spanned 25 min to accommodate the relatively slower process of longhand writing that all the learning methods similarly involved, it is plausible that deliberate erring can be more efficiently applied in learners' study routines while still preserving its effectiveness.

In addition, whereas the present research has showcased the benefits of committing and correcting deliberate conceptual errors for knowledge retention and higher-order application, it will be important for future work to rigorously assess the extent to which these benefits persist over longer retention intervals and generalize across a wider range of error types and criterion tasks. For instance, besides conceptual errors, procedural errors are relevant in domains such as software training, motor skill acquisition, and problem solving. Notably, relative to errorless learning, inducing inadvertent errors when learners initially attempt and fail to solve challenging problems has been found to facilitate spontaneous transfer, whereby learners are more likely to successfully apply the source solution to solve novel analogous problems that are structurally similar but with nonidentical surface features (Gick & McGarry, 1992). By extension, deliberate procedural errors when developing incorrect solutions may yield learning gains, particularly when these errors are analogous to those that would typically be generated to the novel target problem at hand, thereby facilitating noticing or retrieval of the source problem for successful knowledge application and transfer. Testing the derring effect with diverse error types will shed light on how deliberate errors can be effectively implemented in various educational tasks.

Another open question pertains to the extent that the efficacy of deliberate erring is moderated by student characteristics such as one's level of expertise or prior knowledge. For instance, whereas learners with less domain knowledge have been found to benefit from studying only correct solutions via worked examples, these appear to be redundant for learners with greater domain knowledge, who benefit more from studying both incorrect and correct solutions (Große & Renkl, 2007). In view that students in our study benefited from deliberate erring even when they possessed relatively low self-reported prior knowledge of the studied material, the derring effect may well be amplified for students with high prior knowledge who may, in turn, be better equipped to relate their existing knowledge to the studied material

when formulating plausible incorrect responses. At the same time, it is possible that learners with no prior knowledge may gain less from deliberate erring particularly when the to-be-learned material is highly complex, in view that such interactions between the nature of the material and learners' expertise may induce cognitive overload that interferes with learning (Sweller et al., 1998, 2019). Investigating the generality of the derring effect across student characteristics at different stages of learning will further illuminate the contexts in which deliberate errors are more (or less) likely to be helpful.

## Conclusion

Errors are natural and inevitable events in human life, as with how chips and cracks on objects convey their history of being used. Yet, embracing and learning from our errors is often fraught with challenges in educational contexts, particularly when errors have traditionally been viewed as aversive events to be avoided or, at best, serendipitous accidents. If such setbacks can be circumvented, however, the comeback may be compelling. In reconsidering current approaches to errors, we have demonstrated that deliberate erring, a novel and counterintuitive strategy, effectively promotes errors as intentional events to enhance meaningful learning. In *kintsugi*, the beauty of an artifact lies precisely in its being broken—likewise, in ~~rejeeting~~ deliberately embracing imperfection, learning may ultimately be perfected.

## 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>
- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. <https://doi.org/10.1037/a0021017>
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214. <https://doi.org/10.3102/00346543070002181>
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. Holt, Rinehart and Winston.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*(5), 610–632. [https://doi.org/10.1016/0749-596X\(89\)90016-8](https://doi.org/10.1016/0749-596X(89)90016-8)
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201–210. <https://doi.org/10.3758/BF03193441>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Bloom, B. (1956). *Taxonomy of educational objectives: The classification of educational goals*. McKay.
- Bretzing, B. H., & Kulhavy, R. W. (1979). Notetaking and depth of processing. *Contemporary Educational Psychology, 4*(2), 145–153. [https://doi.org/10.1016/0361-476X\(79\)90069-9](https://doi.org/10.1016/0361-476X(79)90069-9)
- Brown, A. L., & Smiley, S. S. (1978). The development of strategies for studying texts. *Child Development, 49*(4), 1076–1088. <https://doi.org/10.2307/1128747>
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21–32.
- 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>
- Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*(3), 248–263. <https://doi.org/10.1016/j.cedpsych.2003.09.001>
- 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](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>
- de Ries, K. E., Schaap, H., van Loon, A.-M. M. J. A. P., Kral, M. M. H., & Meijer, P. C. (2021). A literature review of open-ended concept maps as a research instrument to study knowledge and learning. *Quality & Quantity: International Journal of Methodology*. Advance online publication. <https://doi.org/10.1007/s11135-021-01113-x>
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*(4), 374–380. <https://doi.org/10.3758/BF03210921>
- 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>
- 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>
- Fass, W., & Schumacher, G. M. (1978). Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of Educational Psychology, 70*(5), 803–807. <https://doi.org/10.1037/0022-0663.70.5.803>
- 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>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111*(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Frese, M., & Altmann, A. (1989). The treatment of errors in learning and training. In L. Bainbridge & S. A. R. Quintanilla (Eds.), *Developing skills with information technology* (pp. 65–86). Wiley.

