Journal of Cognitive Psychology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pecp21

Retrieval practice effect and individual differences: current status and future directions

Marcos Felipe Rodrigues de Lima & Luciano Grüdtner Buratto

To cite this article: Marcos Felipe Rodrigues de Lima & Luciano Grüdtner Buratto (16 May 2024): Retrieval practice effect and individual differences: current status and future directions, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2024.2354963

To link to this article: https://doi.org/10.1080/20445911.2024.2354963

Published online: 16 May 2024.

JOURNAL OF COGNITIVE PSYCHOLOGY
https://doi.org/10.1080/20445911.2024.2354963

Routledge
Taylor & Francis Group

REVIEW ARTICLE

Retrieval practice effect and individual differences: current status and future directions

Marcos Felipe Rodrigues de Lima and Luciano Grüdtner Buratto

Departamento de Processos Psicológicos Básicos, Instituto de Psicologia, University of Brasília, Brasília, Brazil

ABSTRACT

Retrieval practice – the act of retrieving information from memory – can be implemented into educational contexts to improve learners' long-term retention. Here, we provide a narrative review that explores whether individual-difference variables moderate the retrieval practice effect. Studies have examined cognitive, personality, motivational, and attitudinal factors. No consistent relationships between retrieval practice effect and individual-difference variables were identified. Interpreting this literature is challenging due to methodological heterogeneity across studies. We describe the dual-memory framework and the working memory dependent dual-process account, two theoretical accounts that can generate theoretically-driven predictions for future individual-difference studies. In addition, we propose the following research agenda: (a) investigate the test-retest reliability of the retrieval practice effect under experimental paradigms typically used in studies on individual differences; (b) explore the relationship between individual differences and retrieval dynamics in free-recall tests; (c) pursue close replications; and (d) conduct investigations in real-life classrooms.

ARTICLE HISTORY
Received 20 December 2023
Accepted 7 May 2024

KEYWORDS
Retrieval practice; testing effect; learning; education; individual differences

How can learners optimize their learning in educational contexts? Instructors need to cover a substantial amount of course content, making inexpensive and easy-to-implement learning strategies preferable (Roediger & Pyc, 2012). Retrieval practice – the act of retrieving information from memory – could meet this need. Methods for practicing retrieval include self-testing, explaining concepts to a colleague, and summarising texts in a closed-book format. Multiple findings suggest that retrieval practice enhances long-term retention in both laboratory and classroom experiments (Abel et al., 2017; Carpenter et al., 2009; Jaeger et al., 2015; Roediger & Karpicke, 2006). Based on these findings, cognitive scientists have advocated for incorporating this learning technique into educational contexts (Agarwal & Bain, 2019; Brown et al., 2014; Dunlosky et al., 2013).

While retrieval practice improves retention more than other learning techniques on average, it may not benefit all learners equally. If this holds true, recommendations for the use of retrieval practice in educational contexts must be nuanced. Do individual-difference variables moderate the retrieval practice effect? In five main sections, this narrative review focuses on the relevant literature exploring this question. First, we characterise the phenomenon and delineate a justifiable recommendation based on the extant literature. Next, we summarise findings from individual-difference research on the retrieval practice effect, highlighting important methodological considerations for scholars. We then describe the dual-memory framework (Rickard & Pan, 2018) and the working memory dependent dual-process account (Zheng et al., 2023), two theoretical accounts that can generate theoretically-driven predictions for future individual-difference studies. Next, we discuss two important concepts for research on individual differences: reliability and portability. Lastly, in our concluding comments, we outline a research agenda for future studies.

The retrieval practice effect: experimental evidence

Consider the following classroom scenario. During the Sensation and Perception lecture, the instructor introduced concepts related to the biological basis of vision. In the final part of the lecture, students received sheets containing a series of statements about the topic. Some of these statements were complete (e.g. "The fovea, a region with the highest density of cone photoreceptors in the retina, relates to focal vision."), while others had

CONTACT Marcos Felipe Rodrigues de Lima lima.piraju@gmail.com Departamento de Processos Psicológicos Básicos, Instituto de Psicologia, Universidade de Brasília, Campus Darcy Ribeiro, ICC Sul, Sala ASS-012/5, CEP 70.910-900, Brasília, DF, Brasil

blank spaces (e.g. "The _____ can change its shape to adjust the focus of objects located at different distances." [lens]). The students were instructed to restudy the complete statements either by rereading it or by attempting to fill in the blanks for the incomplete ones, practicing retrieval of factual knowledge. After 15 min, the instructor collected the sheets. Two days later, at the beginning of the next lecture, students took an unexpected closed-book test, being asked to write down everything they remembered from the previous lecture.

Experiments on retrieval practice mirror certain aspects of this description: Participants encode a to-be-learned material (e.g. content about the biological basis of vision), practice it through either rereading or retrieval practice, and, after a given retention interval (e.g. 2 days), take a memory test. In the given example, the retrieval practice effect would be evident if, on average, students recalled more information on the memory test for material they practiced through retrieval compared to the material they reread (or compared to other control condition).

Using variations of this general procedure, experiments have demonstrated that retrieval practice enhances long-term retention compared with rereading (e.g. Carpenter, 2009; Pyc & Rawson, 2010) or semantic elaboration tasks (e.g. Coane, 2013; Karpicke & Blunt, 2011; but see Yang et al., 2021). Researchers have found the retrieval practice effect even after longer retention intervals (e.g. 16 weeks; Carpenter et al., 2009) and with educationally relevant materials (e.g. prose passages: Roediger & Karpicke, 2006; verbatim vs. inference questions: Karpicke & Blunt, 2011), indicating the ecological validity of the effect and its potential to improve not only the recall of materials, but also inference based on the recall of those materials.

Is this evidence sufficient to support retrieval practice recommendations in educational contexts? The majority of the extant evidence for the retrieval practice effect is experimental. Experimental psychologists design tightly controlled environments – using random assignment, counterbalancing, and holding other variables constant (Roediger & Yamashiro, 2020) – to make cause-and-effect claims possible (Cronbach, 1957). They assume homogeneous mechanisms underlie experimental effects across participants (Logie, 2018) and that the individuals in their sample are interchangeable instances with others from the same population (Borsboom et al., 2009). In other words, under this reasoning, it would make no difference whether, in a given experiment, Mary took or did not take part in the study along with other *n* participants. This is because process inference – rather than population inference – often is the goal in experimental psychology (cf. Hayes, 2022, pp. 65–67).

The situation changes slightly when applying a psychological principle. For instance, from Mary's perspective, what matters is whether retrieval practice is useful for her, not for instances supposedly interchangeable with her. Therefore, returning to the question from the previous paragraph, there is sufficient evidence to legitimately recommend retrieval practice in educational contexts to improve students' learning and retention, compared with no/filler activity, testing with fewer questions, and rereading (Yang et al., 2021). Instructors suggesting retrieval practice can anticipate an overall improvement in retention. However, concluding that retrieval practice will improve retention for *each* individual would be an ecological fallacy (McDermott, 2021). The assumption of homogeneity of the retrieval practice effect, often only implicit in experimental research, is likely unrealistic. For example, while experiments demonstrate the retrieval practice effect at the group level, there are also studies reporting that a sizeable number of participants did not benefit from retrieval practice (e.g. Minear et al., 2018).

