Journal of Experimental Psychology: Applied

Do Working Memory Capacity and Test Anxiety Modulate the Beneficial Effects of Testing on New Learning?

Chunliang Yang, Bukuan Sun, Rosalind Potts, Rongjun Yu, Liang Luo, and David R. Shanks

Online First Publication, June 4, 2020. http://dx.doi.org/10.1037/xap0000278

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

http://dx.doi.org/10.1037/xap0000278

Do Working Memory Capacity and Test Anxiety Modulate the Beneficial Effects of Testing on New Learning?

Chunliang Yang
Beijing Normal University

Bukuan Sun
Fuqing Branch of Fujian Normal University

Rosalind Potts
University College London

Rongjun Yu
National University of Singapore

Liang Luo
Beijing Normal University

David R. Shanks
University College London

Although testing has repeatedly been shown to be one of the most effective strategies for consolidating retention of studied information (the backward testing effect) and facilitating mastery of new information (the forward testing effect), few studies have explored individual differences in the beneficial effects of testing. The current study recruited a large sample (1,032 participants) to explore the potential roles of working memory capacity and test anxiety in the enhancing effects of testing on new learning, and the converse influence of testing on test anxiety. The results demonstrated that administering interim tests during learning appears to be an effective technique to potentiate new learning, regardless of working memory capacity and test anxiety. At a final test on all studied materials, individuals with low working memory capacity benefited more from interim testing than those with high working memory capacity. These testing effects are minimally modulated by levels of trait/state test anxiety, and low-stake interim testing neither reduced nor increased test anxiety. Overall, the results imply that low-stake interim tests can be administered to boost new learning irrespective of learners' level of WMC, test anxiety, and of possible reactive effects of testing on test anxiety.

Public Significance Statement

Although testing is frequently regarded as an assessment tool, the current study demonstrates that it also appears to be an effective technique to boost new learning across individuals with different levels of working memory capacity and test anxiety. In addition, retrieval practice can more effectively consolidate retention of studied information for individuals with low working memory capacity. Regardless of the concern that frequent tests may increase text anxiety, this research finds that frequent tests tend to exert minimal influence on test anxiety.

Keywords: backward and forward testing effects, working memory capacity, test anxiety, individual differences

Over the last century, since the pioneering study by Abbott (1909), hundreds of experiments have repeatedly demonstrated that retrieval practice (i.e., testing) can more effectively facilitate long-term learning and memory retention than other learning strat-

Chunliang Yang, Institute of Developmental Psychology, School of Psychology, Beijing Normal University; Bukuan Sun, School of Education, Fuqing Branch of Fujian Normal University; Rosalind Potts, Division of Psychology and Language Sciences, University College London; Rongjun Yu, Department of Psychology, National University of Singapore; Liang Luo, Institute of Developmental Psychology, School of Psychology and Collaborative Innovation Centre of Assessment Toward Basic Education Quality, Beijing Normal University; David R. Shanks, Division of Psychology and Language Sciences, University College London.

This research was supported by the Fundamental Research Funds for the Central Universities (2019NTSS28), the Educational Research Project of Young and Middle-Aged Teachers of Fujian Province (JAS170619), the Natural Science Foundation of China (31671130), and the United Kingdom Economic and Social Research Council (ES/S014616/1). The data from this study are at https://osf.io/76wm8/ under an embargo that will be lifted on publication of a follow-up article reporting additional analyses.

Correspondence concerning this article should be addressed to Chunliang Yang, Institute of Developmental Psychology, School of Psychology, Beijing Normal University, 19 Xinjiekouwai Street, Haidian District, Beijing 100875, China. E-mail: chunliang.yang@bnu.edu.cn

egies, such as restudying (Roediger & Karpicke, 2006), note-taking (Rummer, Schweppe, Gerst, & Wagner, 2017), or creating concept maps (Karpicke & Blunt, 2011). Most of this research has explored the benefits of testing on memory for previously studied information, an effect we term the *backward testing effect* (BTE), following Pastötter and Bäuml (2014) and Yang, Potts, and Shanks (2018). Besides this direct benefit of testing to the studied/tested material, an indirect benefit of retrieval practice is that testing of studied information, by comparison with restudying or doing nothing, can more effectively enhance subsequent learning and retention of new information. This phenomenon is termed the *forward testing effect* (FTE) and is the focus of the current research.

The backward and forward advantages of testing jointly make a tempting case that learners and instructors should use retrieval practice as a practical strategy to enhance learning and teaching activity. Indeed, many publications (including academic articles, book chapters, and policy and news reports) have promoted the use of testing in educational settings (e.g., Brown, Roediger, & McDaniel, 2014; Pashler et al., 2007; Putnam, Sungkhasettee, & Roediger, 2016; Roediger & Karpicke, 2006; Yang, Potts, & Shanks, 2018). However, before recommending testing to practitioners, it is critical to determine whether the benefits of retrieval practice are generalizable across individuals with diverse cognitive abilities (e.g., working memory capacity [WMC]) and traits (e.g., test anxiety).

WMC refers to an individual differences construct reflecting the span for temporarily holding information available for current processing (Wilhelm, Hildebrandt, & Oberauer, 2013) and test anxiety captures a combination of anxious symptoms (such as overarousal, tension, worry about test failure, etc.) before or during taking a test (Spielberger, 1980). WMC is measured by a variety of tests that require the updating and maintenance of information (Wilhelm et al., 2013), whereas test anxiety is measured by a standardized questionnaire instrument. Both WMC and test anxiety have been assumed to be key modulators of the beneficial effects of testing (Tse & Pu, 2012) but have mainly been studied in relation to the BTE. Here we especially explore their roles in relation to the forward effects of testing. Below we briefly review the literature on the FTE and research findings derived from previous studies regarding the roles of WMC and test anxiety in the BTE, and then we introduce the rationale and design of the current study.

The FTE

Szpunar, McDermott, and Roediger (2008, Experiment 1B) conducted a classic study demonstrating the FTE using a multilist procedure. They instructed two groups (test/no-test) of participants to study five 18-word lists, which were studied one word at a time (2 s each) and list-by-list. After studying each of Lists 1–4, the test group took a free recall test in which they recalled words from the just-studied list, whereas the no-test group solved math problems. After studying List 5, both groups were instructed to freely recall as many List 5 words as they could. The results showed that the test group ($M = 7.00$ out of 18) correctly recalled twice as many List 5 words as the no-test group ($M = 3.50$), clearly demonstrating the FTE: Interim testing on Lists 1–4, by comparison with no testing, substantially potentiated the learning and retrieval of List 5. Furthermore, Szpunar et al. (2008) observed that their no-test group ($M = 3.70$) suffered from over 10 times as many prior list intrusions (incorrectly recalling List 1–4 words when instructed to recall List 5) as their test group ($M = 0.30$), indicating that, besides potentiating new learning, interpolated testing can also effectively prevent the build-up of prior list intrusions across lists.

In recent years, an emerging body of studies has been conducted to explore the FTE on different types of learning as well as its robustness, limits, and theoretical basis (for reviews, see Chan, Meissner, & Davis, 2018; Pastötter & Bäuml, 2014; Yang, Potts, & Shanks, 2018). It has been shown that the FTE generalizes to various types of learning, such as the learning of single items (e.g., Chan, Manley, Davis, & Szpunar, 2018; Szpunar et al., 2008; Weinstein, Gilmore, Szpunar, & McDermott, 2014; Yang, Potts, & Shanks, 2017), paired associates (e.g., Weinstein, McDermott, & Szpunar, 2011; Yang et al., 2017), text passages (e.g., Wissman, Rawson, & Pyc, 2011), lecture videos (e.g., Jing, Szpunar, & Schacter, 2016; Szpunar, Khan, & Schacter, 2013), artists' painting styles (e.g., Lee & Ahn, 2018; Yang & Shanks, 2018), and spatial information (Bufe & Aslan, 2018). C. L. Yue, Soderstrom, and Bjork (2015) and Yang, Chew, Sun, and Shanks (2019) demonstrated transfer of the FTE. C. L. Yue et al. (2015, Experiment 2), for instance, observed that testing of a studied lecture video on one topic (e.g., the life cycle of a star) potentiated the learning of a new video on a completely different topic (e.g., lightning formation). Yang et al. (2019, Experiment 3) observed that testing on text statements about artists' contributions to art facilitated subsequent visual learning of different artists' painting styles.

Recently, researchers have turned to exploring the FTE in different populations, such as older adults (Pastötter & Bäuml, 2019), young children (Aslan & Bäuml, 2016), and patients with traumatic brain injury (Pastötter, Weber, & Bäuml, 2013). But little research has been conducted to explore whether the FTE generalizes across individuals with different levels of WMC and test anxiety. As will be discussed below, both WMC and test anxiety have been assumed to be key modulators of test-enhanced learning. Hence, the primary aim of the current study is to investigate their potential roles in the FTE.

Potential Roles of WMC and Test Anxiety in the BTE and FTE

A few studies have explored the modulating roles of WMC and test anxiety in the BTE, but with inconsistent and conflicting findings. For instance, Tse and Pu (2012) demonstrated a weak positive correlation between WMC and the BTE, reflecting the fact that individuals with high WMC benefitted more from testing than ones with low WMC. In contrast, Agarwal, Finley, Rose, and Roediger (2017) recently observed that the BTE was attenuated with increasing WMC, and they concluded that retrieval practice is more beneficial for individuals with low WMC than for ones with high WMC. Inconsistent with both of these outcomes, Brewer and Unsworth (2012), Nordstrand (2018), and Wiklund-Hörnqvist, Jonsson, and Nyberg (2014) found null relationships between the BTE and WMC, implying that testing reliably enhances retention of studied information regardless of WMC.

