

PLAT 20(1) 2021: Enhancing Student Learning in Research and Educational Practice: The Power of Retrieval Practice and Feedback

Psychology Learning & Teaching

2021, Vol. 20(1) 1–20

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1475725720976462

journals.sagepub.com/home/plj

Abstract

Students and instructors are looking for effective study and instructional strategies that enhance student achievement across a range of content and conditions. The current Special Issue features seven articles and one report, which used varied methodologies to investigate the benefits of practising retrieval and providing feedback for learning. This editorial serves as an introduction and conceptual framework for these papers. Consistent with trends in the broader literature, the research in this Special Issue goes beyond asking whether retrieval practice and feedback enhance learning, but rather, when, for whom, and under what conditions. The first set of articles examined the benefits of retrieval practice compared to restudy (i.e., the testing effect) and various moderators of the testing effect, including participants' cognitive and personality characteristics (Bertilsson et al., 2021) as well as the timing of the practice test and sleep (Kroneisen & Kuepper-Tetzel, 2021). The second set of articles examined the efficacy of different types of feedback, including complex versus simple feedback (Enders et al., 2021; Pieper et al., 2021) and positively or negatively valenced feedback (Jones et al., 2021). Finally, the third set of articles to this Special Issue examined practical considerations of implementing both retrieval practice and feedback with educationally relevant materials and contexts. Some of the practical issues examined included when students should search the web to look for answers to practice problems (Giebl et al., 2021), whether review quizzes should be required and contribute to students' final grades (den Boer et al., 2021), and how digital learning environments should be designed to teach students to use effective study strategies such as retrieval practice (Endres et al., 2021). In short, retrieval and feedback practices are effective and robust tools to enhance learning and teaching, and the papers in the current Special Issue provide insight into ways for students and teachers to implement these strategies.

Keywords

Education, testing effect, feedback, formative assessment, knowledge acquisition

Teachers at all levels of education, including readers of *Psychology Learning and Teaching*, aim to provide high-quality instruction and educational experiences to their students. From very different research backgrounds, researchers are eager to contribute to a broad empirical basis for the decisions that teachers must make in order to enhance student learning and achievement. We are very fortunate to highlight articles in the current Special Issue representing various empirical approaches from education, psychology, and cognitive neuroscience to improve learning and teaching. Despite the diversity in their methodologies and contexts, the contributions in this Special Issue have one key feature in common: they examine the learning benefits brought about by practising retrieving one's knowledge and receiving feedback. Given that we, psychology instructors, teach our students the value of evidence-based practice, the way that we teach should be grounded in empirical evidence as well. The articles presented in this Special Issue contribute to a broad literature on the science of learning and teaching by providing new evidence regarding how to implement retrieval practice and feedback in educational settings in a manner that maximizes student learning.

The Testing Effect

The testing effect refers to the finding that tests are not merely opportunities to assess one's learning but are potent learning opportunities themselves. More than a century of research has revealed that retrieving previously studied information enhances subsequent memory for that information (Abbott, 1909; for reviews, see Roediger & Butler, 2011; Roediger & Karpicke, 2006; Rowland, 2014). For example, Roediger and Karpicke (2006) had participants read a series of brief passages (e.g., 'The Sun', 'Sea Otters') and then reread the passages or try to retrieve as much information from the passages as possible. On subsequent test days or weeks later, participants recalled significantly more information from the passages if they had practiced retrieval rather than rereading. Indeed, a meta-analysis of lab experiments determined that, on average, information is 2.5 times more likely to be recalled on subsequent tests if it was reviewed through retrieval practice compared to restudying (Rowland, 2014). The testing effect is not limited to the lab, though. Other meta-analyses have found similar sized benefits of incorporating retrieval-practice activities (e.g., flashcards, multiple-choice quizzes, short answer review questions) into classes with different aged students and different types of course content (Adesope et al., 2017; Schwieren et al., 2017; Sotola & Crede, 2020). For example, Sotola and Crede (2020) found that, across 52 independent classroom studies, students were 2.5 times more likely to pass a course if their instructor incorporated frequent low-stakes quizzing than if the instructor did not. Interestingly, the benefits of frequent low-stakes quizzing were higher in psychology courses than other types of courses.

Despite the strength of the testing effect and the myriad of contexts in which it has been observed, many moderators have been identified. Thus, research on the testing effect is no longer asking whether retrieval practice enhances learning, but rather, for whom does retrieval practice enhance learning, for which types of content, and under what conditions. The studies reported in this Special Issue contribute to these testing effect research priorities by identifying potential moderators and characterizing their effects on learning with a variety of retrieval-practice based tasks and materials (true-false quizzes, computer programming tasks, written reflections, formative vs. summative assessments, foreign-language translations, probabilistic learning) in diverse settings (lab, in-person materials science course, online psychology course, pre-service teacher course).

Moderators of the Testing Effect: An Overview

Potential moderators of the testing effect can broadly be classified as pertaining to learner characteristics (e.g., working memory capacity), the materials (e.g., simple definitions vs. complex math problems), the retrieval task (e.g., free recall vs. multiple choice), and the learning context (e.g., the timing and use of practice tests in a course; Figure 1). The lines connecting the four classes of moderators signify the interactions between moderators. For example, Butler and colleagues (2013) examined whether the effect of the retrieval task (e.g., elaborate feedback versus correct-answer feedback) depended upon the materials (e.g., facts versus concepts). The examples provided in Figure 1 are meant to illustrate a classification of potential moderators of the testing effect but are not comprehensive of all the potential moderators that are worth examining. Furthermore, not all of the potential moderators listed in Figure 1 have necessarily been investigated or supported by empirical research. A comprehensive review of the moderators of the testing effect is outside the scope of this Introduction. Our aim with Figure 1 is to provide a classification of potential moderators as a way to organize existing research (for a similar approach, see Dunlosky et al., 2013), inspire future research, and facilitate broader conclusions about moderators of the testing effect and interactions between moderators.

The studies reported in this Special Issue investigated all four types of potential moderators (indicated by an * in Figure 1). In terms of features of the retrieval task, multiple articles compared different types of feedback, including feedback versus no feedback (Pieper et al., 2021), correct-answer feedback with explanations versus correct-answer only feedback (Enders et al., 2021), and positive versus negative feedback (Jones et al., 2021). Furthermore, three studies in the current Special Issue examined characteristics of the learning context. Den Boer and colleagues examined whether the benefits of retrieval practice varied depending upon whether the retrieval practice was required for a course. Others in this Special Issue examined the optimal timing of the retrieval-practice task. For optimal learning, should a retrieval-practice activity be administered immediately after learning or two hours later (Kroneisen & Kuepper-Tetzl, 2021)? Can a retrieval task enhance learning if it is administered as a pre-test (i.e., before all of the relevant content has been taught; Giebl et al., 2021)? Finally, one study in the Special Issue also examined a feature of the materials: Enders and colleagues examined item difficulty (i.e., whether the student could answer the question correctly; see Table 1).