The retrieval practice effect at the participant level is defined as a difference score: recall performance on retrieval practice condition minus recall performance on rereading condition (see, e.g. Agarwal et al., 2017, Figure 2; Brewer & Unsworth, 2012, Figures 1–2; Pan et al., 2015, Figure 1). For example, if a participant recalled .70 and .50 of retrieval practice and rereading items, respectively, we can say that their retrieval practice effect was .20. Considering data from five combined experiments (Brewer & Unsworth, 2012; Pan et al., 2015, Experiments 1 and 2; Robey, 2019, Experiments 1 and 2; $N_{combined} = 739$; data made publicly available by Brewer et al., 2021), a group-level retrieval practice effect was observed, $M_{Difference} = .11$, 95% CI [.09, .12] 28.55, and 71.45% of participants benefitting from retrieval. Importantly for our discussion, 28.55% of participants in Brewer et al.'s data did not benefit from retrieval practice. A natural question arises: Do individual-difference variables moderate the retrieval practice effect?

Around a decade ago, researchers had already questioned whether retrieval practice would benefit different learners equally (Brewer & Unsworth, 2012). This question holds both practical and theoretical significance. From a practical standpoint, if researchers identify a characteristic that predict whether learners will benefit from retrieval practice, it implies that recommendations for using retrieval practice in educational contexts should be qualified based on this boundary condition. From a theoretical standpoint, identifying profiles of learners who consistently do not benefit from retrieval practice would inform contemporary accounts in the field. For example, take fluid intelligence (gF) – the

ability to solve novel problems, engage in inductive, sequential, and quantitative reasoning, which is typically measured by nonverbal and supposedly culture-free tasks (Walrath et al., 2020) – as an example. If individuals with lower gF do not benefit from retrieval practice, then: (a) instructors should use other learning techniques with learners with this profile; and (b) researchers would need to propose or revise hypotheses that account for this moderator variable.

In summary, the retrieval practice effect has been well established in experimental research (for meta-analytic reviews, see Rowland, 2014; Yang et al., 2021). The phenomenon has been replicated across various learning conditions, materials, and criterion tests, making retrieval practice a highly useful learning technique in educational contexts (Dunlosky et al., 2013). However, this utility is expected to manifest at the group level, potentially obscuring the fact that retrieval practice may not work for some learners. It is this aspect that individual-difference research aims to explore.

Individual-difference variables as potential moderators of the retrieval practice effect

Brewer and Unsworth (2012) outline three potential relationships between individual-difference variables and the retrieval practice effect, but these relationships are not exhaustive (they could be nonlinear, for example). Brewer and Unsworth primarily focus on cognitive abilities. However, we extend their description by including the term *trait* to encompass personality, motivational, and attitudinal variables that have interested researchers.

The first scenario supports the universal recommendation of retrieval practice for all learners: The participant-level retrieval practice effect is consistently unrelated to individual-difference variables. Learners across the spectrum of these variables are equally likely to benefit practicing retrieval. The second scenario suggests that retrieval practice is more beneficial for those learners who already possess a higher level of latent ability or trait. In this case, learners who employ their cognitive resources suboptimally might benefit less from retrieval practice. If this holds true, instructors would need to explore alternative methods to enhance the learning of students with lower latent abilities or traits. Lastly, the third scenario suggests that learners with higher latent abilities or traits are already employing their cognitive resources optimally and, consequently, benefit less – or do not benefit at all – from retrieval practice. Conversely, learners with lower latent abilities or traits, presumed to use suboptimal encoding strategies, could reap greater benefits from retrieval practice.

Individual differences in the retrieval practice effect

Tables 1 and 2 provide summaries of characteristics and results, respectively, for studies on individual differences in the retrieval practice effect. To clarify, this narrative review focuses on individual-difference variables as potential moderators of the retrieval practice effect in healthy individuals. Therefore, we did not consider studies investigating whether the retrieval practice effect would emerge in different populations, such as comparing a clinical group with an age-matched healthy control group (e.g. Sumowski et al., 2010; for a review, see Lima et al., 2020). Similarly, studies involving *between*-subject manipulations of learning technique (i.e. retrieval practice vs. control condition) were excluded from these tables since they did not measure the retrieval practice effect at the participant level (e.g. Jaeger et al., 2015; Stenlund et al., 2017).

Most studies used word pairs as the to-be-learned material and cued-recall tests during both the practice and the final-test phases, and rereading as the control condition (Table 1). Regarding the number of practice cycles and the provision of feedback, the scenario was more heterogeneous. Studies incorporating corrective feedback or multiple practice cycles introduce indirect benefits of retrieval practice (Karpicke, 2017). Given this heterogeneity, it should be emphasized that there is no single retrieval practice effect, but rather multiple different effects, depending on the experimental paradigm adopted.

Some studies focused on personality variables, such as grit, the person's perseverance and passion for pursuing long-term goals (Duckworth et al., 2007); need for cognition, the person's tendency to enjoy spending time in effortful cognitive activities (Cacioppo & Petty, 1982); and test anxiety, a situation-specific personality trait related to test taking (Spielberger, 2010). These studies were predominantly conducted by the same research team, and they did not report any significant associations (Bertilsson et al., 2021; Wiklund-Hörnqvist et al., 2022).[1] A notable exception was an experiment that observed that participants with lower test anxiety

[1]Minear et al. (2018) also measured need for cognition and grit, along with measures of the Big Five constructs, academic entitlement, academic self-efficacy, test anxiety, and stress (see their Footnote 1). However, Minear et al. did not report the relationships observed between these variables and the participant-level retrieval practice effect.

Table 1. Characteristics of studies on individual differences in the retrieval practice effect.

