# Do Practice Tests (Quizzes) Reduce or Provoke Test Anxiety? A Meta-Analytic Review

## Author Note

All data are publicly available at OSF, and the web-link is provided on the cover page.

## Abstract

Practice testing is a powerful tool to consolidate long-term retention of studied information, facilitate subsequent learning of new information, and foster knowledge transfer. However, practitioners frequently express the concern that tests are anxiety-inducing and that their employment in the classroom should be minimized. The current review integrates results across 24 studies (i.e., 25 effects based on 3,374 participants) to determine the effect of practice tests (quizzes) on test anxiety (TA) and explore potential moderators of the effect. The results show strong Bayesian evidence ($BF_{10}$ > 25,000) that practice tests appreciably reduce TA to a medium extent (Hedges' $g$ = -0.52), with minimal evidence of publication bias. Easy practice tests tend to be more effective in mitigating TA than difficult ones. These findings support a recommendation for instructors to incorporate quizzes into their curriculum. However, instructors should be aware that quizzes themselves may be more stressful than other learning activities. Methods to make quizzes less stressful and more enjoyable are discussed. Research on the effect of practice tests on TA is still in its infancy, and future research directions are highlighted.


***Keywords***: Practice testing; Test anxiety; Test difficulty; Quiz pressure; Meta-analysis

What interventions are available to enhance students' academic learning and performance? The search for such interventions has long been an important pursuit for researchers and practitioners (including students, parents, instructors, and educational policy-makers). A promising principle from hundreds of studies in cognitive and educational psychology is that practice testing (i.e., practice retrieval) is one of the most effective strategies to consolidate long-term retention of studied information and facilitate subsequent learning of new information, a phenomenon labeled the *testing effect*, the *retrieval practice effect*, or *test-enhanced learning* (Carpenter et al., 2022; Pan & Rickard, 2018; Roediger & Butler, 2011; Shanks et al., 2023; Yang et al., 2021). It has been firmly established that retrieval practice is more beneficial by comparison with many other learning strategies, such as restudying (Roediger & Karpicke, 2006b), note-taking (Heitmann et al., 2018; Rummer et al., 2017), concept-mapping (Karpicke & Blunt, 2011) and other elaborative strategies (Larsen et al., 2013).

Although the mnemonic benefits of practice testing are substantial, researchers frequently express dismay that testing (quizzing) has not been more fully adopted and embedded in educational settings to boost students' attainment (Roediger & Karpicke, 2006b; Yang et al., 2021). Policy-makers even endeavor to minimize the use of tests in the classroom. For instance, President Barack Obama signed the Every Student Succeeds Act (2015) which included guidelines to cap the amount of testing permitted in US schools (i.e., no more than 2% of class time; Zernike, 2015). A potential obstacle for implementing tests in the classroom is the legitimate concern that tests may provoke test anxiety (TA), which in turn impedes students' academic performance. However, there are also reasons to believe that practice tests may reduce TA. For instance, tests may inform students about the formats and contents of later assessments, reduce uncertainty and hence alleviate TA (see below for detailed discussion).

Over the past few decades, some studies have been conducted to investigate whether practice tests (quizzes) attenuate or increase TA, but the findings have been inconsistent and fragmented, and little clear guidance has emerged for practitioners. The current article primarily aims to conduct a meta-analysis of previous findings in order to (1) determine the effect of practice tests on TA, and (2) explore potential moderators (and boundary conditions) of any such effects.

Below, we first briefly review empirical evidence about test-enhanced learning to highlight the merits of tests in promoting learning. Next, we introduce TA and its detrimental effects, and then review previous findings about the effect of practice tests (quizzes) on TA. Finally, we outline the rationale of the current meta-analysis.

**Test-enhanced learning**

Although testing has been widely-viewed as an assessment *of* learning, it has also been established as an assessment *for* learning (Roediger & Karpicke, 2006a; Yang et al., 2021). Specifically, numerous studies observed that, by comparison with many other strategies (e.g., restudying, note-taking, concept-mapping, self-explaining), testing can not only more effectively consolidate long-term retention of studied information, a phenomenon termed the *backward testing effect* (for reviews, see Roediger & Karpicke, 2006a; Rowland, 2014), but also more effectively facilitate subsequent learning of new information, known as the *forward testing effect* (for reviews, see Chan et al., 2018; Pastötter & Bäuml, 2014; Shanks et al., 2023; Yang, Potts, et al., 2018).