Understanding the features of the materials, retrieval task, and learning context that moderate the testing effect has important practical implications. If researchers can identify, say, the optimal type of retrieval task, timing of the retrieval task, and type of feedback to provide in a given context, then these insights can be shared with instructors to maximize student learning. However, just as important as the features of the retrieval task are the characteristics of the student. Predicting who will benefit from retrieval practice is key, if retrieval practice enhances learning for only a subset of students (e.g., Carpenter et al., 2016; Minear et al., 2018).

Many of the articles in this Special Issue also examined characteristics of the learner, including grit, working memory capacity, need for cognition, self-efficacy, anxiety, and sleep (see Table 1). These characteristics can be measured and observed about the learner, regardless of the retrieval-practice task or content with which they are engaging. However, critically, many characteristics of a learner are not static traits, but vary depending upon the materials and learning task conditions. For example, a learner does not always have low

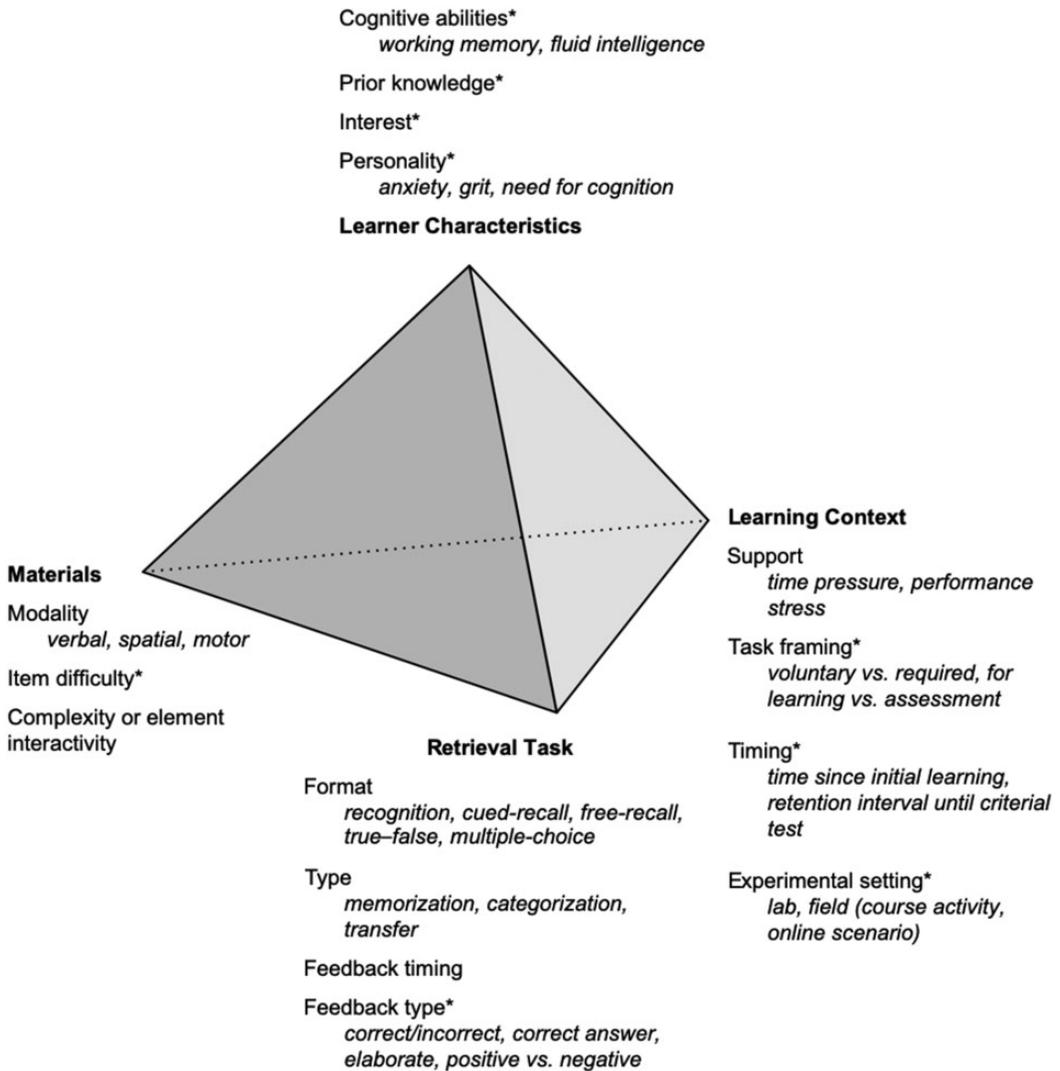


Figure 1. Potential moderators of the testing effect can broadly be classified as pertaining to learner characteristics, the materials, the retrieval task, and the learning context. The examples provided are illustrative and not comprehensive. The interactions among any of the classes of moderators may be examined, too, as indicated by the lines connecting the four classes of moderators. This model was inspired by Jenkins' (1979) tetrahedral model of memory experiments. * Indicates that the potential moderator was investigated in at least one of the studies in this Special Issue.

levels of interest or knowledge, but rather, their interest and knowledge varies across courses and tasks. Den Boer and colleagues examined students' interest and study time with content from a Materials Science course. Similarly, Giebl and colleagues examined students' prior content knowledge related to computer programming.

An important and timely testing effect research question is how these moderators interact. For example, Minear and colleagues (2018) examined how fluid intelligence (gF)

Table 1. Summary of Papers in the Special Issue.