Similar to the research on the potential role of WMC, only a few studies have explored the role of test anxiety in the BTE and the research findings are similarly conflicting. Tse and Pu (2012) found that, for low WMC individuals, there was a negative correlation between test anxiety and the BTE, and Mok and Chan (2016) demonstrated that individuals with high test anxiety benefited less from testing than ones with low test anxiety. In addition, Hinze and Rapp (2014) showed that increasing test performance pressure correspondingly decreased the BTE. Such findings imply that high test anxiety (or pressure) may reduce the beneficial effects of testing, or that individuals with relatively higher test anxiety may even suffer from testing (for an illustration, see the top-right panel in Tse and Pu's [2012] Figure 2, p. 259). By contrast, Clark, Crandall, and Robinson (2018) recently reported a positive relationship between test anxiety and the BTE in their nonincentive condition, indicating that testing is more beneficial for individuals with high test anxiety.

To date, only Tse and Pu (2012) have explored how WMC and test anxiety jointly modulate the BTE. These researchers found that, for high WMC participants, test anxiety had little impact on the magnitude of the BTE. Critically, for low WMC participants, the magnitude of the BTE decreased as a function of test anxiety. Tse and Pu proposed that, for individuals with low WMC and high test anxiety, their limited working memory resources might be expended on buffering their high test anxiety, leading to fewer resources deployed for learning and a smaller positive (or even a negative) effect of testing on memory for studied information.

As discussed, research findings regarding the potential modulating roles of WMC and test anxiety in the BTE have been somewhat inconsistent and equivocal. One possible reason for the inconsistency might be that the sample sizes in those studies, ranging from 61 (Clark et al., 2018) to 160 (Tse & Pu, 2012), were relatively small. It is well known that exploring individual differences in cognitive behaviors requires large sample sizes, and small sample sizes frequently lead to false positive (Type I error) and negative (Type II error) results. Indeed, Agarwal et al. (2017, p. 770) acknowledged that their sample size (156 participants), despite being one of the larger samples in studies exploring individual differences in the BTE, might be inadequate to unravel the relationship between WMC and the BTE, and they emphasized the need for future research to further explore this with larger sample sizes. With this in mind, the current research (which examined the effects of WMC and test anxiety on the forward rather than backward testing effect) aimed to recruit a large sample (over 1,000 participants).

With one exception, the roles of WMC and test anxiety in the FTE have not been explored. Pastötter and Frings (2019, Experiment 2) recruited 240 participants to examine the relationship between WMC and the FTE. They found no relationship between these two factors, indicating that the FTE generalizes to individuals with different levels of WMC. The present research goes beyond Pastötter and Frings's (2019) study by measuring both WMC and test anxiety in a high-powered study and assessing their individual and joint modulating effects on the FTE.

According to Tse and Pu's (2012) findings (i.e., a negative correlation between test anxiety and the BTE for individuals with low WMC) and their hypothesis (i.e., individuals with high test anxiety and low WMC have insufficient cognitive resources to concurrently buffer high test anxiety and perform the ongoing learning task, leading to a smaller positive or even negative BTE), it is reasonable to assume that such individuals may worry about subsequent interim tests and devote their limited cognitive resources to buffering their test anxiety. Concurrently buffering high test anxiety and encoding new information may lead to a smaller positive or even negative FTE for individuals with low WMC and high test anxiety.

Potential Effects of Frequent Tests on Test Anxiety

Test anxiety is a major cause of learning deficits (e.g., difficulty concentrating, poor test performance), and how to mitigate test anxiety has long been an important goal for learners, educators, and psychologists (Hembree, 1988). Although there is a concern that taking tests may increase test anxiety (Steele, 2011), two studies, using questionnaires, demonstrated that students tend to believe that frequent tests can alleviate test anxiety (Agarwal, D'Antonio, Roediger, McDermott, & McDaniel, 2014; Sullivan, 2017). For instance, Sullivan (2017) asked 353 undergraduates and graduates if they agreed with the statement: "The option to retake a quiz reduced my test anxiety." Over 90% of students agreed with this statement. Agarwal et al. (2014) asked 1,408 middle and high school students to answer the question: "Did clicker quizzes make you more or less nervous for unit tests?," and students were offered a choice of three answers: "more," "about the same," or "less." Over 70% of students chose "less," and only 6% reported "more." Although these studies have the advantage of having been conducted in real classroom settings, they relied on students' self-report and elicited students' beliefs about the value of repeated testing rather than directly measuring actual test anxiety before a test in the context of prior testing versus no prior testing of the material. In addition, these studies explored test anxiety solely in the context of the BTE.

Szpunar et al. (2013) experimentally examined whether frequent tests can reduce test anxiety in a forward testing situation. They instructed participants to study a four-segment lecture video. An experimental group took an interim test after studying each segment; by contrast, the control group restudied Segments 1–3 and took an interim test on Segment 4. After the Segment 4 interim test, all participants reported how anxious they were about the final test on the whole video. Besides obtaining an FTE (i.e., tested participants recalled more items in the Segment 4 interim test than those in the control groups), Szpunar et al. (2013) also observed that their experimental group reported lower test anxiety regarding the final test than their control group.

Although the two questionnaire studies described above demonstrated that students tend to believe that testing mitigates test anxiety, we highlight that, in many situations, survey reports can be inaccurate. For an illustration, consider the well-known font size effect. People believe that words presented in a large font (48-pt) will be remembered substantially better than ones in a small (18-pt) font, but in fact font size has little influence on memory retention (M. L. Mueller, Dunlosky, Tauber, & Rhodes, 2014; Rhodes & Castel, 2008; Yang, Huang, & Shanks, 2018). Although Szpunar et al. (2013) experimentally demonstrated that frequent tests reduce test anxiety, obviously one such study is insufficient to draw a firm conclusion. Hence, the current experiment also aims to explore the effect of testing on test anxiety.

Overview of the Current Study

As discussed above, little research has explored the roles of WMC and test anxiety in the FTE, and hence the current study primarily aims to assess the modulating roles of WMC and test anxiety in the FTE. Although one study has experimentally demonstrated that testing may reduce test anxiety, further research is required to confirm (or refute) this finding, which is the second aim of the current study.

Previous studies have obtained conflicting results regarding the roles of WMC and test anxiety in the BTE, and the sample sizes in those studies were relatively small, signaling a need to recruit a large sample of participants to measure individual differences in the FTE. Accordingly, the current experiment recruited over 1,000 participants.

Participants were instructed to learn five word lists (see Figure 1 for a schematic illustration of the experimental procedure). Before undertaking the learning task, all participants completed the Test Anxiety Inventory (TAI; Chinese version; X. Yue, 1996) to measure their trait test anxiety (i.e., the tendency to perceive test situations as threatening). Then they studied five word lists, and the procedure of this learning task was adapted from Szpunar et al. (2008; for details, see the flowchart in Figure 1). During the multilist learning task, participants also reported their state test anxiety regarding the upcoming tests. Finally, participants completed the widely used Operation Span (OSPAN; Chinese version) task to measure their WMC (Unsworth, Heitz, Schrock, & Engle, 2005).

In summary, the current study measured participants' trait test anxiety (measured by the TAI), state test anxiety regarding the List 5 interim test, List 5 interim test recall (an index of the FTE), state test anxiety regarding the cumulative test, cumulative test recall, and WMC (measured by the OSPAN task). To investigate the roles of test anxiety and WMC in the FTE, we ask whether test anxiety and WMC moderate the effect of interim testing on List 5 interim test recall. To explore whether interpolated testing alleviates test anxiety, we compare state test anxiety reports between a group that takes interim tests on Lists 1–4 and one not taking these tests.

Method

Participants

We decided to continue data collection either until we obtained over 1,000 participants' data or until the end of the academic semester (i.e., the Fall 2018 semester). Based on the effect size (Cohen's $d = 1.43$) for the FTE documented in Yang et al. (2017, Experiment 4), the estimated power to detect a significant ($\alpha = .05$) FTE with such a large sample size is greater than .950 (calculated using G^*Power; Faul, Erdfelder, Lang, & Buchner, 2007). In addition, this sample has power greater than .950 for detecting a small effect ($f^2 = .02$) in a multiple linear regression analysis (i.e., the relationships among the FTE, WMC, and test anxiety).