Study	Sample	Material	Retrieval practice task	Cycles	Feedback	Retention interval	Final test
Agarwal et al. (2017)	College students	110 general knowledge facts	CR test	1	Yes and no	10 min or 2 days	CR test
Bertilsson et al. (2021)	Upper secondary-level students	60 Swahili-Swedish word pairs	CR test	6	Yes	5 min, 1 week, or 4 weeks	CR test
Brewer and Unsworth (2012)	College students	40 weakly associated English-English word pairs	CR test	1	Yes	1 d	CR test
Cogliano et al. (2019)	College students	Topics in an educational psychology course	MC test	1	Yes	1 week	Chapter exam, MC test
Glaser and Richter (2023)	Students enrolled in a teacher-training program	One of five topics from psychology curriculum	Short-answer questions	1	Yes	1 week	Short-answer questions and MC test
Jonsson et al. (2021), Exp. 1	Upper secondary-level students	60 Swahili-Swedish word pairs	CR test	6	Yes	5 min, 1 week, or 4 weeks	CR test
Jonsson et al. (2021), Exp. 2	Upper secondary-level students	60 Swahili-Swedish word pairs	CR test	6	Yes	5 min, 1 week, or 4 weeks	CR test
Minear et al. (2018)	College students	48 Swahili-English word pairs	CR test	4	Yes	2 days	CR test
Moreira et al. (2019), Exp. 1	6th grade students	One encyclopedic text	Fill-in-the-blank test	1	No	1 week	Fill-in-the-blank and MC tests
Moreira et al. (2019), Exp. 2	4th grade students	One encyclopedic text	Fill-in-the-blank test	2	No	1 week	Fill-in-the-blank and MC tests
Pan et al. (2015), Exp. 1	Adults from the Amazon Mturk worker pool	40 weakly associated English-English word pairs	CR test	1	Yes	1 d	CR test
Pan et al. (2015), Exp. 2	College students	40 weakly associated English-English word pairs	CR test	1	Yes	1 d	CR test
Robey (2019), Exp. 1	College students	40 English-English word pairs from five categories[a]	CR test	2	Yes	30 min	CR test
Robey (2019), Exp. 2	College students	40 English-English word pairs from five categories[a]	CR test	2	Yes	15 min	CR test
Tse et al. (2019), Exp. 1	College students	80 general knowledge facts	CR test	2	Yes	Immediate and 2 days	CR test
Tse et al. (2019), Exp. 2	College students	80 general knowledge facts	CR test	2	Yes	Immediate and 2 days	CR test
Tse and Pu (2012)	College students	40 Swahili-English word pairs	CR test	12	No[b]	1 week	CR test
Wenzel and Reinhard (2019), Exp. 2	College students	Textbook chapter on the brain's lateralisation	MC and open-ended questions	1	Yes	1 week	MC and open-ended questions
Wiklund-Hörnqvist et al. (2022)	Upper secondary-level students	60 Swahili-Swedish word pairs	CR test	6	Yes	1 week	CR test
Yang et al. (2020)	College students	Five 18-word lists	FR test	1	No	5 min	Cumulative FR test
Zheng et al. (2023)	College students	64 Fribble-Chinese associations	CR test	1	No	1 d	CR test

Note: With the exception of Cogliano et al. (2019), who used a no-test control condition, all other studies adopted restudy/rereading as the control condition. Cycles represent the number of practice rounds for a given material. CR = cued-recall. MC = multiple-choice. FR = free-recall.

[a]Related-high imageability nouns, related-low imageability nouns, unrelated-high imageability nouns, unrelated-low imageability nouns, and nonsense words.

[b] Although there was no feedback after retrieval attempts, the repeated retrieval practice condition (S-T-S-T-S-T-S-T-S-T-S-T) was compared with the repeated study condition (S-S-S-T-S-S-S-T-S-S-S-T), where S stands for study blocks and T stands for test (retrieval practice) blocks. [c] Retention interval estimated based on the article text.

benefited more from retrieval practice (Tse & Pu, 2012). However, this finding was not replicated (Tse et al., 2019). Tse and Pu's focus was on a three-way interaction, which we will discuss below.

The majority of studies have focused on cognitive abilities, such as gF; attentional control, encompassing the range of processes that allow learners to selectively focus and actively maintain task-relevant information, guiding their thoughts and actions in face of distractors (Unsworth et al., 2021); crystallized intelligence, referring to the application of learned skills and knowledge accumulated over a lifetime (Walrath et al., 2020); reading skills, including word and pseudoword naming abilities and sentence comprehension (Moreira et al., 2019); and general cognitive abilities, a latent factor combining different abilities (Jonsson et al., 2021). These studies indicated that the retrieval practice effect is independent of these variables (Brewer & Unsworth, 2012; Jonsson et al., 2021; Minear et al., 2018; Moreira et al., 2019). It is important to note, however, that these constructs have been explored by no more than two studies.

Table 2. Summary of studies on individual differences in the retrieval practice effect.

Study	AC	Cog	EM	Error	gC	gF	Grit	Motivation	NFC	Prior knowledge	Reading	Retrieval	Test anxiety	WMC
Agarwal et al. (2017)														–[b]
Bertilsson et al. (2021)									X					X
Brewer and Unsworth (2012)	X		–			–								X
Cogliano et al. (2019)										–				
Glaser and Richter (2023)				X				X	X	X		X	X	
Jonsson et al. (2021), Exp. 1		X												
Jonsson et al. (2021), Exp. 2		X												
Minear et al. (2018)					X	X								X
Moreira et al. (2019), Exp. 1											X			
Moreira et al. (2019), Exp. 2						X					X			X
Pan et al. (2015), Exp. 1			X											
Pan et al. (2015), Exp. 2			X											
Robey (2019), Exp. 1			X			X								
Robey (2019), Exp. 2			X			X								
Tse et al. (2019), Exp. 1													X	X
Tse et al. (2019), Exp. 2													X	–[c]
Tse and Pu (2012)													–	X
Wenzel and Reinhard (2019), Exp. 2						+								
Wiklund-Hörnqvist et al. (2022)									X					
Yang et al. (2020)														–
Zheng et al. (2023)														+[d]

Note: Exp. = experiment. AC = attentional control. Cog = cognitive ability (composite index including measures of episodic memory, fluid intelligence, working memory capacity, visuospatial short-term memory, and updating). EM = episodic memory. Error = attitudes towards errors. gC = crystalized intelligence. gF = fluid intelligence. Grit = perseverance to achieve long-term goals. Motivation = learning and performance motivation. NFC = need for cognition. Reading = reading ability. Retrieval = retrievability (performance during practice phase). WMC = working memory capacity. The X denotes no effect, the plus symbol denotes a positive effect, and the minus symbol denotes a negative effect.

[a]Prior knowledge was an induced (i.e. manipulated) individual-difference variable. [b]Agarwal et al. (2017) observed a negative significant correlation and three nonsignificant ones. [c]Tse et al. (2019, Experiment 2) observed two negative regression coefficients (one significant and one marginally significant) and two nonsignificant ones. [d]The relationship was positive only for the low-frequency condition.

Nine studies assessed working memory capacity (WMC), the ability to simultaneously storage information and process additional information (Conway et al., 2005). Four of them found weak, nonsignificant relationships between WMC and the retrieval practice effect (*r*s ranging from –.12 and .17; Bertilsson et al., 2021; Brewer & Unsworth, 2012; Minear et al., 2018; Moreira et al., 2019, Experiment 2). Studies investigating test anxiety, WMC, and the retrieval practice have shown inconsistent results. Tse and Pu (2012) found that test anxiety negatively correlated with the retrieval practice effect for participants with lower WMC ($r = -.39$) but this correlation was nonsignificant for higher WMC ($r = -.14$). However, in eight hierarchical regression models, Tse et al. (2019) overall did not replicate this pattern.