Authors	Learning Context	Materials	Retrieval Task: Complexity	Testing Effect Moderator(s) Category: description	Timescale of the Study
Bertilsson et al., 2021	Lab	Foreign-language translations	Memorization	Learner characteristics: working memory capacity, grit, and need for cognition	Weeks
Kronenisen & Kuepper-Tetzel, 2021	Lab	Foreign-language translations	Memorization	Retrieval task: delay between initial study and retrieval practice, and sleep before final test	Hours to days
Giebl et al., 2021	Lab	Computer programming	Transfer	Learner characteristics: sleep	Minutes
Jones et al. 2021	Lab	Probabilistic learning with arbitrary pictures	problem-solving Probabilistic learning	Learner characteristics: prior knowledge	Minutes
Enders et al., 2021	Field: online university psychology course	Psychology facts	Memorization	Retrieval task: positive vs. negative feedback Learner characteristics: self-reported anxiety	Days to weeks
Pieper et al. 2021	Field: pre-service teachers in training	Written reflections of teaching practice	Transfer	Retrieval task: correct answer vs. elaborate feedback	Weeks
den Boer et al., 2021	Field: university materials-science course	Engineering facts, calculations, and examples	Memorization calculation transfer	Retrieval (and reflection) task: elaborate feedback vs. no feedback Learning context: retrieval practice for assessment vs. learning	Months
Endres et al., 2021	Field: online university course	Learning strategies	Transfer	Learner characteristics: prior knowledge	Months

interacted with item difficulty to moderate the testing effect. High *gF* participants showed a larger testing effect for difficult compared to easy items; in contrast, low *gF* participants showed a larger testing effect for easy compared to difficult items. Articles reported in this Special Issue also examined the interaction between different classes of moderators. For example, Kroneisen and Kuepper-Tetzel asked whether the effect of the timing of the task (immediately after learning vs. two hours later) depended upon when one would sleep (e.g., immediately after retrieval practice vs. several hours after retrieval practice); Jones and colleagues examined whether the valence of feedback (positive vs. negative) depended upon the learner's level of anxiety; and Giebl and colleagues examined whether the benefit of the retrieval task (pre-testing vs. no pre-testing) depended on prior knowledge.

Lab-Based Investigation of Moderators in the Special Issue

One approach to investigating moderators of the testing effect is to conduct tightly controlled lab-based experiments. The benefits of a lab-based approach to researching moderators is that it allows researchers to manipulate or observe specific potential moderators, while controlling for other confounding variables. For example, in a study of the testing effect in a university biology classroom, Carpenter and colleagues (2016) found that higher performing students benefited from retrieval practice more than copying down information, but middle and lower performing students benefited more from copying than retrieval practice. Although prior course performance seemingly moderated the testing effect, course performance could have been a proxy for a range of other individual difference variables besides prior knowledge (e.g., working memory capacity, interest in the course material, average amount of sleep). Therefore, lab-based studies offer an important complement to the more ecologically valid classroom studies because they allow for more precise examination of individual moderators. Despite increased experimental control, lab-based studies can still closely reflect the type of learning that students do in genuine classes. Indeed, studies in this Special Issue have taken such an approach, conducting lab-based experiments of the testing effect with educationally relevant materials such as foreign-language transitions (Bertilsson et al., 2021; Kroneisen & Kuepper-Tetzel, 2021) and computer programming (Giebl et al., 2021; see Table 1).

Bertilsson and colleagues examined individual differences in cognitive and personality traits, specifically working memory capacity, grit (i.e., perseverance and passion for long-term goals), and need for cognition (the tendency for an individual to engage in and enjoy thinking), which have all been shown to relate to academic achievement (e.g., Cowan, 2014; Duckworth & Quinn, 2009; Sadowski & Gulgoz, 1992). Participants learned Swahili–Swedish translations and then reviewed the translations through restudy or retrieval practice with corrective feedback. On a final test of the Swahili words five minutes, one week, and four weeks after review, participants correctly recalled significantly more Swedish translations in the retrieval practice than the restudy condition. Thus, a testing effect emerged. However, the size of the testing effect was not moderated by the participant's working memory capacity, grit, or need for cognition.

Bertilsson and colleagues' finding contributes to the mixed literature on the relationship between working memory capacity and the testing effect, which has revealed a positive association (Tse & Pu, 2012), a negative association (Agarwal et al., 2017), and no association (Bertilsson et al., 2017; Brewer & Unsworth, 2012; Minear et al., 2018; Wiklund-Hörnqvist et al., 2014) between working memory capacity and the size of the testing

effect. Thus, future research should aim to identify the circumstances under which working memory is related to the size of the testing effect, which will have both practical and theoretical implications. We encourage future research on working memory capacity and the testing effect to closely consider measurement issues. The degree to which one can estimate the effect of working memory capacity on the testing effect is limited by the reliability of the working memory estimate. Therefore, future research should aim to collect a sample with a large range of working memory capacities, use multiple measures of working memory capacity, and treat these measures as indicators of working memory capacity as a latent variable. Through structural equation modelling, one could then determine the relation of working memory capacity and the testing effect, free of the measurement error associated with estimating working memory (e.g., Engle & Kane, 2004; Kane et al., 2005). This analytic approach should allow for more precision, and perhaps consistency in the literature, in understanding how working memory moderates the size of the testing effect.

In addition to contributing to the relatively nascent literature on cognitive differences and the testing effect, Bertilsson and colleagues' study is among the first to examine personality differences and the testing effect. Consistent with prior research, their study found that grit and need for cognition did not moderate the magnitude of the testing effect. However, examining the association between personality and the size of testing effect may be a worthwhile direction for future research since it has been repeatedly shown that personality traits can predict academic performance above and beyond measures of cognitive functioning (e.g., O'Connor & Paunonen, 2007). Until new findings emerge, though, students and teachers can feel relatively confident that retrieval practice will benefit learning – at least of vocabulary definitions or translations – regardless of various student traits (see Table 2).

Beyond differences in learners' cognitive abilities and personalities, the testing effect may be moderated by learners' lifestyles, such as their study and sleep schedules. Kroneisen and Kuepper-Tetzel examined the degree to which learning from a practice test depended upon when participants took the practice test and whether they slept during retention. Participants studied Polish–German vocabulary translations, reviewed the translations through a practice test without feedback, and then took a final test 12 hours later. Participants took the practice test either immediately after initial learning or two hours later; participants also either stayed awake for the 12 hours between initial learning and the final test or slept during those 12 hours. When controlling for practice test performance, both timing of the practice test and sleep affected final test performance. Participants remembered more on the final test following an immediate practice test and if they slept during the delay before the final test. Kroneisen and Kuepper-Tetzel's results highlight the importance of considering possible interactions between moderators. The authors found that the beneficial effects of sleep on final test performance were larger in the immediate than the delayed practice test condition. Thus, this study in the Special Issue supports the important role of sleep in learning, memory, and academic performance (e.g., Curcio et al., 2006; Diekelmann & Bjorn, 2010; Lowe et al., 2017) and initiates an interesting new line of research focused on how sleep and study schedules interact to affect learning from tests. Thus, students should focus not just on how they study, but when they study, and prioritize incorporating sleep before an exam (see Table 2).