Accordingly, we recruited 1,075 Chinese participants from the Fuqing Branch of Fujian Normal University before the end of the semester. They were either recruited from Psychology classes or through campus advertisement. Because of computer errors, data

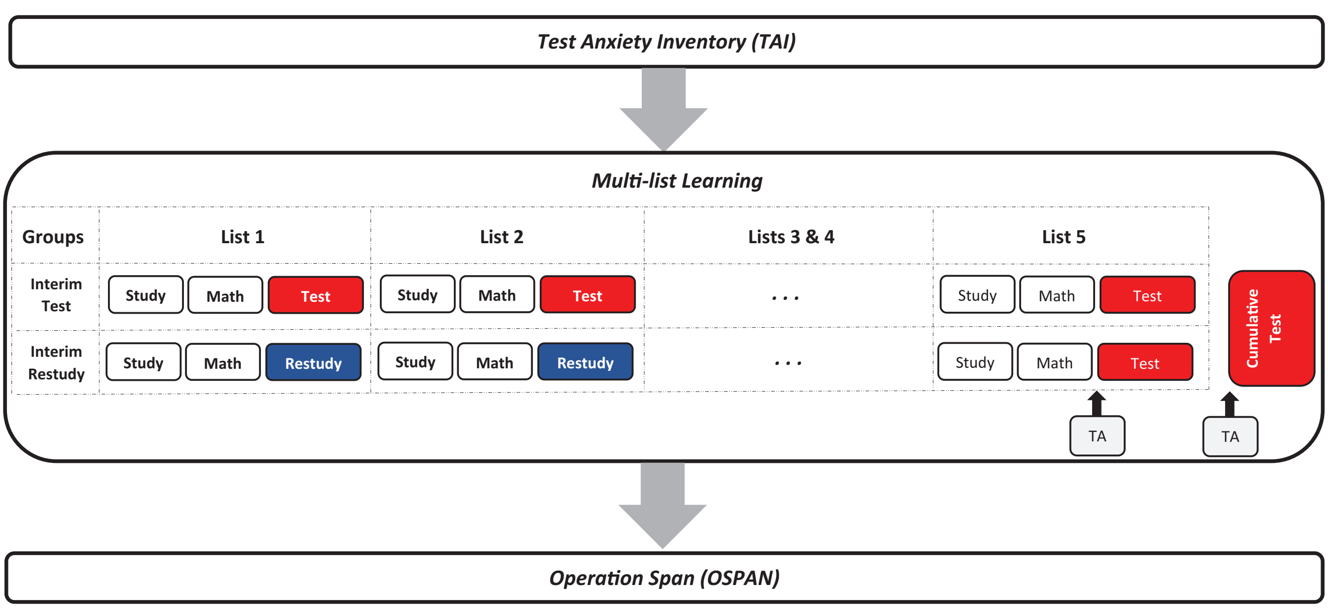

Figure 1. Experimental design schema for the interim test and interim restudy groups. Both groups completed the Test Anxiety Inventory followed by the multilist learning task. After studying List 1 and solving math problems for 1 min, the interim test group recalled List 1 words whereas the interim restudy group restudied List 1. Lists 2–4 were identical to List 1 in each group, but with new words. After studying List 5 and solving math problems, both groups reported their test anxiety (TA) regarding the List 5 interim test and then recalled as many List 5 words as they could. Both groups then reported their TA regarding the cumulative test and recalled as many words as they could from all five lists in the cumulative test. Finally, both groups performed the operation span (OSPAN) task. See the online article for the color version of this figure.

from 42 participants were not recorded, leaving data from a final total of 1,033 participants. One participant was removed from the data analyses because this individual took notes during the multilist learning task, violating the experimental instructions.

The remaining 1,032 participants were randomly allocated to the interim test (518 participants) and interim restudy (514 participants) groups. Their mean age was 18.63 (SD = 1.10; 96 did not report their age). Six hundred and 59 were female, 284 were male, and the remaining 89 did not report their gender. All were native Chinese speakers. They were tested either individually or in groups of up to 20 in a quiet laboratory room.[1] They participated either for course credit, for monetary compensation, or voluntarily. The Ethics Committee at the School of Education, Fuqing Branch of Fujian Normal University, approved the experiment.

Materials

The 20-item TAI (Chinese version) was used to measure participants' trait test anxiety (X. Yue, 1996). Cronbach's alpha of the TAI in our sample was .889, indicating good internal consistency. For the multilist learning task, 90 two-character high-frequency Chinese words were selected from Liu and Reichle (2017; available at https://osf.io/fp3yw/). Word frequency ranged from 51.98 to 768.09 per million (M = 132.47; SD = 111.00), and the number of strokes ranged from 10 to 21 (M = 14.88; SD = 2.68). To prevent any item selection effects, for each participant, the computer randomly separated the words into five lists and assigned them to Lists 1–5. The widely used OSPAN task (Chinese version) was used to measure WMC, and this task has been estabished as valid and reliable for measuring WMC (2-week test-retest reliablity = .83; Unsworth et al., 2005).

Procedure

Figure 1 schematically depicts the procedure. After arriving at the laboratory, all participants first completed the 20-item TAI. The 20 items, which cover a variety of test anxiety symptoms (such as "During tests I feel very tense"), were presented one by one in a random order. Participants reported how frequently they experienced these symptoms on a 4-point scale ranging from 1 (*never*) to 4 (*always*).[2]

Following the completion of the questionnaire, participants initiated the multilist learning task. At the beginning of the learning task, they were instructed to study five lists of words in preparation for a final cumulative test, during which they would be asked to recall as many words as they could from all five lists. They were also informed that, after studying each individual list, the computer would randomly decide either to give them a memory test or to offer them a restudy opportunity on that list. In fact, the test decisions were predetermined. As depicted in Figure 1, the interim test group took an interim test on each of Lists 1–5, whereas the interim restudy group restudied Lists 1–4 and took an interim test on List 5.

In the List 1 study phase, 18 words were presented one by one, for 2 s each, in a random order. A cross sign was presented for 0.5 s after the presentation of each word to mark the interstimulus interval. After studying the words, both groups solved as many simple math problems (e.g., 32 + 49 = ?) as they could for 1 min in a distractor task. Immediately following the distractor task, the interim test group was instructed to recall as many words as they could from the just-studied list (List 1) in 1 min. By contrast, all words were presented again, one by one, for 2 s each, in a new random order for the interim restudy group to restudy.[3] The procedures for Lists 2–4 were the same as for List 1, except that participants studied new words in each list.

After the completion of List 4, both groups studied List 5 and solved math problems for 1 min. Then both groups were informed that the computer had decided to test them on List 5 (i.e., they would be required to recall as many List 5 words as they could), and they reported how anxious they were regarding the upcoming test on a scale ranging from 1 (*not anxious at all*) to 9 (*extremely anxious*). Immediately following this report of state test anxiety, both groups took the List 5 interim test, during which they had unlimited time to recall as many List 5 words as they could.[4] Following the List 5 interim test, both groups reported how anxious they were about the final cumulative test on the same 9-point scale, and then took the cumulative test, during which they recalled as many words as they could from all five lists. The cumulative test was self-paced. No feedback was provided in the interim and cumulative tests.

Finally, participants completed the OSPAN task to measure their WMC (for detailed description of the OSPAN procedure, see Unsworth et al., 2005). In each trial, participants viewed a series of letters which were presented one by one. After viewing each letter, they solved a two-operator arithmetic problem (e.g., 4 + 6 = ?), and after viewing all letters and solving all arithmetic problems in a trial, they recalled the just-viewed letters in the same sequence as they appeared in the encoding phase. Participants' task was to remember as many letters as they could and concurrently to maintain their arithmetic answer accuracy higher than 85%. Following Tse and Pu (2012), we took the absolute OSPAN score as an index of WMC.

Participants were allowed to take self-determined rest intervals between the questionnaires, the multilist learning task, and the OSPAN task. The entire experiment took about 60 min. At the end, participants were debriefed.

Results and Discussion

All participants completed the TAI and the multilist learning task. Eighty-four participants (30 in the interim test and 54 in the

[1] In each group test session, participants were randomly assigned to the interim test and interim restudy conditions.

[2] During the questionnaire phase, participants also completed a battery of other questionnaires, including the Adult ADHD Self-Report Scale (ASRS) Symptom Checklist, the Deliberate and Spontaneous Mind Wandering Scales (Carriere, Seli, & Smilek, 2013), the Theories of Intelligence Scale—Self Form for Adults (Dweck, 1999), and the View of Failure Scale (Haimovitz & Dweck, 2016). Given that we administered these questionnaires only for exploratory purposes and these data were not related to the main aims of the current research, we do not discuss them further.

[3] The interim restudy group spent 45 s restudying words and hence it would have been possible to allocate the same amount of time on interim tests to completely equate the task duration between groups. But because participants might be unable to complete their recall in 45 s, we extended recall time to 1 min, following Szpunar et al. (2008).

[4] We limited the test durations (i.e., 1 min) of the List 1–4 interim tests to roughly equate the task duration between groups. We did not control the duration of the List 5 interim test due to the concern that the interim restudy group, having no experience with the interim test, might not be able to complete their recall within 1 min.

interim restudy groups) failed to complete the OSPAN task, and their WMC scores were treated as missing values in the following analyses. For 111 participants (55 in the interim test and 56 in the interim restudy groups), their overall arithmetic accuracy in the OSPAN task was below 85%. Following Unsworth et al.'s (2005) recommendation, we removed their WMC scores and treated them as missing values. The final sample contains WMC scores for 837 participants (433 in the interim test and 404 in the interim restudy groups).

There were no differences between groups in age, trait test anxiety, or WMC, as reflected by Bayes factors (all $BF_{10} \leq 0.14$). About 68% and 72% of participants were female in the interim test and interim restudy groups, respectively. A proportion (chi-square) test revealed no difference in gender between the groups, $\chi^2 = 1.75$, $p = .19$. Table 1 provides the correlation matrix among variables. As an aid, in Table 2 we summarize the main findings (and conclusions) regarding the key outcome measures and the modulating roles of test anxiety and WMC.