Agarwal et al. (2017) assessed the effect of an individual-difference variable contingent on other design features. Specifically, Agarwal et al. found that participants with lower WMC benefited more from retrieval practice with feedback in a 2-day retention interval – but not in other three conditions, retrieval practice with feedback in a 5-min retention interval, retrieval practice without feedback in a 5-min retention interval, and retrieval practice without feedback in a 2-day retention interval. Yang et al. (2020) were primarily interested in the indirect benefit of retrieval practice on subsequent learning of new material – the *forward testing effect*. Although not their main focus, Yang et al. found greater retrieval practice effects for participants with lower WMC, acknowledging the possibility that this finding might be partially driven by the forward testing effect. Finally, Zheng et al. (2023) simultaneously considered WMC and task demands. They found that both that low – and high-WMC participants benefitted from retrieval practice for high-frequency items (which presumably recruits fewer working memory resources), but only high-WMC participants benefitted from retrieval practice for low-frequency items (which recruits more working memory resources).

Studies examining the relationship between the retrieval practice effect and episodic memory abilities, as well as gF, have yielded mixed results. Brewer and Unsworth (2012) observed that participants with lower episodic memory abilities and lower gF scores, as opposed to those with higher abilities and scores, benefited the most from retrieval practice. However, in two replication attempts, Pan et al. (2015) did not observe the same pattern of results concerning the episodic memory measure.

Other studies also failed to find significant correlations between the retrieval practice effect and

measures of episodic memory or gF (Moreira et al., 2019; Robey, 2019). Wenzel and Reinhard (2019, Experiment 2) reported a result pattern that was the opposite of that reported by Brewer and Unsworth (2012): Retrieval practice benefited participants with average and above-average gF, but did not benefit participants with below-average gF. Finally, Minear et al. (2018), who also considered item difficulty, identified a three-way interaction: The retrieval practice effect for easy items was greater for lower gF participants, while the retrieval practice effect for difficult items was greater for higher gF participants. This result, however, only emerged when the analyses were restricted to individuals who benefited from retrieval practice.

Studies in this section primarily took place in laboratory settings, using simple stimuli as the to-be-learned material (see Table 1). Given the interest in implementing retrieval practice in educational contexts, it is essential to investigate whether learners' characteristics moderate the retrieval practice effect in authentic classrooms (Trumbo et al., 2021). We are aware of only two studies that have explored this question. Cogliano et al. (2019) examined whether the benefits of retrieval practice are moderated by prior topic knowledge – assessed through pre-tests – in an educational psychology course. Specifically, they found that the advantage of retrieval practice over a no-test control condition was greater for low prior knowledge than those with high prior knowledge topics. Glaser and Richter (2023) investigated whether retrievability (i.e. performance during practice phase), prior knowledge, need for cognition, learning and performance motivation, test anxiety, and attitudes toward errors moderate the magnitude of the retrieval practice effect in a higher-education teacher-training program. In a series of analyses conducted separately for each moderator of interest, Glaser and Richter found that all interaction terms were nonsignificant, suggesting that the retrieval practice effect was independent of these learners' characteristics.

In summary, studies on individual differences in the retrieval practice effect have examined a wide range of cognitive, personality, motivational, and attitudinal factors. At this point, consistent relationships between individual-difference variables and the retrieval practice effect remain to be identified. At most, observed associations tend to be null or negative – individuals with lower scores in a given variable tend to benefit more from retrieval practice than those with higher scores (but see Zheng et al., 2023). Interpreting this literature is challenging due to methodological heterogeneity across studies, such as the types of materials used (e.g. word pairs, educationally relevant texts), retention intervals (ranging from immediate to 4 weeks), and even analytical methods (e.g. bivariate correlation, multiple regression). Therefore, it is possible that one or more of these differences contributed to the diverse outcomes across the studies.

Methodological heterogeneity across studies

In human memory research, the interplay between material characteristics, task complexity, and individual differences can lead to the recruitment of different cognitive processes (Healey & Kahana, 2014; Logie, 2018), thus contributing to inconsistent results. If the relationship between an individual-difference variable and the retrieval practice effect can be influenced by other variables (i.e. higher-order interactions), these variables need to be carefully considered in future investigations (for a related argument, see Zheng et al., 2023). The following illustrative examples aim to point out important aspects of methodological heterogeneity that need to be considered by researchers.

First, Agarwal et al. (2017) and Brewer and Unsworth (2012) examined whether WMC moderates the retrieval practice effect. While their experiments varied in several dimensions, we focus on one here: The number of intervening items between successive presentations of the same item (i.e. lag), which ranged from 0 to 9 in the Agarwal et al. study. These lags imposed different demands on learners' capacity to keep the information active in the face of distractions, such as the interference of subsequent items (Conway et al., 2005). In contrast, the Brewer and Unsworth study had a lag of at least 20 items, thus exceeding learners' capacity, and rendering WMC irrelevant. During the practice phase, participants in the Agarwal et al. study performed better (approximately 80%) than those in the Brewer and Unsworth study (46%). Of course, differences in material difficulty could also account for the different results. Nonetheless, exploring this three-way interaction (Learning Technique $\times$ WMC $\times$ Lag) could indeed open up an intriguing research avenue.

Second, the number of retrieval opportunities varies across studies (see Table 1). In experiments using word pairs, participants might transition from mediated to direct retrieval with extended practice (Crutcher & Ericsson, 2000; Kole & Healy, 2013). Consequently, the varying number of retrieval practice opportunities could engage different cognitive mechanisms. If different learners have distinct learning trajectories, the relationship between individual-difference variables and the retrieval practice effect might change over time. The key point here is that researchers should convert between-study differences into independent

variables to identify the so-called "hidden moderators" (Klein et al., 2018).

The final example pertains to how items are presented during the practice phase (Abel et al., 2017). Figure 1 illustrates three different designs (cf. Gupta et al., 2024). Items assigned to one condition can be temporally separated from those in another condition. This order – retrieval-practice-first blocked or rereading-first blocked designs – might (e.g. Robey, 2019) or might not be (e.g. Brewer & Unsworth, 2012) counterbalanced across participants. Alternatively, rereading and retrieval practice items can be randomly intermixed during the practice phase (mixed practice design; e.g. Pan et al., 2015).

Crucially, in a study involving a 5-list learning, Pastötter et al. (2011) found that retrieving information from episodic, semantic, or short-term memory after each one of the four initial lists, as opposed to rereading them, enhanced the encoding of the fifth list, as measured by a later memory test. Likewise, Gupta et al. (2024). found that studies using retrieval-practice-first blocked designs are partially confounded by this forward testing effect (see Figure 1), that is, the retrieval practice effect appears smaller, possibly because the blocked presentation benefits the subsequent encoding of the material in the rereading condition (for an independent, but similar argument, see Mulligan et al., 2022).

The significance of these findings lies in their indication that the experimental design could impact the retrieval practice effect at the group level, thereby influencing the functional relationship between this effect and individual-difference variables. Notably, studies using blocked designs observed smaller retrieval practice effects (Brewer & Unsworth, 2012; Minear et al., 2018; Robey, 2019) compared to those using mixed practice designs (Pan et al., 2015). Importantly, Gupta et al. (2024) claimed that higher ability learners are likely to benefit more from this confounding forward testing effect. If this claim is true, then it is possible that such methodological characteristic adds noise in the retrieval practice effect at the participant level (e.g. by changing the rank order of participants in terms of benefits from retrieval practice). Additionally, other findings suggest that even unrelated tasks, if administered *before* the memory task, might also lead to this confounding effect, improving encoding in the rereading condition (Pastötter et al., 2011).