More broadly, Kroneisen and Kuepper-Tetzel's study is an example of an important direction for future applied research on the testing effect. Testing has been shown to enhance learning, but there are many other beneficial study techniques (Soderstrom et al., 2015). The most apparent combination is between retrieval practice and spaced learning. Many of the

Table 2. Practical Advice for Students and Teachers from Papers in the Special Issue.

Authors	Advice for Students	Advice for Teachers
Bertilsson et al., 2021	Practice testing yourself, even when it feels difficult.	Create opportunities for retrieval practice, even if your students will have difficulties at times.
Kroneisen & Kuepper-Tetzl, 2021	If you do not have time to sleep after learning, delay retrieval practice after initial learning to enhance long-term memory.	Emphasize the value of sleep and delaying retrieval practice but acknowledge that the value of delaying retrieval practice may depend on sleep.
Giebl et al., 2021	Try to solve a problem before you look up the answer.	Use questions or problems to introduce a lesson, not just for practice after the lesson.
Jones et al., 2021	See the benefits of negative feedback: as an anxious person you might learn more than from positive feedback.	Provide both positive and negative feedback in order to support different students.
Enders et al., 2021	Seek explanations for questions you answered incorrectly.	Provide elaborate feedback, even for simple true–false questions.
Pieper et al., 2021	Seek expert feedback on your self-reflections of your performance on complex tasks.	Provide elaborate feedback on performance, even on self-reflections.
den Boer et al., 2021	Test yourself on the course materials before the next lecture in order to stay motivated and assess your current level of knowledge.	Incorporate low-stakes tests (rather than high-stakes tests) throughout the term to enhance students' motivation and long-term learning.
Endres et al., 2021	Take part in an online learning-strategy module when transitioning to university.	Implement practice schedules to challenge students with low, medium, and high prior learning-strategy knowledge at retrieval.

Note: The advice for students and teachers in this table is grounded in the results of seven articles and one report in this Special Issue. However, some of this advice is based on results from the laboratory, whereas other advice is based on results from more ecologically valid field studies. More research is needed to systematically evaluate this advice with a variety of course content and students.

retrieval-practice studies include spaced retrieval in the experimental designs. Future research should examine whether the benefits of testing can be combined with other effective learning strategies to produce even larger learning gains (e.g., Kubik et al., 2020; Miyatsu & McDaniel, 2019). Relevant to Kroneisen and Kuepper-Tetzl's study, Bäuml and colleagues (2014) compared retrieval practice to restudying when combined with sleep or wakefulness. Although sleep has repeatedly been shown to enhance memory (Ashworth et al., 2013; Diekelmann & Born, 2010), Bäuml and colleagues (2014) found that sleep enhanced memory after restudying, but only had a minimal impact memory after retrieval practice. That is, adding another learning benefit (brought about by sleep) to testing conferred little additional benefit to memory (but see Abel et al., 2019; Mazza et al., 2016). More research is needed to better understand when retrieval practice should be combined with other productive techniques to optimize students' learning.

Taken together, the lab-based studies of moderators of the testing effect in this Special Issue (see Table 1) contribute to the growing body of research on moderators of the testing

effect. However, we emphasize that more research is needed before we can make empirically sound recommendations to teachers and students regarding who should engage in retrieval practice and when. Furthermore, we encourage using these studies as a guide for future investigations of the testing effect. Specifically, future research should continue to test moderators of the testing effect, but also examine potential interactions between moderators as well as the combined effect of testing and other effective study strategies. As the studies described in the next section demonstrate, the interaction between feedback and the testing effect is a particularly rich area for future research.

Maximizing the Benefits of Formative Testing: Providing Feedback

One powerful tool to maximize learning is to provide feedback to students (Butler & Woodward, 2018; Wisniewski et al., 2020). Feedback typically contains information about one's current task performance or level of understanding and can in turn enhance learning (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Wisniewski et al., 2020). Many of the articles in this Special Issue examined the benefits of feedback in combination with retrieval practice (den Boer et al., 2021; Enders et al., 2021; Jones et al., 2021; Pieper et al., 2021) or included feedback to enhance learning, even when it was not the main objective of the study (den Boer et al., 2021; Endres et al., 2021).

Feedback as an Effective Complement to Test-based Learning

As reported in the previous section, retrieval practice robustly enhances long-term learning, and students and instructors can capitalize on this testing benefit through various forms of formative assessment. However, retrieval practice can also have negative consequences. For example, multiple-choice practice tests may induce errors and misconceptions from exposure to the incorrect alternative answer choices (cf. Butler & Woodward, 2018). However, prior meta-analytic studies revealed that feedback ameliorated these potential negative side effects of testing and magnified its benefits in both lab-based studies (Rowland et al., 2014) and more applied classroom studies (Phelps, 2012, 2019; Schwieren et al., 2017; Wisniewski et al., 2020). The benefit of feedback pertains to myriad educational levels, subjects (including psychology, see Schwieren et al., 2017), feedback timing (immediate vs. delayed), and task materials (e.g., simple vs. complex; e.g., Van der Kleij et al., 2015). When considering 435 studies, a recent meta-analysis (Wisniewski et al., 2020) revealed a medium positive effect size of feedback ($d = 0.48$) on student learning. To conclude, it is generally advisable to provide feedback when students take formative tests, particularly when feedback can be conveniently delivered digitally (Enders et al., 2021) – a requirement for remote education during the current Covid-19 pandemic.

What is the Optimal Type of Feedback? Considering Levels of Feedback and Moderators

A recent meta-analytic study revealed a huge variability in the effectiveness of feedback (Wisniewski et al., 2020). It often fosters students' learning and achievement, for example, when it is geared towards task performance (Enders et al., 2021; for an overview, see Kleij et al., 2015) or self-regulation (Pieper et al., 2021). However, feedback can also be detrimental to learning, for example, when it involves personal evaluations of abilities (e.g., praising the students' mathematical talent; Brooks et al., 2019). As a result, students or children may rather attribute their success to their stable abilities, and eventually become

less motivated to increase effort or to improve specific study strategies, specifically when facing negative feedback (Gundersson et al., 2017).

As feedback can both engender positive and negative effects on learning and can be implemented in various forms, feedback must be treated as a complex construct (cf. Wisniewski et al., 2020). Feedback can be viewed as providing information on various levels (Hattie & Timperley, 2007):

- task performance (the *task level*; Enders et al., 2021; Jones et al., 2021), such as ‘This response was correct.’
- students’ strategies on how to perform or solve the task (*processing level*, see Pieper et al., 2021), such as ‘This strategy worked well for this type of task.’
- students’ habits to monitor, evaluate, and regulate task strategies to attain the learning goal (*self-regulation level*, see Pieper et al., 2021), such as ‘You can more often monitor and evaluate whether you have made progress toward your original learning goals, considering the remaining time for the task.’
- students’ abilities (*self-level*), such as ‘You are mathematically gifted!’