Recall Performance in the Multilist Learning Task

For the interim test group, recall in the List 1–5 interim tests was 7.11 ($SD = 3.22$), 7.14 ($SD = 3.47$), 7.23 ($SD = 3.62$), 7.36 ($SD = 3.90$), and 7.17 ($SD = 4.15$), respectively. A repeated-measures analysis of variance (ANOVA) showed no main effect of List (1–5), $F(4, 2068) = 0.68$, $p = .606$, $\eta_p^2 = .001$, indicating that recall was maintained at a comparable level across lists.

List 5 interim test recall and prior list intrusions (incorrectly recalling words from Lists 1–4 in the List 5 interim test) for both groups are shown in Figures 2A and 2B, respectively. The interim test group correctly recalled nearly three times as many List 5 words as the interim restudy group ($M = 2.65$, $SD = 3.80$), difference = 4.51, 95% confidence interval (CI) [4.03, 5.00], $t(1030) = 18.21$, $p < .001$, Cohen's $d = 1.13$, $BF_{10} = 5.6e + 60$, demonstrating a strong FTE. Consistent with previous studies (Szpunar et al., 2008; Weinstein et al., 2014; Yang et al., 2017), we also observed that the interim restudy group ($M = 5.98$, $SD = 5.20$) suffered from about eight times as many prior list intrusions as the interim test group ($M = 0.75$, $SD = 1.40$) in the List 5 interim test, difference = 5.23 [4.76, 5.69], $t(1030) = 22.10$, $p < .001$, $d = 1.38$, $BF_{10} = 1.2e + 85$, indicating that interim testing effectively reduces prior list intrusions across lists.

For the cumulative test, following Yang et al. (2017), we separately analyzed recall of items from Lists 1–4 and List 5, because the interim test group was tested on Lists 1–4 whereas the interim restudy group restudied these lists, but both groups took an interim test on List 5. Any difference in List 1–4 cumulative test recall could be attributed to two potential factors: the BTE (i.e., testing on studied lists enhances their retention more effectively than restudying) and the FTE (i.e., testing on prior lists potentiates subsequent learning of new lists).

As shown in Figure 2C, in the cumulative test, the interim test group ($M = 13.30$, $SD = 8.16$) correctly recalled more List 1–4 words than the interim restudy group ($M = 11.22$, $SD = 10.04$), difference = 2.08 [0.96, 3.20], $t(1030) = 3.65$, $p < .001$, $d = 0.23$, $BF = 47.51$. It is unknown whether the superior List 1–4 cumulative test recall in the interim test group originated from the BTE or FTE (or a combination of both), and the current study was not designed to disentangle them. Hence, in the following discussions we use the term *retrieval practice effect* as a descriptive label for the positive testing effect(s) on List 1–4 cumulative test recall.

Cumulative test recall for each of Lists 1–4 was also analyzed. Because these results were not the key research interest, they are not reported here but are available on the Open Science Framework (https://osf.io/nqpvm/), and corresponding theoretical implications are discussed. In brief, there was a significantly positive effect of retrieval practice on List 3 and 4 cumulative test recall, whereas a negative effect on List 1 was detected, which means that the overall enhancing effect of interim testing on List 1–4 cumulative test recall reported above is largely attributable to its facilitating effect on Lists 3 and 4, which cancelled out the negative effect on List 1.

As shown in Figure 2D, similar to their List 5 interim test recall, in the cumulative test the interim test group ($M = 5.67$, $SD = 3.79$) correctly recalled about twice as many List 5 words as the interim restudy group ($M = 2.53$, $SD = 3.41$), difference = 3.15 [2.71, 3.59], $t(1030) = 14.02$, $p < .001$, $d = 0.87$, $BF_{10} = 3.7e + 37$.

State Test Anxiety for the List 5 Interim Test and Cumulative Test

We now turn to the question of whether frequent tests attenuate test anxiety. The TAI measured participants' trait test anxiety, and there was no difference in trait test anxiety between groups before the multilist learning task. Before the List 5 interim and cumulative tests, all participants reported their state test anxiety regarding the upcoming tests.

As shown in Figures 2E and 2F, there was no meaningful difference between the two groups' state test anxiety regarding the List 5 interim test (interim test: $M = 2.86$, $SD = 2.27$; interim

Table 1
Correlation Matrix of Variables

Variables	1	2	3	4	5	6	7	8
1. Trait test anxiety	—							
2. State test anxiety (List 5)	0.291***	—						
3. List 5 interim test recall	−0.008	−0.078*	—					
4. Prior list intrusions in List 5 interim test	−0.022	0.004	−0.437***	—				
5. State test anxiety (cumulative)	0.324***	0.665***	−0.054	−0.044	—			
6. List 1–4 cumulative test recall	0.006	−0.030	0.459***	0.086**	−0.059	—		
7. List 5 cumulative test recall	−0.020	−0.070*	0.418***	0.559***	−0.079*	0.422***	—	
8. WMC	−0.055	−0.004	0.160***	−0.011	0.024	0.165***	0.119***	—

Note. Trait test anxiety = test anxiety measured by the TAI; State test anxiety (List 5) = state test anxiety ratings for the List 5 interim test; State test anxiety (cumulative) = state test anxiety ratings for the cumulative test; WMC = working memory capacity.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2
Main Findings Regarding the Modulating Effects of Trait/State Test Anxiety and Working Memory Capacity (WMC) on Measures of Recall

Variables	List 5 interim test recall (FTE)	List 1–4 cumulative test recall (retrieval practice effect)
Interim task	Yes: Interim testing boosted List 5 interim test recall.	Yes: Interim testing boosted List 1–4 cumulative test recall.
Trait test anxiety	No: Trait test anxiety did not significantly correlate with List 5 interim test recall.	No: Trait test anxiety did not significantly correlate with List 1–4 cumulative test recall.
State test anxiety	Yes: There was a weakly negative correlation between state test anxiety and List 5 interim test recall.	No: State test anxiety did not significantly correlate with List 1–4 cumulative test recall.
WMC	Yes: Greater WMC was associated with superior List 5 interim test recall.	Yes: Greater WMC was associated with superior List 1–4 cumulative test recall.
Interim Task × Trait Test Anxiety	No: Trait test anxiety did not modulate the FTE.	No: Trait test anxiety did not modulate the retrieval practice effect.
Interim Task × State Test Anxiety	No: State test anxiety did not modulate the FTE.	No: State test anxiety did not modulate the retrieval practice effect.
Interim Task × WMC	No: WMC did not modulate the FTE.	Yes: The retrieval practice effect was larger for participants with low than those with high WMC.
Interim Task × Trait Test Anxiety × WMC	No: Trait test anxiety and WMC did not jointly modulate the FTE.	No: Trait test anxiety and WMC did not jointly modulate the retrieval practice effect.
Interim Task × State Test Anxiety × WMC	No: State test anxiety and WMC did not jointly modulate the FTE.	No: State test anxiety and WMC did not jointly modulate the retrieval practice effect.

Note. FTE = forward testing effect.

restudy: $M = 2.86$, $SD = 2.47$), difference <0.001 [−0.29, 0.29], $t(1030) = -0.006$, $p = .996$, $d < -0.001$, $BF_{10} = 0.07$, and the same was true for state test anxiety regarding the cumulative test (interim test: $M = 3.16$, $SD = 2.34$; interim restudy: $M = 3.07$, $SD = 2.57$), difference = 0.09 [−0.21, 0.39], $t(1029) = 0.62$, $p = .54$, $d = 0.04$, $BF_{10} = 0.08$.[5] Overall these results suggest that, at least under the current conditions, testing neither increases nor reduces test anxiety.

Roles of Trait/State Test Anxiety and WMC in the FTE

Regression analyses were conducted to examine the question of main research interest: whether trait/state test anxiety and WMC individually or jointly modulated the FTE. In these analyses, List 5 interim test recall was entered as the dependent variable. In the first step, we entered participants' age, gender (male = 0; female = 1), and Interim task (restudying = 0; testing = 1) as the independent variables in the regression model. In the second step, trait test anxiety and WMC scores were added. In the third step, we entered the product of Interim task and WMC scores, the product of Interim task and trait test anxiety scores, the product of WMC and trait test anxiety scores, and the product of Interim task, WMC, and trait test anxiety scores (i.e., the interaction variables).

The detailed regression results are summarized in Table 3 and are visually shown in Figures 3A–C. At the first step, there was a main effect of Interim task, $\beta = 4.387$, $p < .001$. At the second step, WMC scores positively predicted List 5 interim test recall, $\beta = 0.045$, $p < .001$, indicating that individuals with better WMC learned more words in List 5 (Figure 3A), but trait test anxiety had no significant influence, $\beta = -0.009$, $p = .637$. At the third step, no variables (apart from interim task) had a main effect on List 5 interim test recall. Furthermore, there was no interaction between interim task and WMC, $\beta = -0.079$, $p = .329$, indicating that the FTE generalizes across individuals with differing levels of WMC. Similarly, there was no interaction between Interim task and trait test anxiety, $\beta = -0.115$, $p = .367$, demonstrating that the FTE generalizes across individuals with different levels of trait test anxiety. Finally, and most importantly, there was no interaction between Interim task, trait test anxiety, and WMC, $\beta = 0.002$, $p = .405$, indicating that trait test anxiety and WMC did not jointly moderate the FTE.[6]

We repeated all the above analyses but replaced trait test anxiety scores with state test anxiety scores (i.e., the test anxiety ratings before the List 5 interim test) as a predictor of List 5 interim test recall, to ascertain whether state test anxiety modulates the FTE.