- Frese, M., & Keith, N. (2015). Action errors, error management, and learning in organizations. *Annual Review of Psychology*, *66*, 661–687. <https://doi.org/10.1146/annurev-psych-010814-015205>
- Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2008). Negative knowledge: Understanding professional learning and expertise. *Vocations and Learning*, *1*(2), 87–103. <https://doi.org/10.1007/s12186-008-9006-1>
- Gick, M. L., & McGarry, S. J. (1992). Learning from mistakes: Inducing analogous solution failures to a source problem produces later successes in analogical transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 623–639. <https://doi.org/10.1037/0278-7393.18.3.623>
- Glover, J. A., Plake, B. S., Roberts, B., Zimmer, J. W., & Palmere, M. (1981). Distinctiveness of encoding: The effects of paraphrasing and drawing inferences on memory from prose. *Journal of Educational Psychology*, *73*(5), 736–744. <https://doi.org/10.1037/0022-0663.73.5.736>
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2019). The effects of comprehension-test expectancies on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 1066–1092. <https://doi.org/10.1037/xlm0000634>
- 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>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hay, D., Kinchin, I., & Lygo-Baker, S. (2008). Making learning visible: The role of concept mapping in higher education. *Studies in Higher Education*, *33*(3), 295–311. <https://doi.org/10.1080/03075070802049251>
- 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](https://doi.org/10.1016/S0022-5371(78)90393-6)
- Jansen, R. S., Lakens, D., & IJsselstein, W. A. (2017). An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review*, *22*, 223–233. <https://doi.org/10.1016/j.edurev.2017.10.001>
- Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*(1), 48–59. <https://doi.org/10.1037/a0021977>
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, *26*(3), 379–424. <https://doi.org/10.1080/07370000802212669>
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299. <https://doi.org/10.1080/00461520.2016.1155457>
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. <https://doi.org/10.1080/10508406.2011.591717>
- Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317–326. <https://doi.org/10.1007/s10648-015-9309-3>
- Karpicke, J. D., & Blunt, J. R. (2011a). Response to comment on “Retrieval practice produces more learning than elaborative studying with concept mapping.” *Science*, *334*(6055), Article 453. <https://doi.org/10.1126/science.1204035>
- Karpicke, J. D., & Blunt, J. R. (2011b). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. <https://doi.org/10.1126/science.1199327>
- Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology*, *90*(4), 677–691. <https://doi.org/10.1037/0021-9010.90.4.677>
- Kiewra, K. A. (1985). Investigating notetaking and review: A depth of processing alternative. *Educational Psychologist*, *20*(1), 23–32. [https://doi.org/10.1207/s15326985ep2001\\_4](https://doi.org/10.1207/s15326985ep2001_4)
- Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review*, *1*(2), 147–172. <https://doi.org/10.1007/BF01326640>
- Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, *36*(2), 416–428. <https://doi.org/10.3758/MC.36.2.416>
- Koriat, A., & Ma’ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478–492. <https://doi.org/10.1016/j.jml.2005.01.001>
- 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>
- Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of notetaking and test expectancy on the learning of text material. *The Journal of Educational Research*, *68*(10), 363–365. <https://doi.org/10.1080/00220671.1975.10884802>
- 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](https://doi.org/10.1016/0361-476X(88)90020-3)
- Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition*, *43*(1), 14–26. <https://doi.org/10.3758/s13421-014-0452-8>
- Lorenzet, S. J., Salas, E., & Tannenbaum, S. I. (2005). Benefiting from mistakes: The impact of guided errors on learning, performance, and self-efficacy. *Human Resource Development Quarterly*, *16*(3), 301–322. <https://doi.org/10.1002/hrdq.1141>
- Mayer, R. E. (1984). Aids to text comprehension. *Educational Psychologist*, *19*(1), 30–42. <https://doi.org/10.1080/00461528409529279>
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, *8*(4), 357–371. <https://doi.org/10.1007/BF01463939>
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice*, *41*(4), 226–232. [https://doi.org/10.1207/s15430421tip4104\\_4](https://doi.org/10.1207/s15430421tip4104_4)
- McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *39*(3), 462–476. <https://doi.org/10.3758/s13421-010-0035-2>
- McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review*, *1*(2), 113–145. <https://doi.org/10.1007/BF01326639>
- Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. <https://doi.org/10.3758/PBR.15.1.174>
- Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, *138*, Article 107296. <https://doi.org/10.1016/j.neuropsychologia.2019.107296>
- Metcalfe, J., & Xu, J. (2018). Learning from one’s own errors and those of others. *Psychonomic Bulletin & Review*, *25*(1), 402–408. <https://doi.org/10.3758/s13423-017-1287-7>

- Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science, 13*(3), 390–407. <https://doi.org/10.1177/1745691617710510>
- 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>
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2*(4), 267–271. <https://doi.org/10.1111/j.1467-9280.1991.tb00147.x>
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*(3), 413–448. <https://doi.org/10.3102/00346543076003413>
- Novak, J. D. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research in Science Education, 35*(1), 23–40. <https://doi.org/10.1007/s11165-004-3431-4>
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139173469>
- Pan, S. C., Sana, F., Samani, J., Cooke, J., & Kim, J. A. (2020). Learning from errors: Students' and instructors' practices, attitudes, and beliefs. *Memory, 28*(9), 1105–1122. <https://doi.org/10.1080/09658211.2020.1815790>
- Peterson, S. E. (1991). The cognitive functions of underlining as a study technique. *Reading Research and Instruction, 31*(2), 49–56. <https://doi.org/10.1080/19388079209558078>
- 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>
- Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27*(1), 91–109. [https://doi.org/10.1207/s15326985ep2701\\_7](https://doi.org/10.1207/s15326985ep2701_7)
- Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 69–80. <https://doi.org/10.1037/0278-7393.28.1.69>
- Reason, J. (1990). *Human error*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139062367>
- Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137*(1), 131–148. <https://doi.org/10.1037/a0021705>
- Roediger, H. L., III. (2013). Applying cognitive psychology to education: Translational educational science. *Psychological Science in the Public Interest, 14*(1), 1–3. <https://doi.org/10.1177/1529100612454415>
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review, 30*(2), 431–455. <https://doi.org/10.1007/s10648-017-9403-9>
- Skinner, B. F. (1958). Teaching machines: From the experimental study of learning come devices which arrange optimal conditions for self instruction. *Science, 128*(3330), 969–977. <https://doi.org/10.1126/science.128.3330.969>
- 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>
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89. [https://doi.org/10.1207/s1532690xci0201\\_3](https://doi.org/10.1207/s1532690xci0201_3)
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Terrace, H. S. (1963). Discrimination learning with and without "errors." *Journal of the Experimental Analysis of Behavior, 6*(1), 1–27. <https://doi.org/10.1901/jeab.1963.6-1>
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. <https://doi.org/10.1037/0022-0663.95.1.66>
- Wang, X., Mayer, R. E., Zhou, P., & Lin, L. (2021). Benefits of interactive graphic organizers in online learning: Evidence for generative learning theory. *Journal of Educational Psychology, 113*(5), 1024–1037. <https://doi.org/10.1037/edu0000606>
- Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist, 24*(4), 345–376. [https://doi.org/10.1207/s15326985ep2404\\_2](https://doi.org/10.1207/s15326985ep2404_2)
- Wong, S. S. H., & Lim, S. W. H. (2019). Prevention–permission–promotion: A review of approaches to errors in learning. *Educational Psychologist, 54*(1), 1–19. <https://doi.org/10.1080/00461520.2018.1501693>
- Wong, S. S. H., & Lim, S. W. H. (2021). The derring effect: Deliberate errors enhance learning. *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0001072>
- Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145*(7), 918–933. <https://doi.org/10.1037/xge0000177>
- Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1073–1092. <https://doi.org/10.1037/xlm0000363>

Received April 26, 2021

Revision received August 31, 2021

Accepted September 24, 2021 ■