How can these potential interpretive problems be mitigated? Researchers must be explicit about the specific effect they intend to investigate. If the focus is on the direct effect of retrieval practice, it is preferable that the experiment adopts a mixed design with only one practice cycle and without feedback after retrieval practice. Alternatively, for those interested in a combination of direct and indirect benefits of retrieval practice (as can be the case for applied researchers in authentic educational contexts), a mixed practice design may be preferred. Finally, to prevent interpretive issues arising from the confounding forward testing effect, researchers should avoid retrieval-practice-first blocked practice designs and the application of cognitive tasks (e.g. gF tasks) before the main experiment.

Theoretical accounts

We do not aim to present a comprehensive review of theoretical accounts of the retrieval practice effect (for this purpose, see Karpicke et al., 2014). Instead, we seek to assess how these accounts can contribute to research on individual differences. To date, influential accounts of the retrieval practice effect – such as the mediator-effectiveness hypothesis (Pyc & Rawson, 2010), the relational-processing hypothesis (Rawson & Zamary, 2019), and the episodic context account (Lehman et al., 2014) – typically predict group-level patterns, being, at least in their original formulations, silent about potential individual differences in the retrieval practice effect. As a consequence, studies on individual differences often lack a clear theoretical orientation. However, contemporary accounts might be capable of accommodating evidence of the moderating role of individual differences. We provide some possibilities below.

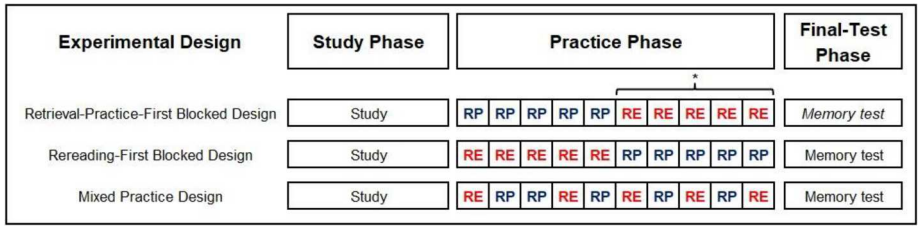

Figure 1. The potential confounding role of the forward testing effect in studies on the retrieval practice effect. Note. RE = rereading. RP = retrieval practice. Asterisk indicates which set of items is expected to benefit from the forward testing effect. Italics denote the memory test in which is expected a better performance in the rereading condition (in retrieval-practice-first blocked design), compared with this same condition in the rereading-first blocked design. Based on Gupta et al. (2024).

Consider the elaborative retrieval hypothesis (Carpenter, 2009), which posits that during the search for a target, retrieval practice activates information related to retrieval cues, thereby activating a semantic network that may facilitate subsequent retrieval through multiple pathways. In a individual-difference study, Minear et al. (2018, p. 1476) have argued: "One might speculate that individuals higher in crystallized intelligence (e.g. vocabulary knowledge) would have more elaborate semantic networks and this would be most evident for the more difficult items, yielding a larger [retrieval practice] effect on difficult items for individuals high in this measure than those scoring low." In the same vein, Buchin and Mulligan (2022) claimed that the elaborative retrieval hypothesis predicts a greater retrieval practice effect for high-prior knowledge information than for low-prior knowledge information. One can argue that the same reasoning applies when prior knowledge is an individual-difference, instead of an experimentally manipulated, variable.

Another theoretical account that can accommodate predictions for studies on individual differences is the dual-memory framework (Rickard & Pan, 2018), which relies on the idea of strength of memory traces. Initially developed to account for findings from experiments using cued-recall tasks, the dual-memory framework posits that rereading and retrieval practice items are recalled if their memory strengths reach a fixed response threshold. Importantly, while rereading items are modelled by a single memory strength dimension, retrieval practice items are modelled by two distinct and independent memory strength dimensions (Rickard & Pan, 2018).

The dual-memory framework predicts the retrieval practice effect based on the probability of correct responses in the rereading condition (rereading proportion correct), PC_R, using a quadratic function: *Retrieval practice effect* $= PC_R - PC_R^2$ (Rickard & Pan, 2018). This function, illustrated as the solid line in Figure 2, suggests that larger retrieval practice effects are expected when $PC_R = .50$, decreasing as PC_R approaches 0 or 1. Now, consider the case where an individual-difference variable (e.g. gF) has a positive correlation with memory performance (for a review providing evidence of latent correlations between gF and long-term memory, see Unsworth, 2019). What will be the relationship between gF and the retrieval practice effect? The dual-memory framework proposes that the correlation between gF and the retrieval practice effect will be "(in part) a joint consequence of (1) the relation between the [individual-difference] variable and [rereading] proportion correct, and (2) the relation between [rereading] proportion correct and the [retrieval practice effect]" (Rickard, 2020, p. 789).

An illustrative example is provided below. Suppose Alice, Bella, and Chloe recall .75, .65, and .55 of the rereading items, respectively, in a hypothetical experiment with a shorter retention interval or an easier final test. Further, assume that their true gF scores were ranked as follows: *Alice* > *Bella* > *Chloe*. In this scenario, the dual-memory framework predicts retrieval practice effects of .19, .23, and .25 for Alice, Bella, and Chloe, respectively (rounding to two decimal places; see the dark gray points in Figure 2). In this example, the average $PC_R > .50$, and gF scores and the retrieval practice effects are negatively correlated, a consequence

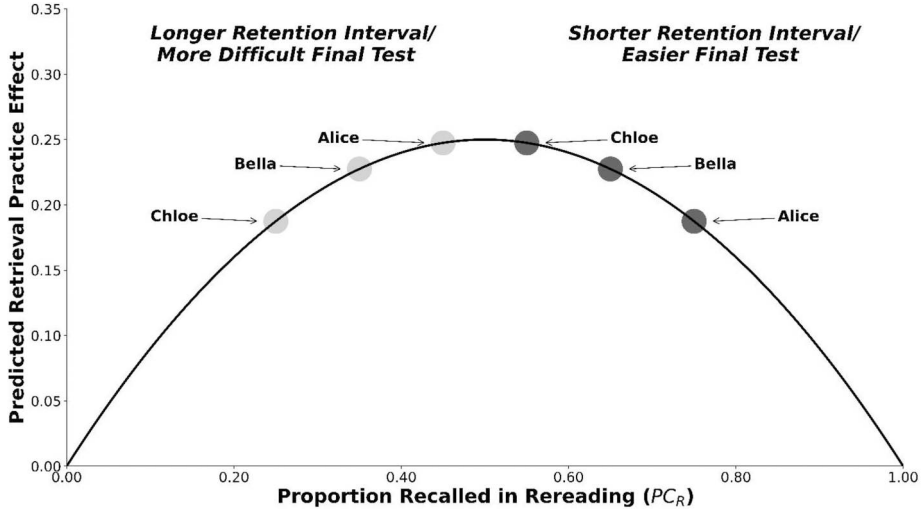

Figure 2. Illustrative Example of the Dual-Memory Framework in Two Hypothetical Experiments with Different Retention Intervals and Difficulties in the Final Test. Note. Solid line represents the predicted retrieval practice effect (y-axis) as a function of the proportion recalled in rereading (x-axis).

of the quadratic relationship between PC_R and the predicted retrieval practice effect described earlier.