Table 4 presents the detailed regression results. All the results showed the same pattern as those based on trait test anxiety, except that at the second step, state test anxiety negatively correlated with List 5 interim test recall, $\beta = -0.128$, $p = .028$. Participants reporting higher state anxiety learned fewer words in List 5 (Figure 3C). Similar to trait test anxiety, state test anxiety did not modulate

[5] Because of an unknown computer error, one participant's test anxiety rating for the cumulative test was not recorded in the interim restudy group. This participant's test anxiety regarding the cumulative test was treated as a missing value in the following analyses.

[6] Following Tse and Pu (2012), we divided participants into two (high and low WMC) subsets to further explore whether trait test anxiety affects the FTE differently for individuals with different levels of WMC. The results remained the same: no modulating role of WMC or trait/state test anxiety in the FTE. Median split analyses were also performed to replicate the results from the following regression analyses, and all results showed the same patterns. Considering the limitations and potential pitfalls of median split analysis (Iacobucci, Posavac, Kardes, Schneider, & Popovich, 2015; Liben-Nowell, Strand, Sharp, Wexler, & Woods, 2019; Maxwell & Delaney, 1993) and for the sake of brevity, we do not further discuss those results.

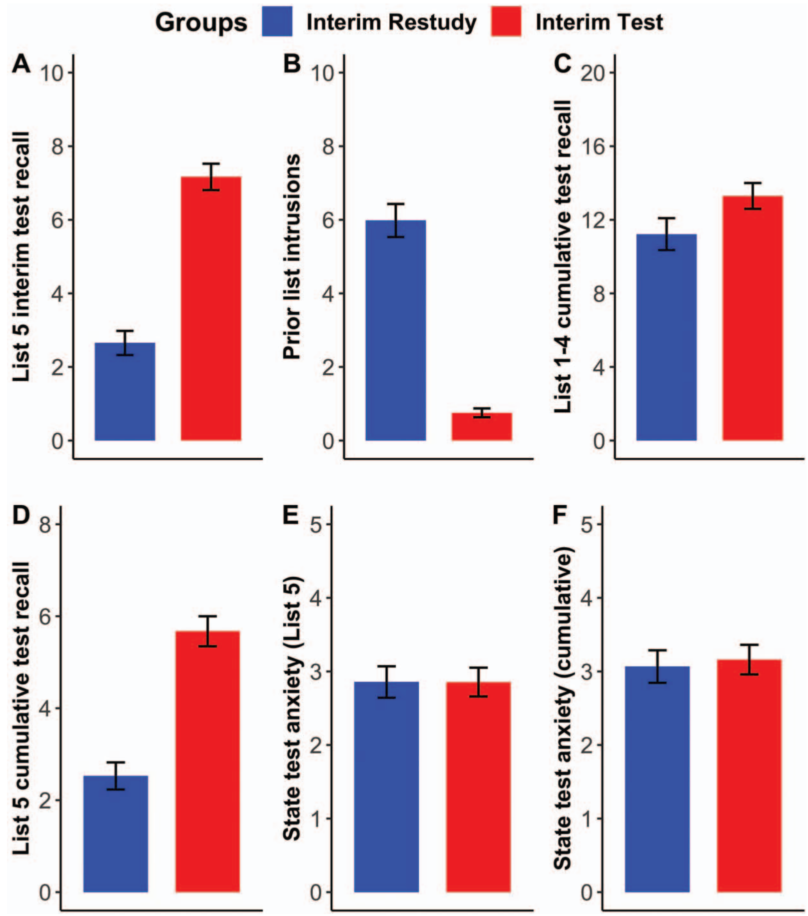

Figure 2. A: List 5 interim test recall; B: Prior list intrusions in the List 5 interim test; C: List 1–4 cumulative test recall; D: List 5 cumulative test recall; E: State test anxiety for the List 5 interim test; F: State test anxiety for the cumulative test. Error bars represent 95% CI. See the online article for the color version of this figure.

the FTE in that there was no interaction between Interim task and state test anxiety, $\beta = -0.467$, $p = .280$.

Overall, these results demonstrate that interim tests tend to reliably potentiate learning and recall of new information regardless of WMC and trait/state test anxiety.

Roles of Trait/State Test Anxiety and WMC in the Retrieval Practice Effect on List 1–4 Cumulative Test Recall

In this section, we address the question of whether trait/state test anxiety and WMC moderated the retrieval practice effect (i.e., the effect of interim testing on List 1–4 cumulative test recall). Similar regression analyses, in which the dependent variable was changed to List 1–4 cumulative test recall, were conducted to explore the potential roles of trait test anxiety and WMC in the retrieval practice effect.

Table 5 lists the detailed regression results and the regression relationships are visually depicted in Figures 3D–F. At the first step, Interim task had a main effect on List 1–4 cumulative test recall, $\beta = 2.418$, $p < .001$. It is worth noting that gender also had a main effect, $\beta = 3.977$, $p < .001$: female participants recalled more List 1–4 words in the cumulative test than male participants (for related findings, see Herlitz, Nilsson, & Bäckman, 1997). At the second step, WMC positively predicted List 1–4 cumulative test recall (individuals with better WMC recalled more words overall), $\beta = 0.085$, $p < .001$, but trait test anxiety did not, $\beta = -0.024$, $p = .553$. At the third step, there were no main effects or interactions among Interim task, trait test anxiety, and WMC.

As clearly shown in Figure 3D, List 1–4 cumulative test recall in participants with low WMC tended to benefit more from interim testing than that of participants with high WMC. We therefore conducted a further regression analysis with only interim task and WMC scores as the independent variables (trait test anxiety now excluded) and List 1–4 cumulative test recall as the dependent variable. The results show a main effect of interim task, $\beta = 7.675$, $p < .001$, a main effect of WMC, $\beta = 0.153$, $p < .001$, and a significant interaction between interim task and WMC, $\beta = -0.109$, $p = .010$.

We repeated the above analyses but replaced trait test anxiety scores with state test anxiety scores for the cumulative test (i.e., test anxiety ratings prior to the cumulative test). The regression results are reported in Table 6. At the third step, the interaction

Table 3
Regression Coefficients of Interim Task, Working Memory Capacity (WMC), and Trait Test Anxiety on List 5 Interim Test Recall

Predictors	β	*SE*	*t*	*p*	R^2	*F*
Step 1: List 5 interim test recall ~ Age + Gender + Interim Task					.236	84.32***
Age	−0.166	0.139	−1.197	.232		
Gender	0.316	0.306	1.034	.302		
Interim task	4.387	0.277	15.852	<.001		
Step 2: List 5 interim test recall ~ Age + Gender + Interim Task + Trait Test Anxiety + WMC					.256	10.94***
Age	−0.078	0.138	−0.567	.571		
Gender	0.186	0.306	0.608	.543		
Interim task	4.340	0.274	15.846	<.001		
Trait test anxiety	−0.009	0.019	−0.472	.637		
WMC	0.045	0.010	4.610	<.001		
Step 3: List 5 interim test recall ~ Age + Gender + Interim Task + Trait Test Anxiety + WMC + Interim Task × Trait Test Anxiety + Interim Task × WMC + Trait Test Anxiety × WMC + Interim Task × Trait Test Anxiety × WMC					.250	0.40
Age	−0.087	0.141	−0.614	.539		
Gender	0.188	0.306	0.613	.540		
Interim task	8.696	4.202	2.070	.039		
Trait test anxiety	0.015	0.090	0.165	.869		
WMC	0.062	0.057	1.096	.274		
Interim Task × Trait Test Anxiety	−0.115	0.127	−0.903	.367		
Interim Task × WMC	−0.079	0.08	−0.977	.329		
Trait Test Anxiety × WMC	<.001	0.002	−0.191	.848		
Interim Task × Trait Test Anxiety × WMC	0.002	0.002	0.833	.405		

Note. The *F* values in the right most column report the model comparison results. Specifically, the column lists the *F* values for Step 1 model vs. null model, Step 2 model vs. Step 1 model, and Step 3 model vs. Step 2 model, respectively.
*** $p < .001$.

between interim task and WMC was significant, $\beta = -0.180$, $p = .006$, confirming the above finding that individuals with low WMC benefitted more from retrieval practice in the List 1–4 cumulative test than did those with high WMC. In addition, there was a two-way interaction between state test anxiety and WMC ($\beta = -0.026$, $p = .018$). However, there was no three-way interaction among interim task, state test anxiety, and WMC ($\beta = 0.025$, $p = .161$).

Overall, the above results demonstrate that interim testing was more effective in enhancing List 1–4 cumulative test recall for low WMC participants than for high WMC ones. Trait/state test anxiety appeared not to moderate the retrieval practice effect on studied information, nor did it moderate the effect in conjunction with WMC.

General Discussion

Although testing has been repeatedly shown to be highly effective at helping to consolidate learning, little research as yet has explored individual differences in test-enhanced learning. The current research recruited a large sample of participants to explore the roles of WMC and trait/state test anxiety in the beneficial effects of testing on new learning and its possible converse influences on test anxiety. Table 2 serves as a descriptive summary of the main findings.