Being such a multifaceted phenomenon, current research no longer examines the global benefits of feedback but rather addresses specific feedback levels. The task level has been relatively well-studied but the self-regulation level has hardly been investigated (Brooks et al., 2019; Butler & Woodward, 2018). Several studies reported in the Special Issue contributed in this regard by examining feedback on the task level (Enders et al., 2021; Jones et al., 2021) as well as the levels of processing and self-regulation (Pieper et al., 2021).

Across and within these four feedback levels, one major research question is to understand the heterogeneity of feedback benefits, and to identify the critical moderators. For example, it has been shown that feedback is more efficient when it is delayed (compared to immediate feedback, Butler et al., 2007; Rowland, 2014), and when the outcome of interest is cognitive (e.g., learning performance) compared to motivational (e.g., intrinsic motivation; Wisniewski et al., 2020). Beside these and other moderators (see Kluger & DeNisi, 1996; Shute, 2008; Van der Kleij et al., 2015; Wisniewski et al., 2020), it is critical to examine the optimal design of feedback – aspects that are inherent to the specific feedback procedure.

How can feedback most effectively improve test-based learning? Several studies in this Special Issue investigated two major components: the valence of the feedback (i.e., negative versus positive feedback) and the complexity of the feedback message (right/wrong feedback, corrective-answer feedback vs. correct-answer feedback with explanations; Butler et al., 2013). Surprisingly, there is sparse research on the relative benefits of these different types of feedback on the task level and the self-regulation level (cf., Brooks et al., 2019; Butler & Woodward, 2018; Wisniewski et al., 2020). To this end, the studies reported in the Special Issue make valuable contributions to the field, covering both lab-based investigations and applied field studies using digital formats.

The article by Jones and colleagues examined the valence of feedback in a lab-based study at the task level. They implemented a simple probabilistic learning task with anxious versus non-anxious learners and assessed learners’ reactions to positive versus negative feedback through a specific component of the event-related potential (ERP measured via the electroencephalogram). Strikingly, the results revealed that highly anxious people expected negative feedback, as indicated by reduced ERP change, and predominantly learned from

negative feedback, whereas non-anxious learners expected and profited more from positive feedback. This contribution highlights the importance of learner characteristics as potential moderators of the effects of different feedback types. Future research is encouraged to more systematically include interindividual difference variables to unravel potential aptitude–treatment interactions that help to understand why positive versus negative feedback may be beneficial in one population but not in the other. Furthermore, this study of the Special Issue suggests an innovative approach to instructional research that more systematically includes brain-physiological indicators to further inform both underlying mechanisms of feedback processing as well as practical guidelines for learning and teaching (see also Jonsson et al., 2020; Wiklund-Hörnqvist et al., 2017).

In addition to the valence of feedback, current research attempts to enhance the efficacy of feedback by increasing the complexity of the feedback message on the task level (Butler et al., 2013; Van der Kleij et al., 2015). The article by Enders and colleagues takes up this timely issue and examines whether elaborate (correct-answer feedback with explanation) feedback versus corrective (correct-answer only) feedback benefits fact-based knowledge acquisition above and beyond potential direct effects of formative testing in a real-world digital educational setting. More specifically, students in a psychology course at a distance university answered verbal statements with true–false questions, and immediately received corrective or elaborate feedback. On a final test after a self-regulated delay, students profited more from elaborate feedback than corrective feedback, particularly for questions that they had initially answered incorrectly. The study by Enders et al. exemplifies an important line of research aimed at maximizing the benefits of retrieval practice and feedback in formative testing settings. It is notable that students increasingly acquire fact-based knowledge through quizzing in addition or even as alternative to attending the lectures and reading the course materials (Marsh et al., 2007; Roediger & Marsh, 2005). Future research needs to generalize the positive effects of elaborate feedback with different materials and outcome variables (learning transfer, cf. Butler et al., 2013).

Beyond providing diagnostic information on task performance, feedback can also include hints on how to solve the task at hand and suggestions to improve the self-regulated use of various task strategies. Pieper and colleagues (2021) adopted this broader perspective by examining whether student teachers can improve their reflection skills based on expert feedback via a digital platform. In a field experiment, student teachers wrote two reflective journals about their teaching practices, upon which they either obtained high-information feedback on their journal entry or not. For example, students received elaborate feedback about their reflections on the goals of the lessons they taught, which aspects of their teaching were sufficiently addressed in their reflections, on the analysis of their teaching quality, and on their plan to improve their lessons going forward. The results showed that high-information feedback on reflective journaling improved the quality of the student teachers' later reflections as well as their conceptual knowledge about reflective journaling. Future research should examine how helping student teachers improve their self-reflections can translate into improved lessons and student learning. Future research should also consider interactions with other moderators more systematically such as the timing of feedback and learners' prior expertise (cf. Nückles et al., 2020; Roelle et al., 2011). This article in the Special Issue exemplifies an important yet underexplored line of research; it highlights the effectiveness of incorporating feedback on processing and self-regulation, not just task performance, in the context of journal-writing-to-learn (cf. Nückles et al., 2020). It is often important for learners to receive feedback on the ultimate level of task performance and

learning that they are aiming for as well as strategies for how to achieve that desired target state of learning, not just feedback on how their current state of learning does meet the target state. To know ‘Where am I going?’ and ‘Where to next?’ (Hattie & Timperley, 2007, p. 87) may be particularly beneficial for fostering more complex skills (e.g., reflections) in comparison to acquiring simple facts or vocabulary.

Why Does Feedback Enhance Test-based Learning?