What if this hypothetical experiment were instead one with a longer retention interval or a more difficult final test? In this new scenario, some forgetting would be expected. For instance, let us assume that Alice, Bella, and Chloe recall .45, .35, and .25 of the rereading items, respectively. In this hypothetical situation, the participants' rank-order in restudy performance was preserved, consistent with studies indicating a high correlation between immediate and delayed recall (Jonsson et al., 2014). The dual-memory framework now predicts retrieval practices of .25, .23, and .19 for Alice, Bella, and Chloe, respectively, reversing the order of the magnitude of the retrieval practice effects (see the light gray points in Figure 2). Now, with an average $PC_R < .50$, assuming that the true gF scores remain the same, the gF scores and the retrieval practice effects are positively correlated. Of course, this pattern will change depending on the strength of the correlation between PC_R and an individual-difference variable (see Rickard, 2020). A potential avenue for future studies is to randomly assign participants to tasks that induce an average PC_R above or below .50 (e.g. easier and more difficult tests, respectively) and explore whether the correlations between an individual-difference variable and the retrieval practice effect align with the predictions of the dual-memory framework.

Recently, Liu and colleagues proposed an alternative dual-process account, the working memory dependent dual-process account (Liu et al., 2018; Zheng et al., 2023). This model posits that retrieval practice enhances memory through both the retrieval attempt itself (often referred to as a *direct effect*; cf. Karpicke, 2017) and a post-retrieval re-encoding of retrieved information. Zheng et al. used a modified source of activation confusion (SAC) model to computationally implement their dual-process account. Here, we provide only a summary of the model. The SAC model represents items as nodes in a localist associative network (Diana et al., 2006). Each node has a base-level strength (degree of learning or storage), which increases as its respective node is activated (Popov & Reder, 2020). Details about the rules governing the increase in strength levels are beyond the scope of this review (for a comprehensive description, see Popov & Reder, 2020).

One of the assumptions of the SAC model is that the working memory resource pool, which is assumed to differ across learners, is depleted and recovered over time. Resources are depleted during retrieval attempts (Popov & Reder, 2020). Items with weaker base-level strength (e.g. low-frequency items) deplete more working memory resources than items with stronger base-level strength (e.g. high-frequency items). In the context of retrieval practice experiments, an additional assumption is that during retrieval practice, extra working memory resources are consumed to an extent that is inversely proportional to the item's current base-level strength (Zheng et al., 2023).

As the hypothesised processes – retrieval attempt and post-retrieval re-encoding – occur sequentially, when working memory resources are depleted during the retrieval attempt, they may be insufficient during the post-retrieval re-encoding process, thereby preventing low-WMC learners from fully benefiting from retrieval practice. The implication is straightforward: WMC will moderate the retrieval practice effect in tasks that working memory demands surpasses working memory resources for some, but not all, participants. As described earlier, Zheng et al. (2023) obtained initial support for their predictions. One differential in the Zheng et al. study is the absence of feedback after retrieval practice. A follow-up study would include feedback after retrieval practice. According to the model, feedback can be thought as a working memory aid (Zheng et al., 2023), especially learners with fewer resources (e.g. Tse et al., 2010). If this is the case, then providing feedback should restore the retrieval practice effect for low-WMC participants in the low-frequency condition.

As a final note, ancillary measures on the final-test phase play relevant roles in some theoretical accounts, such as the shape of cumulative recall over time; and the production, shift, and retrieval of mediators; and retrieval organisation (Lehman et al., 2014; Pyc & Rawson, 2010; Rawson & Zamary, 2019; for a discussion on ancillary measures, see Karpicke, 2017). For example, in some situations, semantic organization mediates the relationship between learning condition (rereading, retrieval practice) and final-test performance (Cavendish et al., 2022; Rawson & Zamary, 2019). However, to date, there is no study on individual differences investigating whether participants' characteristics are related with semantic organisation or others ancillary measures (two studies investigate self-reported measures; see Minear et al., 2018; Robey, 2019). Given that might be multiple ways to achieve high performance in cognitive tasks (Logie, 2018), this appears to be an interesting avenue for future research.

In summary, while most theoretical accounts do not make predictions about the moderating role of individual-difference variables in the retrieval practice effect, they may be able to accommodate such effects through additional assumptions. Two dual-process accounts – the dual-memory framework (Rickard, 2020; Rickard & Pan, 2018) and the working memory

dependent dual-process account (Liu et al., 2018; Zheng et al., 2023) – have the advantage of being capable to describing how different empirical patterns emerge from the complex interaction between learner characteristics, materials, and tasks.

Two important concepts for research on individual differences

In methodology, certain concepts exhibit polysemic characteristics. In experimental research, effects are said to be *reliable* when they are replicable across participants or situations (e.g. Carpenter, 2009, p. 1563) or even when they are statistically significant (e.g. Roediger & Karpicke, 2006, p. 252). In individual-difference research, scores of a measure are deemed reliable, in a psychometric sense, if they consistently yield error-free scores (Nunally & Bernstein, 1994). Importantly, each sense of reliability requires its own evidence, and demonstrations of reliable experimental effects do not guarantee test-retest reliability (Hedge et al., 2018; Logie et al., 1996). Here we specifically advocate for test-retest reliability, as opposed to other forms of reliability (e.g. internal consistency), because poor test-retest reliability could indicate that conflicting results (e.g. Brewer & Unsworth, 2012 vs. Pan et al., 2015) are due to participants inconsistently benefiting more or less from retrieval practice.

The retrieval practice effect is reliable in the experimental sense (Rowland, 2014; Yang et al., 2021), but its test-retest reliability is only supported by preliminary evidence (Lima & Buratto, 2023), which was limited to a 5-min retention interval. Minear et al. (2018) found a non-significant correlation ($r = .39$) between the retrieval practice effect for easy and difficult items, suggesting that it may not behave in a trait-like manner (McDermott, 2021). This is an important issue since individual-difference research assumes, at least implicitly, that the retrieval practice effects at the participant level are consistent over time. More studies are needed to investigate whether the retrieval practice effect exhibits test-retest reliability across various experimental paradigms, including different retention intervals, tasks, and materials (Brewer & Unsworth, 2012; McDermott, 2021).

Closely related, the retrieval practice effect is not a singular phenomenon but rather comprises multiple distinct effects. In other words, the retrieval practice lacks *portability*, meaning that due to the methodological heterogeneity in individual-difference studies, retrieval practice effects estimated in these studies do not represent a fixed unit (Rouder & Haaf, 2019) – they can be thought as random effects, akin to the meta-analysis literature (Hedges & Vevea, 1998). This variability in effects across different procedures raises the possibility that these effects may have different relationships with individual-difference variables. For example, holding other variables constant, retrieval practice effects from two experiments, one with a single practice cycle and another with 20 practice cycles, might have varying correlations with individual-difference variables. This is because different experimental designs recruit cognitive abilities at different levels (cf. Logie, 2018).