Minimal Role of WMC and Trait/State Test Anxiety in the FTE

Several regression analyses showed that neither WMC nor trait test anxiety tended to modulate the FTE. We repeated these analyses but replaced trait test anxiety with state test anxiety regarding the List 5 interim test. Again, state test anxiety and WMC played little role in the FTE. Overall, these results imply that interim testing appears to reliably enhance subsequent learning and retention of new information regardless of trait/state test anxiety and WMC. The absence of any detectable modulating role of WMC in the FTE is consistent with what Pastötter and Frings (2019, Experiment 2) observed.

Even though word list learning and free recall tests have been extensively used to assess the influence of test anxiety on memory in previous studies (e.g., J. H. Mueller, 1978), the level of state test anxiety reported in the current study was relatively low. Hence, it may be premature to conclude that test anxiety and its conjunction with WMC do not modulate the FTE, because the task used in the current study might be unable to evoke sufficient levels of test anxiety to detect their influences. The following analyses speak to this possibility.

We first selected 100 participants whose state anxiety reports regarding the List 5 interim test were highest in the interim test group to form a high anxiety/interim test group ($M = 6.53$, $SD = 1.74$), and 100 participants whose state test anxiety scores were lowest in the interim test group to form a low anxiety/interim test group ($M = 1.00$, $SD = 0$). The same was done for the interim restudy group to form a high anxiety/interim restudy group ($M = 6.88$, $SD = 1.82$) and a low anxiety/interim restudy group ($M = 1.00$, $SD = 0$). Then, a regression analysis was conducted, with WMC, state test anxiety (high vs. low), study method (test vs. restudy) and their products as the independent variables, and with List 5 interim test recall as the dependent variable. Again, the results showed no significant interaction between state test anxiety

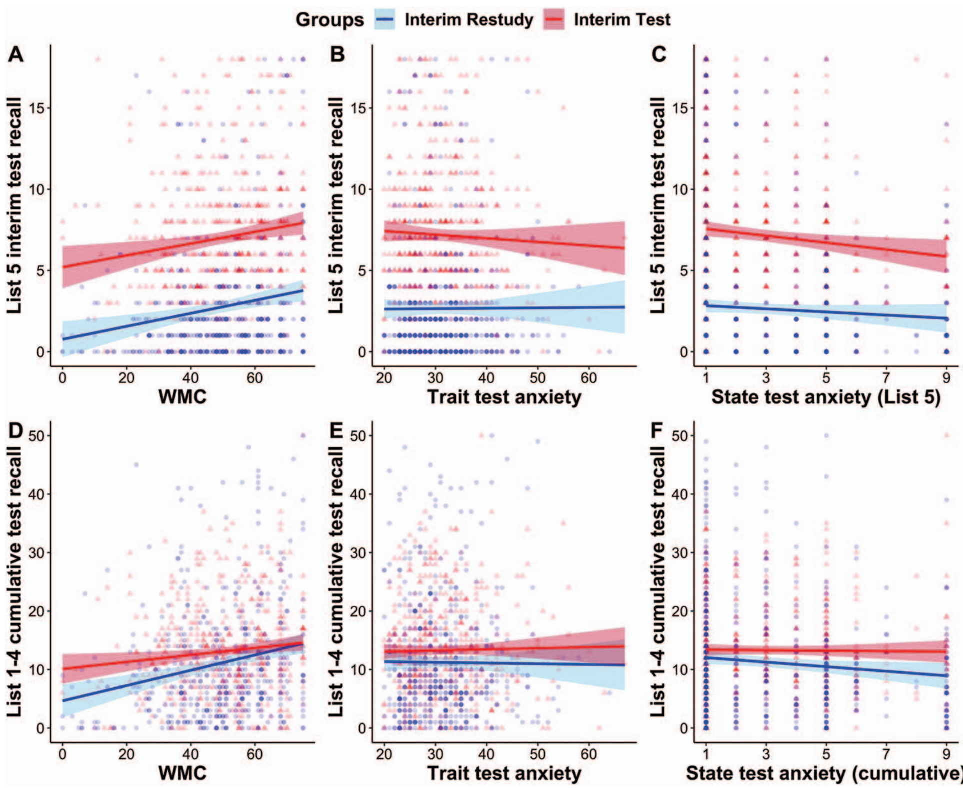

Figure 3. A: Regressions of working memory capacity (WMC) on List 5 interim test recall; B: Regressions of trait test anxiety on List 5 interim test recall; C: Regressions of state test anxiety on List 5 interim test recall; D: Regressions of WMC on cumulative test recall; E: Regressions of trait test anxiety on cumulative test recall; F: Regressions of state test anxiety on cumulative test recall. Error regions represent 95% CI. See the online article for the color version of this figure.

and study method (p = .327), and no three-way interaction among WMC, state test anxiety and study method (p = .292). Overall, these results imply that the FTE is minimally modulated by test anxiety and its interaction with WMC.[7]

WMC but Not Trait/State Test Anxiety Modulates the Retrieval Practice Effect on List 1–4 Cumulative Test Recall

Linear regression analyses demonstrated that in the List 1–4 cumulative test, low WMC participants tended to benefit more from interpolated testing than high WMC participants. These results are consistent with Agarwal et al.'s (2017) proposal that retrieval practice benefits retention of studied information more for individuals with low WMC. There are at least two possible explanations for this phenomenon: episodic memory and attentional processing.

Episodic memory refers to memory of specific episodes from personally encountered events (Herlitz et al., 1997), such as memory about where, when, and how a given item is studied. Many studies have established a positive correlation between episodic memory ability and WMC, showing that individuals with high WMC are typically better able to remember the episodic details of studied items (e.g., Kane & Engle, 2000). Testing is known to boost retention of studied information (at least partially) by conferring context changes, which make studied/tested information more vivid and accessible during retrieval (Karpicke, Lehman, & Aue, 2014). Because individuals with low WMC tend to have poorer episodic memory, more room is available for testing to enrich contextual information of studied materials, leading to a larger retrieval practice effect (Brewer & Unsworth, 2012).

Another possible explanation is attentional processing. Individuals with low WMC are less able to sustain their attention during a prolonged study phase (Miller, Gross, & Unsworth, 2019) and are more susceptible to mind wandering (i.e., focusing on things unrelated to the ongoing task; McVay & Kane, 2012). Numerous behavioral and neural studies have shown convincing evidence that testing can boost retention through enhancing learning engagement (for reviews, see van den Broek et al., 2016; Yang, Potts,

[7] Similarly, according to state test anxiety reports regarding the cumulative test, four groups of participants (with 100 in each group) were selected: high anxiety/interim test (M = 7.25, SD = 1.70), low anxiety/interim test (M = 1.00, SD = 0), high anxiety/interim restudy (M = 6.79, SD = 1.62), and low anxiety/interim restudy (M = 1.00, SD = 0). A regression analysis, with List 1–4 cumulative test recall as the dependent variable, again showed no significant interaction between state test anxiety and study method (p = .220), and no three-way interaction amongst WMC, state test anxiety, and study method (p = .783). These results imply that the retrieval practice effect is unlikely to be modulated by test anxiety and its interaction with WMC.

Table 4
Regression Coefficients of Interim Task, Working Memory Capacity (WMC), and State Test Anxiety on List 5 Interim Test Recall

Predictors	β	*SE*	*t*	*p*	R^2	*F*
Step 1: List 5 interim test recall ~ Age + Gender + Interim Task					.236	84.32***
Age	−0.166	0.139	−1.197	.232		
Gender	0.316	0.306	1.034	.302		
Interim task	4.387	0.277	15.852	<.001		
Step 2: List 5 interim test recall ~ Age + Gender + Interim Task + State Test Anxiety + WMC					.260	13.31***
Age	−0.062	0.138	−0.448	.654		
Gender	0.190	0.303	0.628	.530		
Interim task	4.336	0.273	15.888	<.001		
State test anxiety	−0.128	0.058	−2.197	.028		
WMC	0.045	0.010	4.659	<.001		
Step 3: List 5 interim test recall ~ Age + Gender + Interim Task + State Test Anxiety + WMC + Interim Task × State Test Anxiety + Interim Task × WMC + State Test Anxiety × WMC + Interim Task × State Test Anxiety × WMC					.262	0.44
Age	−0.069	0.139	−0.500	.618		
Gender	0.195	0.303	0.641	.521		
Interim task	6.328	1.570	4.031	<.001		
State test anxiety	0.115	0.268	0.430	.667		
WMC	0.065	0.020	3.276	.001		
Interim Task × State Test Anxiety	−0.467	0.432	−1.081	.280		
Interim Task × WMC	−0.038	0.030	−1.271	.204		
State Test Anxiety × WMC	−0.005	0.005	−0.905	.366		
Interim Task × State Test Anxiety × WMC	0.009	0.008	1.058	.291		

Note. State test anxiety reported in this table represents the state test anxiety ratings regarding the List 5 interim test.
*** $p < .001$.

& Shanks, 2018). Because individuals with low WMC are less able to maintain their learning engagement, more room is available for testing to facilitate their attentional encoding, leading to a larger retrieval practice effect.