As discussed above, feedback is a broad, multifaceted instructional strategy, including many variants, which likely affect learning through different mechanisms. Task-level feedback provides diagnostic information for the learner regarding whether the to-be-learned contents were correctly recalled. Such metacognitive information help learners to reduce the discrepancy between actual and desired states of learning and to enhance subsequent restudy activities (cf. Butler & Woodward, 2018; Hattie & Timperley, 2007). For example, providing feedback helps students to correct previous errors (e.g., Pashler et al., 2005), to maintain the correct answers (Butler et al., 2008), or to focus on closing specific knowledge gaps (i.e., non-recalled learning contents) in subsequent learning opportunities. With more complex materials, feedback that encourages students to close knowledge gaps may not only lead students to correct errors related to isolated facts or details but may also induce broader conceptual change such as revising misconceptions (Corral & Carpenter, 2020). Thus, feedback after a retrieval-practice opportunity not only directly benefits learning, but also supports accurate metacognition and improved subsequent studying (McDaniel & Little, 2019; Roediger et al., 2011). Complex or high-information feedback also addresses the processing and self-regulation levels, include multiple, cyclic stages of continuous learning, assessment, and feedback processing (cf. Bangert-Drowns et al., 1991). Such a macro-theory of feedback incorporates interactions of various factors of prior knowledge and motivation, belief, and students’ feedback evaluation. Future research should systematically test different types and levels of feedback in order to understand when and why feedback enhances learning. This helps to inform and perhaps connect theories of feedback that are directed towards a specific level (cf. Butler & Woodward, 2018) or attempt to cover all levels and their interactions (cf. Bangert-Drowns et al., 1991).

Taken together, the lab-based studies and applied studies in natural settings of this Special Issue contributed to the prominent research field of feedback practices – a powerful means to boost students’ learning and achievement (Hattie & Timperley, 2007). As reported, feedback seems to be more effective the more information it contains: providing an explanation fosters additional knowledge acquisition and high-information feedback on processing and self-regulation enhances reflection skills. Note also that learner characteristics such as anxiety may play a critical role for the relative effectiveness of different feedback types as demonstrated in the effect of feedback valence. Future research is encouraged to further examine interindividual differences and employ neurophysiological measures of feedback processing. From an educational perspective, implementing feedback procedures involves trading off their efficiency with incurring time costs. Future work should more systematically consider time on task – a limited resource in educational settings – and also to compare feedback procedures within these given time restrictions (cf. Hays et al., 2010).

Applying Retrieval-based and Feedback-based Activities in Educational Contexts

Retrieval practice can be implemented efficiently in class in diverse ways such as discussion questions, quizzes, concept maps (cf. Franzis et al., 2020), or reflecting upon prior performance (Pieper et al., 2021). In lab-studies and in applied settings, the above papers of the Special Issue targeted moderators affecting when and why retrieval practice enhance learning, and how feedback can be used to maximize the impact of retrieval practice. Some papers took a broader perspective on retrieval-based and feedback-based learning in educational contexts, which is timely as learning contexts are expanding and changing due to technological advances. Research on learning and instruction must consider that standard educational contexts now extend beyond the lecture hall and include computer-based readings, lectures, and review activities that students complete on their own time with access to the internet. Web technology shapes information processing habits by making information easily available (e.g., Storm et al., 2017), which can have both positive and negative consequences. On the one hand, access to the internet enables students to bypass retrieval practice and immediately search the internet for answers, perhaps impairing learning (Giebl et al., 2021). On the other hand, web technology creates new opportunities to improve educational contexts by seamlessly incorporating frequent test-based learning opportunities into courses (Enders et al., 2021; cf. Howe et al., 2018; Nevid et al., 2020). With retrieval practice being easy to digitally implement in courses, it begs the question of how retrieval-practice opportunities should contribute to a course grade. In the current Special Issue, den Boer and colleagues provided evidence that retrieval-practice opportunities can be treated as formative assessments (rather than contributing significantly to one's final grade) and still enhance learning. Finally, with significant evidence that retrieval practice enhances learning in both lab and classroom settings, an important question is how to encourage students to use retrieval practice and other effective learning techniques in their own self-regulated studying. Endres and colleagues reported on studies leading to the design of an online tool to teach university students about effective study strategies and how to implement these strategies. Web-technology enables universities to teach a large number of students about effective study strategies at once regardless of whether they are taking courses in a traditional or hybrid format (cf. Powers et al., 2016).

Searching Memory Before Searching the Web

Many contributions of this Special Issue investigated how to support durable long-term memory via effortful retrieval practice and feedback. This research on how to work hard to memorize facts might at first sight seem peculiar and outdated given the wealth of knowledge that can be accessed via the internet. In academic settings, memory is not only useful for storing facts and concepts, but also for storing skills on how to utilize external sources such as the internet in our work. Without having background factual knowledge, we would not know which terms to search for on the internet or sources to examine in order to gain further knowledge. The hen-and-egg problem might materialize as a mouse-and-memory problem. Given the wealth of sources available, at least some accumulation of factual knowledge might be necessary before one can profit from external resources. Accordingly, Giebl and colleagues took into account the availability of easily accessible external resources as a characteristic of current learning environments. Using programming knowledge as a

test case, the authors investigated whether and which participants would profit from being nudged into retrieval practice before consulting the internet for answers. Relative to participants who immediately consulted the internet for answers to the programming problems, participants who first attempted to solve the problem independently before consulting the internet performed better on a final test that included similar problems as well as easier, more basic problems. However, attempting to retrieve from memory first before searching the web seemed more helpful for students with at least some programming knowledge compared to students with no prior programming knowledge. While prior research has identified changes in memory and potential cognitive costs of relying on the web as external memory (e.g., Marsh & Rajaram, 2019; Sparrow et al., 2011; Storm et al., 2017), Giebl and colleagues highlighted how to supplement memory-based searches with internet-based searches in order to maximize learning.

(Not) Making Retrieval Practice Part of the Exam

Across a range of materials and settings, the articles in this Special Issue found that retrieval practice enhances learning. A natural follow-up question is whether teachers should require retrieval practice or merely offer it as an optional learning opportunity. Giebl and colleagues' finding suggests that students should be required to engage in retrieval practice; with the ease of searching for information on the web, students may skip optional retrieval opportunities and thus stunt their learning. Den Boer and colleagues directly tested this issue of compulsory versus optional retrieval practice. Continuous studying (rather than cramming before the final exam) enhances learning and instructors could support such spaced studying by interspersing cumulative assessments throughout a course before the final exam (e.g., Schwieren et al., 2017; Tuckman, 1998). Cumulative compensatory assessments in which each assessment covers all of the content taught thus far can also enhance involvement in the class (Kerdijk et al., 2015). Den Boer and colleagues examined how cumulative assessments should be implemented to maximize students' performance on the cumulative assessments and the high-stakes final exam.