The main challenge in interpreting divergent results in the retrieval practice literature is that these studies often vary in multiple factors simultaneously. The Pan et al. (2015) study represents the only explicit attempt at close replication in this field. In cases of conflicting results of studies using heterogeneous procedures, efforts have not been made to follow up investigating the source for discrepancies. We advocate for close replications, where only one factor is manipulated at a time, to identify "hidden moderators" in the relationship between individual-difference variables and the retrieval practice effect (Klein et al., 2018).

Concluding comments

Some limitations of this study need to be acknowledged. We did not conduct a systematic review or a meta-analysis computing summary statistics of effect size. While systematic and meta-analytical reviews are considered superior to narrative reviews as the number of studies increases (see, e.g. Borenstein et al., 2009), the number of studies reviewed here is comparatively small compared to other reviews in the field ($k = 159$ in Rowland, 2014; $k = 222$ in Yang et al., 2021). This is particularly problematic for analyses of heterogeneity and moderators, as estimates of heterogeneity tend to be imprecise when $k < 20$ (Hedges & Vevea, 1998). The field could benefit from a meta-analysis as the literature matures.

One could argue that the including studies with a heterogeneity of measures is an additional limitation in this review. For example, studies employed cued-recall (e.g. Minear et al., 2018), fill-in-the-blank (Moreira et al., 2019), short-answer (Glaser & Richter, 2023), and multiple-choice tests (Wenzel & Reinhard, 2019). Additionally, one study compared retrieval practice with a no-test control condition (Cogliano et al., 2019). From an applied viewpoint, researchers could either compare retrieval practice with treatment-as-usual standards or include the type of control condition as a categorical moderator in future meta-analytical reviews (see, e.g. Yang et al., 2021). From a theoretical viewpoint, however, the implications of using a no-test control condition for contemporary accounts are not entirely clear. For example, the dual-memory framework could predict greater benefits of retrieval practice when using a no-

test control condition, although this is not very informative. In reviews and in future studies, the decision to include or not include different kinds of final tests and control conditions should be based on the tradeoff of applied and theoretical considerations.

In this manuscript, we have reviewed the literature addressing the moderator role of individual-difference variables in the magnitude of the retrieval practice effect, highlighting methodological, theoretical, and conceptual considerations for researchers. Based on this review, we outlined the following research agenda: (a) examine whether the impact of individual differences on the retrieval practice effect depends on other individual differences (e.g. participants' spontaneous encoding and retrieval strategy use) and contextual factors (e.g. lag, extended practice); (b) investigate the reliability of the retrieval practice effect (e.g. test-retest reliability; reliability across tasks, materials, and retention intervals); (c) explore whether an individual-difference variable correlates with the retrieval practice effect as predicted by the dual-memory framework; (d) pursue close replications to ensure that the observed relationships are indeed reliable; and (e) explore the relationship between individual-difference variables and the retrieval practice effect in real-life classrooms. This agenda has the potential to enhance our understanding of individual differences in the retrieval practice effect and to facilitate better-informed decisions by instructors and learners in educational contexts.

Acknowledgments

The ideas in this study are derived from one chapter included in the first author's doctoral dissertation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Data availability statement

Data availability is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Marcos Felipe Rodrigues de Lima http://orcid.org/0000-0002-5922-2543
Luciano Grüdtner Buratto http://orcid.org/0000-0002-7003-7824

References

Abel, M., & Roediger, H. L., III (2017). Comparing the testing effect under blocked and mixed practice: The mnemonic benefits of retrieval practice are not affected by practice format. *Memory & Cognition*, *45*(1), 81–92. https://doi.org/10.3758/s13421-016-0641-8

Agarwal, P. K., & Bain, P. M. (2019). *Powerful teaching: Unleash the science of learning*. Jossey-Bass. https://doi.org/10.1002/9781119549031

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L., III (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory (Hove, England)*, *25* (6), 764–771. https://doi.org/10.1080/09658211.2016.1220579

Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning & Teaching*, *20*(1), 21–39. https://doi.org/10.1177/14757257220973494

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.

Borsboom, D., Kievit, R. A., Cervone, D., & Hood, S. B. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudhary (Eds.), *Dynamics process methodology in the social and developmental sciences* (pp. 67–97). Springer. https://doi.org/10.1007/978-0-387-95922-1

Brewer, G., Robey, A., & Unsworth, N. (2021). Discrepant findings on the relation between episodic memory and retrieval practice: The impact of analysis decisions. *Journal of Memory and Language*, *116*, Article 104185. https://doi.org/10.1016/j.jml.2020.104185

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, *66*(3), 407–415. https://doi.org/10.1016/j.jml.2011.12.009

Brown, P. C., Roediger, H. L., III, & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Belknap Press of Harvard University Press.

Buchin, Z. L., & Mulligan, N. W. (2022). Retrieval-based learning and prior knowledge. *Journal of Educational Psychology*, *115* (1), 22–35. https://doi.org/10.1037/edu0000773

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35* (6), 1563–1569. https://doi.org/10.1037/a0017021

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*(6), 760–771. https://doi.org/10.1002/acp.1507

Cavendish, B. A., Lima, M. F. R., Perícoli, L., & Buratto, L. G. (2022). Effects of combining retrieval practice and tDCS over long-term memory: A randomized controlled trial. *Brain and Cognition*, *156*, Article 105807. https://doi.org/10.1016/j.bandc.2021.105807

Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition*, *2*(2), 95–100. https://doi.org/10.1016/j.jarmac.2013.04.001

Cogliano, M., Kardash, C. M., & Bernacki, M. L. (2019). The effects of retrieval practice and prior topic knowledge on test performance and confidence judgments. *Contemporary Educational Psychology*, *56*, 117–129. https://doi.org/10.1016/j.cedpsych.2018.12.001

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

Crutcher, R. J., & Ericsson, K. A. (2000). The role of mediators in memory retrieval as a function of practice: Controlled mediation to direct access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1297–1317. https://doi.org/10.1037/0278-7393.26.5.1297

Diana, R., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*(1), 1–21. https://doi.org/10.3758/BF03193807

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Glaser, J., & Richter, T. (2023). The testing effect in the lecture Hall: Does it depend on learner prerequisites? *Psychology Learning & Teaching*, *22*(2), 159–178. https://doi.org/10.1177/14757257221136660

Gupta, M. W., Pan, S. C., & Rickard, T. C. (2024). Interaction between the testing and forward testing effects in the case of cued-recall: Implications for theory, individual difference studies, and application. *Journal of Memory and Language*, *134*, Article 104476. https://doi.org/10.1016/j.jml.2023.104476

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, *143*(2), 575–596. https://doi.org/10.1037/a0033715

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486

Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2015). Test-enhanced learning in third-grade children. *Educational Psychology*, *35*(4), 513–521. https://doi.org/10.1080/01443410.2014.963030

Jonsson, B., Wiklund-Hörnqvist, C., Nyroos, M., & Börjesson, A. (2014). Self-reported memory strategies and their relationship to immediate and delayed text recall and working memory capacity. *Education Inquiry*, *5*(3), 22850–22404. https://doi.org/10.3402/edui.v5.22850

Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology*, *113*(5), 972–985. https://doi.org/10.1037/edu0000627

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

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

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 61, pp. 237–284). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Kole, J. A., & Healy, A. F. (2013). Is retrieval mediated after repeated testing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 462–472. https://doi.org/10.1037/a0028880

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787–1794. https://doi.org/10.1037/xlm0000012

Lima, M. F. R., & Buratto, L. G. (2023). The test–retest reliability of the retrieval practice effect. *Quarterly Journal of Experimental Psychology*, *76*(9), 2028–2036. https://doi.org/10.1177/17470218221141586

Lima, M. F. R., Cavendish, B. A., Deus, J. S., & Buratto, L. G. (2020). Retrieval practice in memory- and language-impaired populations: A systematic review. *Archives of Clinical Neuropsychology*, *35*(7), 1078–1093. https://doi.org/10.1093/arclin/acaa035

Liu, X. L., Tan, D. H., & Reder, L. M. (2018). The two processes underlying the testing effect– Evidence from event-related potentials (ERPs). *Neuropsychologia*, *112*, 77–85. https://doi.org/10.1016/j.neuropsychologia.2018.02.022

Logie, R. (2018). Human cognition: Common principles and individual variation. *Journal of Applied Research in Memory and Cognition*, *7*(4), 471–486. https://doi.org/10.1016/j.jarmac.2018.08.001

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305–321. https://doi.org/10.3758/BF03213295

McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology*, *72*(23), 609–633. https://doi.org/10.1146/annurev-psych-010419-051019

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item

difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Moreira, B. F. T., Pinto, T. S. S., Justi, F. R. R., & Jaeger, A. (2019). Retrieval practice improves learning in children with diverse visual word recognition skills. *Memory (Hove, England)*, *27*(10), 1423–1437. https://doi.org/10.1080/09658211.2019.1668017

Mulligan, N. W., Buchin, Z. L., & Zhang, A. L. (2022). The testing effect with free recall: Organization, attention, and order effects. *Journal of Memory and Language, 125*, Article 104333. https://doi.org/10.1016/j.jml.2022.104333

Nunally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill, Inc.

Pan, S. C., Pashler, H., Potter, Z. E., & Rickard, T. C. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language*, *83*, 53–61. https://doi.org/10.1016/j.jml.2015.04.001

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 287–297. https://doi.org/10.1037/a0021801

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review*, *127*(1), 1–46. https://doi.org/10.1037/rev0000161

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330* (6002), 335–335. https://doi.org/10.1126/science.1191465

Rawson, K. A., & Zamary, A. (2019). Why is free recall practice more effective than recognition practice for enhancing memory? Evaluating the relational processing hypothesis. *Journal of Memory and Language*, *105*, 141–152. https://doi.org/10.1016/j.jml.2019.01.002

Rickard, T. C. (2020). Extension of the dual-memory model of test-enhanced learning to distributions and individual differences. *Psychonomic Bulletin & Review*, *27*(4), 783–790. https://doi.org/10.3758/s13423-020-01734-7

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, *25*(3), 847–869. https://doi.org/10.3758/s13423-017-1298-4

Robey, A. (2019). The benefits of testing: Individual differences based on student factors. *Journal of Memory and Language*, *108*, Article 104029. https://doi.org/10.1016/j.jml.2019.104029

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., III, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 242–248. https://doi.org/10.1016/j.jarmac.2012.09.002

Roediger, H. L., III, & Yamashiro, J. K. (2020). Evaluating experimental research. In R. J. Sternberg & D. F. Halpern (Eds.), *Critical thinking in psychology* (pp. 249–279). Cambridge University Press. https://doi.org/10.1017/9781108684354.012

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Spielberger, C. D. (2010). Test anxiety inventory. In I. B. Weiner & W. E. Craighead (Eds.), *The corsini encyclopedia of psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470479216

Stenlund, T., Jönsson, F. U., & Jonsson, B. (2017). Group discussions and test-enhanced learning: Individual learning outcomes and personality characteristics. *Educational Psychology*, *37*(2), 145–156. https://doi.org/10.1080/01443410.2016.1143087

Sumowski, J. F., Chiaravalloti, N., & DeLuca, J. (2010). Retrieval practice improves memory in multiple sclerosis: Clinical application of the testing effect. *Neuropsychology*, *24*(2), 267–272. https://doi.org/10.1037/a0017533

Trumbo, M. C. S., McDaniel, M. A., Hodge, G. K., Jones, A. P., Matzen, L. E., Kittinger, L. I., Kittinger, R. S., & Clark, V. P. (2021). Is the testing effect ready to be put to work? Evidence from the laboratory to the classroom. *Translational Issues in Psychological Science*, *7*(3), 332–355. https://doi.org/10.1037/tps0000292

Tse, C.-S., Balota, D. A., & Roediger, H. L., III (2010). The benefits and costs of repeated testing on the learning of face–name pairs in healthy older adults. *Psychology and Aging*, *25*(4), 833–845. https://doi.org/10.1037/a0019933

Tse, C.-S., Chan, M. H.-M., Tse, W.-S., & Wong, S. W.-H. (2019). Can the testing effect for general knowledge facts be influenced by distraction due to divided attention or experimentally induced anxious mood? *Frontiers in Psychology*, *10*, Article 969. https://doi.org/10.3389/fpsyg.2019.00969

Tse, C.-S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied*, *18*(3), 253–264. https://doi.org/10.1037/a0029190

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*(1), 79–139. https://doi.org/10.1037/bul0000176

Unsworth, N., Miller, A. L., & Robison, M. K. (2021). Are individual differences in attention control related to working memory capacity? A latent variable mega-analysis. *Journal of Experimental Psychology: General*, *150*(7), 1332–1357. https://doi.org/10.1037/xge0001000

Walrath, R., Willis, J. O., Dumont, R., & Kaufman, A. S. (2020). Factor-analytic models of intelligence. In R. J. Sternberg (Ed.), *The Cambridge handbook of intelligence* (2nd ed., pp. 75–98). Cambridge University Press.

Wenzel, K., & Reinhard, M.-A. (2019). Relatively unintelligent individuals do not benefit from intentionally hindered learning: The role of desirable difficulties. *Intelligence*, *77*, Article 101405. https://doi.org/10.1016/j.intell.2019.101405

Wiklund-Hörnqvist, C., Stillesjö, S., Andersson, M., Jonsson, B., & Nyberg, L. (2022). Retrieval practice is effective regardless of self-reported need for cognition - Behavioral and brain imaging evidence. *Frontiers in Psychology*, *12*, Article 797395. https://doi.org/10.3389/fpsyg.2021.797395

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*, *26*(4), 724–738. https://doi.org/10.1037/xap0000278

Zheng, Y., Sun, P., & Liu, X. (2023). Retrieval practice is costly and is beneficial only when working memory capacity is abundant. *NPJ Science of Learning*, *8*(1), Article 8. https://doi.org/10.1038/s41539-023-00159-w