Overall, there are at least two possible explanations regarding why testing on studied information is more beneficial for individuals with low WMC. To our knowledge, little research has been conducted to assess them. Our understanding about this important

Table 5
Regression Coefficients of Interim Task, Working Memory Capacity (WMC), and Trait Test Anxiety on List 1–4 Cumulative Test Recall

Predictors	β	*SE*	*t*	*p*	R^2	*F*
Step 1: List 1–4 cumulative test recall ~ Age + Gender + Interim Task					.062	18.12***
Age	−0.564	0.303	−1.859	.063		
Gender	3.977	0.669	5.942	<.001		
Interim task	2.418	0.606	3.989	<.001		
Step 2: List 1–4 cumulative test recall ~ Age + Gender + Interim Task + Trait Test Anxiety + WMC					.081	8.12***
Age	−0.395	0.304	−1.298	.195		
Gender	3.748	0.672	5.576	<.001		
Interim task	2.335	0.602	3.878	<.001		
Trait test anxiety	−0.024	0.041	−0.593	.553		
WMC	0.085	0.022	3.935	<.001		
Step 3: List 1–4 cumulative test recall ~ Age + Gender + Interim Task + Trait Test Anxiety + WMC + Interim Task × Trait Test Anxiety + Interim Task × WMC + Trait Test Anxiety × WMC + Interim Task × Trait Test Anxiety × WMC					.089	1.85
Age	−0.350	0.309	−1.133	.257		
Gender	3.724	0.671	5.548	<.001		
Interim task	11.069	9.203	1.203	.229		
Trait test anxiety	−0.027	0.197	−0.137	.891		
WMC	0.157	0.125	1.260	.208		
Interim Task × Trait Test Anxiety	−0.105	0.278	−0.377	.706		
Interim Task × WMC	−0.211	0.176	−1.199	.231		
Trait Test Anxiety × WMC	−0.001	0.004	−0.171	.864		
Interim Task × Trait Test Anxiety × WMC	0.003	0.005	0.624	.533		

*** $p < .001$.

Table 6
Regression Coefficients of Interim Task, Working Memory Capacity (WMC), and State Test Anxiety on List 1–4 Cumulative Test Recall

Predictors	β	*SE*	*t*	*p*	R^2	*F*
Step 1: List 1–4 cumulative test recall ~ Age + Gender + Interim Task					.062	18.12^{***}
Age	−0.564	0.303	−1.859	.063		
Gender	3.977	0.669	5.942	<.001		
Interim task	2.418	0.606	3.989	<.001		
Step 2: List 1–4 cumulative test recall ~ Age + Gender + Interim Task + State Test Anxiety + WMC					.081	8.31^{***}
Age	−0.391	0.304	−1.286	.199		
Gender	3.718	0.667	5.571	<.011		
Interim task	2.321	0.601	3.860	<.001		
State test anxiety	−0.109	0.129	−0.849	.396		
WMC	0.086	0.022	3.984	<.001		
Step 3: List 1–4 cumulative test recall ~ Age + Gender + Interim Task + State Test Anxiety + WMC + Interim Task × State Test Anxiety + Interim Task × WMC + State Test Anxiety × WMC + Interim Task × State Test Anxiety × WMC					.088	3.02^{*}
Age	−0.408	0.304	−1.341	.180		
Gender	3.721	0.665	5.600	<.001		
Interim task	11.630	3.438	3.382	.001		
State test anxiety	1.241	0.587	2.112	.035		
WMC	0.214	0.044	4.903	<.001		
Interim Task × State Test Anxiety	−1.332	0.947	−1.406	.160		
Interim Task × WMC	−0.180	0.065	−2.775	.006		
State Test Anxiety × WMC	−0.026	0.011	−2.369	.018		
Interim Task × State Test Anxiety × WMC	0.025	0.018	1.404	.161		

Note. State test anxiety reported in this table represents the state test anxiety ratings regarding the cumulative test.
$^{*}p < .05$. $^{***}p < .001$.

finding is still in its infancy, and more research on this issue is needed.

The current study found that neither state nor trait test anxiety tend to modulate retrieval practice effects on studied information, which is inconsistent with Tse and Pu's (2012) finding (that is, a negative correlation between test anxiety and BTE for low WMC individuals) and Clark et al.'s (2018) finding (i.e., a positive correlation between test anxiety and the BTE). One difference between the current and previous studies (Clark et al., 2018; Tse & Pu, 2012) is worth highlighting: The current study did not ascertain whether the retrieval practice effect on List 1–4 cumulative test recall came from either the BTE or FTE (or a combination of both); by contrast, both Tse and Pu (2012) and Clark et al. (2018) specifically focused on the role of test anxiety in the BTE. Hence, it is premature to firmly conclude that the above findings conflict with those of Clark et al. and Tse and Pu. Future research should further explore the role of test anxiety in the BTE to resolve the inconsistent findings of Tse and Pu (2012) and Clark et al. (2018).

It is worth noting that previous research (including the current study) has largely focused on individual differences in short-term enhanced retention induced by testing, with only one exception (Agarwal et al., 2017) which explored the role of WMC in the BTE with a long-term interval (2 days) inserted between study and final test. The long-term modulating effects of WMC and test anxiety in the FTE await future investigation.

Little Influence of Interpolated Testing on State Test Anxiety

Bayesian *t* tests on state test anxiety reports regarding the List 5 interim and cumulative tests consistently found a null influence of interim tests on state test anxiety, which is inconsistent with the findings from previous studies (Agarwal et al., 2014; Szpunar et al., 2013). In Szpunar et al.'s interim tests on lecture segments, their test group's recall accuracies were about 90%. Szpunar et al. proposed that their test group reported lower test anxiety as a result of "positive feedback from earlier tests" (p. 4). Specifically, the interim tests in Szpunar et al.'s study were relatively easy and the high test accuracy might have reduced test anxiety. By contrast, the interim tests in the current study were relatively difficult, as reflected by low accuracy: Only about 40% (i.e., 7 out of 18) of words were correctly recalled. Hence, one possible reason for the null influence of interpolated testing on test anxiety may be that the interim tests were too difficult in the current study.

We conjecture, however, that this test difficulty account cannot easily explain these divergent findings for at least three reasons. First, the state test anxiety reports for both groups (about 3 on a 9-point scale) were lower than the medium point (5), indicating that low test performance did not induce high state test anxiety for either group. Second, although the interim test group performed substantially better in the list 5 interim test than the interim restudy group, there was no difference in state test anxiety reports regarding the cumulative test, reflecting the fact that test performance had very little influence on immediate test anxiety. Third, as shown in Table 1 (correlation matrix), the correlation between List 5 interim test recall and test anxiety reports regarding the List 5 interim test was quite weak, $r = -.078$, $p = .012$, and the same was true for cumulative test recall and the test anxiety reports regarding the cumulative test, $r = -.059$, $p = .059$. Along the same lines, the correlation between List 5 interim test recall and the test anxiety reports regarding the cumulative test was modest,

$r = -.054$, $p = .081$, again implying little influence of test performance on state test anxiety.

Considering that the current study included a large sample size and the Bayesian results strongly support the null difference in state test anxiety between groups, we propose that frequent tests appear to neither increase nor mitigate test anxiety. Moreover, we conjecture that test difficulty cannot easily explain the divergent findings between the current study and Szpunar et al. (2013). When interpreting the divergent findings on state test anxiety between the current research and that of Szpunar et al. (2013), it should be noted that there were many differences in experimental procedure and stimuli. For instance, Szpunar et al. instructed participants to watch lecture videos whereas the materials in the current research were word lists; their participants only reported state test anxiety before the cumulative test, while in the current research participants reported state test anxiety before both the List 5 interim test and the cumulative test; Szpunar et al. inserted a 5-min break between their Segment 4 interim test and the cumulative test, but no interval was inserted between the List 5 interim test and the cumulative test in the current study. Such subtle divergences might account for the divergent findings on state test anxiety.

It must also be acknowledged of course that there is a substantial difference between the sort of test anxiety induced in the present laboratory learning task (and in Szpunar et al.'s task), on the one hand, and test anxiety associated with potentially life-changing (high-stake) exams taken at school or college, on the other. As shown in the current study, state test anxiety in laboratory-based experiments tends to be relatively low and likely differs from what occurs in high-stake assessments in real educational settings. Many previous studies exploring the relationship between test anxiety and test-enhanced learning in the laboratory (e.g., Clark et al., 2018; Tse & Pu, 2012; the current study) might lack ecological validity. For instance, tests in laboratory experiments have no grades and participants cannot fail them, contrary to what is typical in classroom quizzes and exams. Therefore, the relationship between test anxiety and test-enhanced learning should be further investigated in real educational settings. In addition, future research examining the converse influence of testing on test anxiety in the classroom is urgently needed.

Practical (Educational) Implications

In educational settings, students are frequently instructed to master a large set of class materials. For instance, students in a medical school have to remember names of bones in the human skeleton. The findings from the current study imply that instructors can strategically administer interim tests after teaching each set of facts, which has the potential to boost learning of new facts. Such an enhancing effect will tend to apply equally to individuals with different levels of WMC and test anxiety.