In line with recent work (Nevid et al., 2020), den Boer and colleagues found performance on the cumulative assessments was higher when they were a compulsory course activity. However, the question remained whether the required cumulative assessments should contribute to students' final grades in the high-stakes final exam. On a practical level, having cumulative quizzes contribute to the final grade implies additional administrative burdens. For example, it requires preparations to prevent fraud and to grant opportunities to retake a test missed due to illness. On a theoretical level, it is important to understand whether incorporating performance on the cumulative assessments into the final grade increases motivation and in turn improves learning in the course in general. Den Boer and colleagues found that making the cumulative assessments a part of students' final grades only increased motivation for the cumulative assessments and not the course in general. Furthermore, making the cumulative assessments a part of students' final grades improved students' performance on the cumulative assessments but did not enhance learning in the course in general, as evidenced by performance on the final exam. Therefore, teachers should concentrate their efforts on offering frequent opportunities for retrieval practice of course content, and students should perceive the use of quizzes as an opportunity to learn by self-testing.

Training to Retrieve and Implement Learning Strategies

The Special Issue articles that have been introduced thus far have addressed the optimal way to implement retrieval practice and feedback in both lab and classroom settings. However, an important question remains: how can students be trained to actually use these techniques in their own studying? Previous research suggests that many students rely on a mix of empirically supported learning strategies and ineffective learning strategies (Bjork et al., 2013; Dunlosky et al., 2013). Computer-based, adaptive learning environments may be effective for training students to recognize which learning strategies to implement, when to implement them, and to actually follow-through with implementation (e.g., Burchard & Swerdzewski, 2009). In comparison to, say, a traditional in-person course on effective study strategies, a computer-based, adaptive program could easily be offered to a large number of students at once and therefore has the potential to enhance learning for many with little time or financial cost. Accordingly, Endres and colleagues suggested that at the beginning of university education, students could benefit from a computer-based adaptive learning environment to teach effective learning strategies and actually change students' study habits. Endres and colleagues reported empirical evidence regarding how to optimally design such a program. In separate experiments, the authors investigated how to optimally design an acquisition phase (i.e., initially learning about effective study strategies), a practice phase (i.e., reviewing effective study strategies), and an application phase (i.e., a phase to support regular use effective study strategies). The authors emphasized the importance of capturing student interest during initial learning: emotionally appealing instructional videos (e.g., featuring human hands sketching relevant symbols) led to more situational interest than neutral videos during the acquisition phase. However, Endres et al. found that the optimal strategy for reinforcing this initial learning during the practice phase depended on students' prior knowledge of effective study strategies before beginning the experiment and thus recommended an adaptive program. Specifically, low-knowledge students benefited equally from restudying key concepts and engaging in retrieval practice on these concepts. In contrast, medium-knowledge and high-knowledge students benefited more from retrieval practice than restudying, and high-knowledge students benefited more from more difficult retrieval practice. Finally, Endres et al. investigated how to optimize the application phase, that is, how to improve students' self-reported study habits, not merely their knowledge of how to study. The key finding was that students needed significant support to translate knowledge into application. Merely instructing students to set implementation intentions (i.e., writing detailed if-then statements regarding how they will study and under what conditions) did not improve students' self-reported use of effective study strategies, perhaps because students' implementation intentions were of low quality. In sum, digital learning environments not only provide a rich opportunity for students to engage in retrieval practice and receive feedback but may also be an invaluable tool for teaching students to implement effective learning techniques more frequently in their typical study routines.

Concluding Remarks

Students in psychology and education learn about learning as a central topic in their curricula. They should be granted the opportunity to do this in a way that implements the state-of-the-art knowledge of their fields. This enables efficiency in knowledge acquisition and helps to maintain self-consistency within the empirical disciplines that endorse the value of

evidence-based practice. This Special Issue furthers our understanding on how to support learning through retrieval practice and feedback (see Tables 1 and 2). Both by laboratory-based research (e.g., by isolating moderators of retrieval benefits) and use-inspired applied research (e.g., by implementing retrieval and feedback practice in university courses). The articles in this Special Issue also showed that knowledge about test-enhanced and feedback-enhanced learning is being implemented and developed in interdisciplinary domains. We can thus be optimistic that evidence-based practice in teaching will continue to gain momentum and to incorporate findings related to the application of psychological and instructional principles in other domains.

We hope that this Special Issue will inspire you to embed and sophisticate retrieval and feedback procedures in your own learning and teaching of psychology. Please note that the current abstracts from *Psychology Teaching Review* (PTR26(2)) and *Teaching of Psychology* (ToP47(4)) can also be found in this issue.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Abbot, E. E. (1909). On the analysis of the memory consciousness in orthography. *The Psychological Review: Monograph Supplements*, 11(1), 127.
- Abel, M., Haller, V., Köck, H., Pötschke, S., Heib, D., Schabus, M., & Bäuml, K. H. T. (2019). Sleep reduces the testing effect—but not after corrective feedback and prolonged retention interval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45, 272–287.
- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701.
- Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, 25, 764–771.
- Ashworth, A., Hill, C. M., Karmiloff-Smith, & Dimitriou, D. (2014). Sleep enhances memory consolidation in children. *Journal of Sleep Research*, 23(3), 304–310.
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Bäuml, K.-H. T., Holterman, C., & Abel, M. (2014). Sleep can reduce the testing effect – It enhances recall of restudied items but can leave recall of retrieved items unaffected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1568–1581.
- 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.
- Bertilsson, F., Wiklund-Hörnqvist, C., Stenlund, T., & Jonsson, B. (2017). The testing effect and its relation to working memory capacity and personality characteristics. *Journal of Cognitive Education and Psychology*, 16(3), 241–259.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64, 417–444.