Besides potentiating learning of new facts, interim testing should also be able to facilitate retention of studied information, especially for students with low WMC. Some instructors may have the concern that students with low WMC benefit less or even suffer from testing on studied information as their test performance is generally lower than that of students with high WMC, leading to no reexposure to unrecalled materials, and encouraging the worry that retrieval practice may exacerbate individual differences in academic achievement between students with high and low WMC. However, the current study and that of Agarwal et al. (2017) jointly demonstrate that this is not the case. Instead, individuals with low WMC benefit more from retrieval practice than those with high WMC. Such findings also imply that implementing tests on studied information is a potential technique to narrow, rather than exacerbate, individual differences in learning efficiency based on WMC.

From a positive perspective, despite the possibility that testing may increase test anxiety (Steele, 2011), the current study provides evidence to allay this concern. However, from a negative perspective, our results did not support Szpunar et al.'s (2013) observation that interim tests mitigated test anxiety. Instead, we found that interim tests appear to have minimal influence on test anxiety. Overall, the current study implies that practitioners can administer interim tests without unduly worrying about their effects on test anxiety.

Concluding Remarks

In summary, the takeaway messages are (a) administering interim tests appears to be an efficient technique to potentiate new learning, regardless of WMC and test anxiety; (b) retrieval practice consolidates retention of studied information more effectively for individuals with low than high WMC; (c) neither trait nor state test anxiety substantially modulates testing effects on studied information; and (d) frequent tests tend to neither increase nor mitigate test anxiety.

References

Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements, 11,* 159–177. http://dx.doi.org/10.1037/h0093018

Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory & Cognition, 3,* 131–139. http://dx.doi.org/10.1016/j.jarmac.2014.07.002

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, 25,* 764–771. http://dx.doi.org/10.1080/09658211.2016.1220579

Aslan, A., & Bäuml, K. T. (2016). Testing enhances subsequent learning in older but not in younger elementary school children. *Developmental Science, 19,* 992–998. http://dx.doi.org/10.1111/desc.12340

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language, 66,* 407–415. http://dx.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*. Cambridge, MA: Belknap Press of Harvard University Press.

Bufe, J., & Aslan, A. (2018). Desirable difficulties in spatial learning: Testing enhances subsequent learning of spatial information. *Frontiers in Psychology, 9,* 1701. http://dx.doi.org/10.3389/fpsyg.2018.01701

Carriere, J. S. A., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67,* 19–31. http://dx.doi.org/10.1037/a0031438

Chan, J. C., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag. *Journal of*

Memory and Language, 102, 83–96. http://dx.doi.org/10.1016/j.jml.2018.05.007

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin, 144,* 1111–1146. http://dx.doi.org/10.1037/bul0000166

Clark, D. A., Crandall, J. R., & Robinson, D. H. (2018). Incentives and test anxiety may moderate the effect of retrieval on learning. *Learning and Individual Differences, 63,* 70–77. http://dx.doi.org/10.1016/j.lindif.2018.03.001

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. London, UK: Routledge.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

Haimovitz, K., & Dweck, C. S. (2016). Parents' views of failure predict children's fixed and growth intelligence mind-sets. *Psychological Science, 27,* 859–869. http://dx.doi.org/10.1177/0956797616639727

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58,* 47–77. http://dx.doi.org/10.3102/00346543058001047

Herlitz, A., Nilsson, L.-G., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25,* 801–811. http://dx.doi.org/10.3758/BF03211324

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology, 28,* 597–606. http://dx.doi.org/10.1002/acp.3032

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). The median split: Robust, refined, and revived. *Journal of Consumer Psychology, 25,* 690–704. http://dx.doi.org/10.1016/j.jcps.2015.06.014

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied, 22,* 305–318. http://dx.doi.org/10.1037/xap0000087

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358. http://dx.doi.org/10.1037/0278-7393.26.2.336

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

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. *Psychology of Learning and Motivation, 61,* 237–284. http://dx.doi.org/10.1016/B978-0-12-800283-4.00007-1

Lee, H. S., & Ahn, D. (2018). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology, 110,* 203–217. http://dx.doi.org/10.1037/edu0000211

Liben-Nowell, D., Strand, J., Sharp, A., Wexler, T., & Woods, K. (2019). The danger of testing by selecting controlled subsets, with applications to spoken-word recognition. *Journal of Cognition, 2,* 2. http://dx.doi.org/10.5334/joc.51

Liu, Y., & Reichle, E. D. (2017). Eye-movement evidence for object-based attention in Chinese reading. *Psychological Science, 29,* 278–287. http://dx.doi.org/10.1177/0956797617734827

Maxwell, S. E., & Delaney, H. D. (1993). Bivariate median splits and spurious statistical significance. *Psychological Bulletin, 113,* 181–190. http://dx.doi.org/10.1037/0033-2909.113.1.181

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141,* 302–320. http://dx.doi.org/10.1037/a0025250

Miller, A. L., Gross, M. P., & Unsworth, N. (2019). Individual differences in working memory capacity and long-term memory: The influence of intensity of attention to items at encoding as measured by pupil dilation. *Journal of Memory and Language, 104,* 25–42. http://dx.doi.org/10.1016/j.jml.2018.09.005

Mok, W. S. Y., & Chan, W. W. L. (2016). How do tests and summary writing tasks enhance long-term retention of students with different levels of test anxiety? *Instructional Science, 44,* 567–581. http://dx.doi.org/10.1007/s11251-016-9393-x

Mueller, J. H. (1978). The effects of individual differences in test anxiety and type of orienting task on levels of organization in free recall. *Journal of Research in Personality, 12,* 100–116. http://dx.doi.org/10.1016/0092-6566(78)90087-9

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70,* 1–12. http://dx.doi.org/10.1016/j.jml.2013.09.007

Nordstrand, D. (2018). *Test-enhanced learning, working memory, and difficulty of material* (Master thesis). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-149176 DiVA database.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M. A., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U. S. Department of Education. Retrieved from https://ies.ed.gov

Pastötter, B., & Bäuml, K. H. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5,* 286. http://dx.doi.org/10.3389/fpsyg.2014.00286

Pastötter, B., & Bäuml, K. T. (2019). Testing enhances subsequent learning in older adults. *Psychology and Aging, 34,* 242–250. http://dx.doi.org/10.1037/pag0000307

Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of Cognition, 2,* 37. http://dx.doi.org/10.5334/joc.82

Pastötter, B., Weber, J., & Bäuml, K. H. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology, 27,* 280–285. http://dx.doi.org/10.1037/a0031797

Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L., III. (2016). Optimizing learning in college: Tips from cognitive psychology. *Perspectives on Psychological Science, 11,* 652–660. http://dx.doi.org/10.1177/1745691616645770

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 615–625. http://dx.doi.org/10.1037/a0013684

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? *Journal of Experimental Psychology: Applied, 23,* 293–300. http://dx.doi.org/10.1037/xap0000134

Spielberger, C. D. (1980). *Test Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Steele, D. (2011, Jan 21). Letter to the Editor: Does test-taking help students learn? *New York Times,* p. A24. Retrieved from https://www.nytimes.com/2011/01/26/opinion/l26test.html

Sullivan, D. (2017). Mediating test anxiety through the testing effect in asynchronous, objective, online assessments at the university level. *Journal of Education and Training, 4,* 107–123. http://dx.doi.org/10.5296/jet.v4i2.10777

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences, 110,* 6313–6317. http://dx.doi.org/10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399. http://dx.doi.org/10.1037/a0013082

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,* 253–264. http://dx.doi.org/10.1037/a0029190

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505. http://dx.doi.org/10.3758/BF03192720

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Karlsson Wirebring, L., Segers, E., Verhoeven, L., & Nyberg, L. (2016). Neurocognitive mechanisms of the “testing effect”: A review. *Trends in Neuroscience and Education, 5,* 52–66. http://dx.doi.org/10.1016/j.tine.2016.05.001

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1039–1048. http://dx.doi.org/10.1037/a0036164

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review, 18,* 518–523. http://dx.doi.org/10.3758/s13423-011-0085-x

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology, 55,* 10–16. http://dx.doi.org/10.1111/sjop.12093

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4,* 433. http://dx.doi.org/10.3389/fpsyg.2013.00433

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18,* 1140–1147. http://dx.doi.org/10.3758/s13423-011-0140-7

Yang, C., Chew, S.-J., Sun, B., & Shanks, D. R. (2019). The forward effects of testing transfer to different domains of learning. *Journal of Educational Psychology, 111,* 809–826. http://dx.doi.org/10.1037/edu0000320

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language, 99,* 99–110. http://dx.doi.org/10.1016/j.jml.2017.11.005

Yang, C., Potts, R., & Shanks, D. R. (2017). The forward testing effect on self-regulated study time allocation and metamemory monitoring. *Journal of Experimental Psychology: Applied, 23,* 263–277. http://dx.doi.org/10.1037/xap0000122

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *npj Science of Learning, 3,* 8. http://dx.doi.org/10.1038/s41539-018-0024-y

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 485–492. http://dx.doi.org/10.1037/xlm0000449

Yue, C. L., Soderstrom, N. C., & Bjork, E. L. (2015). Partial testing can potentiate learning of tested and untested material from multimedia lessons. *Journal of Educational Psychology, 107,* 991–1005. http://dx.doi.org/10.1037/edu0000031

Yue, X. (1996). Test anxiety and self-efficacy: Levels and relationship among secondary school students in Hong Kong. *Psychologia: An International Journal of Psychology in the Orient, 39,* 193–202.

Received September 28, 2019
Revision received March 13, 2020
Accepted March 20, 2020 ■