- 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.
- Burchard, M. S., & Swerdzewski, P. (2009). Learning effectiveness of a strategic learning course. *Journal of College Reading and Learning*, 40(1), 14–34.
- Butler, A. C., & Woodward, N. R. (2018). Toward consilience in the use of task-level feedback to promote learning. *Psychology of Learning and Motivation*, 69, 1–38.
- Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290–298.
- 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.
- Brooks, C., Carroll, A., & Gillies, R. (2019). A matrix of feedback. *Australian Journal of Teacher Education*, 44 (4), 14–32.
- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13, 826–830.
- Corral, D., & Carpenter, S. K. (2020). Facilitating transfer through incorrect examples and explanatory feedback. *Quarterly Journal of Experimental Psychology*, 73 (9), 1340–1359.
- Cowan N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, 26,197–223.
- Curcio, G., Ferrara, M., & De Gennaro, L. (2006). Sleep loss, learning capacity and academic performance. *Sleep Medicine Reviews*, 10(5), 323–337.
- den Boer, A., Verkoeijen, P., & Heijltjes, A. (2021). Comparing formative and summative cumulative assessment: Two field experiments in an applied university engineering course. *Psychology Learning & Teaching*, 20(1), 128–143.
- Dickelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114–126.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment*, 91(2), 166–174.
- 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.
- Enders, N., Gaschler, R., & Kubik, V. (2021). Online quizzes with closed questions in formal assessment: How elaborate feedback can promote learning. *Psychology Learning & Teaching*, 20(1), 91–106.
- Endres, T., Leber, J., Böttger, C., Rovers, S., & Renkl, A. (2021). Improving lifelong learning by fostering students' learning strategies at university. *Psychology Learning & Teaching*, 20(1), 144–160.
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, 44, 145–200.
- Francis, A. P., Wieth, M. B., Zabel, K. L., & Carr, T. H. (2020). A classroom study on the role of prior knowledge and retrieval tool in the testing effect. *Psychology Learning & Teaching*, 19(3), 258–274.
- Giebl, S., Mena, S., Storm, B. C., Bjork, E. L., & Bjork, R. A. (2021). Answer first or google first? Using the internet in ways that enhance, not impair, one's subsequent retention of needed information. *Psychology Learning & Teaching*, 20(1), 58–75.
- Gunderson E. A., Sorhagen N. S., Gripshover S. J., Dweck C. S., Goldinmeadow S., & Levine S. C. (2017). Parent praise to toddlers predicts fourth grade academic achievement via children's incremental mindsets. *Developmental Psychology*, 54, 397–409.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.

- Howe, P. D. L., McKague, M., Lodge, J. M., Blunden, A. G., & Saw, G. (2018). PeerWise: Evaluating the effectiveness of a web-based learning aid in a second-year psychology subject. *Psychology Learning & Teaching, 17*(2), 166–176.
- Jenkins, J. J. (1979). Four points to remember. A tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Erlbaum.
- Jones, D. L., Nelson, J. D., & Opitz, B. (2021). Increased anxiety is associated with better learning from negative feedback. *Psychology Learning & Teaching, 20*(1), 76–90.
- Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2020). A learning method for all: The testing effect is independent of cognitive ability. *Journal of Educational Psychology*. Advance online publication. Available at: <https://doi.org/10.1037/edu0000627>
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle. *Psychological Bulletin, 131*, 66–71.
- Kerdijk, W., Cohen-Schotanus, J., Mulder, B. F., Muntinghe, F. L., & Tio, R. A. (2015). Cumulative versus end-of-course assessment: Effects on self-study time and test performance. *Medical Education, 49*, 709–716.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.
- Kroneisen, M., & Kuepper-Tetzl, C. E. (2021). Using day and night – Scheduling retrieval practice and sleep. *Psychology Learning & Teaching, 20*(1), 40–57.
- Kubik, V., Jönsson, F. U., de Jonge, M., & Arshamian, A. (2020). Putting action into testing: Enacted retrieval benefits long-term retention more than covert retrieval. *Quarterly Journal of Experimental Psychology, 73*(12), 2093–2105.
- Lowe, C. J., Safati, A., & Hall, P. A. (2017). The neurocognitive consequences of sleep restriction: a meta-analytic review. *Neuroscience & Biobehavioral Reviews, 80*, 586–604.
- Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194–199.
- Marsh, E. J., & Rajaram, S. (2019). The digital expansion of mind: Implications of Internet usage for memory and cognition. *Journal of Applied Research in Memory and Cognition, 8*(1), 1–14.
- Mazza, S., Gerbier, E., Gustin, M.-P., Kasicki, Z., Koenig, O., Toppino, T. C., & Magnin, M. (2016). Relearn faster and retain longer: Along with practice, sleep makes perfect. *Psychological Science, 27*(10), 1321–1330.
- McDaniel, M. A., & Little, J. L. (2019). Multiple-choice and short-answer quizzing on equal footing in the classroom: Potential indirect effects of testing. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (p. 480–499). Cambridge University Press.
- Miner, 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.
- Miyatsu, T., & McDaniel, M. A. (2019). Adding the keyword mnemonic to retrieval practice: A potent combination for foreign language vocabulary learning? *Memory & Cognition, 47*(7), 1328–1343.
- Nevid, J. S., Gordon, A. J., Terjesen, M. D., & Hicks, A. (2020). Classroom evaluation of online quizzing and concept building exercises embedded in an integrated learning system. *Psychology Learning & Teaching, 19*(2), 184–193.
- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review, 32*, 1089–1126.

- O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*(5), 971–990.
- Pashler, H., Cepeda, N. J., Wixted, J. T., Rohrer, D., Pashler, H., & Cepeda, N. J. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3–8.
- Phelps, R. P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing, 12*(1), 21–43.
- Phelps, R. P. (2019). Test frequency, stakes, and feedback in student achievement: A meta-analysis. *Evaluation Review, 43*(3–4), 111–151.
- Pieper, M., Roelle, J., vom Hofe, R., Salle, A., & Berthold, K. (2021). Feedback in reflective journals fosters reflection skills of student teachers. *Psychology Learning & Teaching, 20*(1), 107–127.
- Powers, K. L., Brooks, P. J., Galazyn, M., & Donnelly, S. (2016). Testing the efficacy of MyPsychLab to replace traditional instruction in a hybrid course. *Psychology Learning & Teaching, 15*(1), 6–30.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437–447.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.
- Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1155–1159.
- Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier.
- Roelle, J., Berthold, K., & Fries, S. (2011). Effects of feedback on learning strategies in learning journals: Learner-expertise matters. *International Journal of Cyber Behavior, Psychology and Learning, 1*, 16–30.
- 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.
- Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: A meta-analytic perspective. *Psychology Learning & Teaching, 16*(2), 179–196.
- Sadowski, C. J., & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills, 74*(2), 610.
- Soderstrom, N. C., Kerr, T. K., & Bjork, R. A. (2016). The critical importance of retrieval and spacing for learning. *Psychological Science, 27*(2), 223–230.
- Sotola, L. K., & Crede, M. (2020). Regarding class quizzes: A meta-analytic synthesis of studies on the relationship between frequent low-stakes testing and class performance. *Educational Psychology Review*. Available at: <https://doi.org/10.1007/s10648-020-09563-9>.
- Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*, 776–778.
- Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. *Memory, 25*, 717–723.
- 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.
- Tuckman, B. W. (1998). Using tests as an incentive to motivate procrastinators to study. *Journal of Experimental Education, 66*(2), 141–147.

- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511.
- Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, 55(1), 10–16.
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087.

Guest Editors*March 2021***Veit Kubik**

Bielefeld University, Germany

Robert Gaschler

FernUniversität Hagen, Germany

Hannah Hausman

University of California, Santa Cruz, USA