Roediger and Karpicke (2006b) provided a clear demonstration of the backward testing effect. In this study, participants studied two texts, with one text studied once and tested once and the other studied twice in succession. In a cumulative free recall test administered one week later, the tested text was recalled substantially better than the restudied one, reflecting a backward testing effect (that is, testing consolidates retention of the studied materials). Szpunar et al. (2013) provided an illustration of the forward testing effect. In this study, two groups (a test group and a no-test group) of participants watched a lecture video which was divided into four segments. After watching each of Segments 1-3, the test group took an interim test on the just-studied segment, whereas the no-test group undertook a distractor task. After watching Segment 4, both groups took a test on this segment. Szpunar et al. found that performance on the Segment 4 test was significantly better in the test than in the no-test group, demonstrating a forward testing effect (that is, practice tests on Segments 1-3 promoted subsequent learning of Segment 4). Furthermore, not only post-testing (i.e., taking a test on studied information) but also pre-testing (i.e., taking a test on to-be-studied information) can facilitate learning (Brown & Tallon, 2015; Pan et al., 2020).

Test-enhanced learning has been repeatedly observed not only in laboratory but also in classroom settings (e.g., Bates, 2019; Leeming, 2002). In a recent meta-analytic review, Yang et al. (2021) integrated data from over 48,000 students, extracted from 222 classroom studies, to determine whether class quizzes improve students' academic performance. The answer is affirmative: Class quizzes enhance students' academic attainment to a medium extent (Hedges' $g = 0.50$). The classroom testing effect generalizes to students across different educational levels (including elementary school, middle school, high school, and university/college), and across 18 subject categories (e.g., Education, Medicine, Psychology, etc.). More importantly, the results showed that classroom quizzes not only benefit retention of factual knowledge, but also promote concept comprehension and facilitate knowledge transfer in the service of solving applied problems. Test-enhanced knowledge transfer has also been observed in many other studies (for a review, see Carpenter, 2012).

Overall, the beneficial effects of testing on learning are substantial in both laboratory and classroom contexts, and practice testing can be utilized as an effective instrument to enhance learning. The cognitive underpinnings of the testing effect are beyond the scope of the current article, and interested readers can consult recent reviews (e.g., Adesope et al., 2017; Chan et al., 2018; Pan & Rickard, 2018; Shanks et al., 2023; Yang et al., 2021; Yang, Potts, et al., 2018).

**TA and its detrimental effects**

TA is a widespread and mostly detrimental emotion (e.g., worry or fear) experienced before, during and after evaluation situations (Spielberger, 1980). For instance, students frequently experience high TA when they feel they are unprepared for an evaluation (McDonald, 2001). Students struggling with TA tend to worry about poor performance, fear of letting down others (e.g., parents and teachers), and demonstrate severe somatic anxiety symptoms (e.g., shortness of breath, muscle tension, and rapid heartbeat) when faced with test situations (Zeidner, 1998, 2007). TA is measured via one of a large number of standardized instruments such as the Test Anxiety Scale (Spielberger, 1980).

The deleterious effects of TA span cognitive and performance domains. Since the 1950s (Mandler & Sarason, 1952), a large body of research has observed detrimental effects of TA on academic performance. For instance, Ali and Mohsin (2013) found that TA negatively predicted students' academic performance in many subjects, including Physics ($r$ = -.19), Chemistry ($r$ = -.20), Biology ($r$ = -.19), and Mathematics ($r$ = -.22). In an early meta-analysis which integrated results across 126 studies, Seipp (1991) observed a negative correlation ($r$ = -.21) between anxiety and academic performance. This pattern was confirmed in a recent and even larger (238 studies) meta-analysis by von der Embse et al. (2018) who demonstrated that TA negatively predicted a variety of educational outcomes, including grade point average and university entrance exam performance. Furthermore, Spada and Moneta (2014) found that TA relates to students' learning styles: High TA students are more likely to employ superficial encoding strategies. It should be noted that TA also negatively relates to students' subjective well-being and life satisfaction (Tahoon, 2021). Students with high TA generally report high levels of depression (Akinsola & Nwajei, 2013) and low levels of self-esteem (von der Embse et al., 2018) and subjective well-being (Steinmayr et al., 2016).

Overall, high TA undermines students' learning and academic performance. It is hence of critical importance to determine whether practice testing (quizzing), as an effective learning tool, provokes or reduces TA.

**Conflicting evidence on the effect of practice tests on TA**

Students may regard tests as threatening, stressful, and undesirable (Cardozo et al., 2020; Cassady, 2004; Haleem & Saeed, 2021; Hinze & Rapp, 2014). For instance, in a letter to the editor of *The New York Times*, Steele (2011) wrote that "*As a college student… I personally would like to avoid as many tests as possible… Trying to learn in a stressful environment is no way to help retain information… Why put more stress and anxiety on students who already have a lot on their plates?*"

Indeed, it is reasonable to expect that tests or quizzes may provoke TA. For instance, practice testing is generally classified as a kind of "desirable difficulty" (Bjork et al., 2014). Desirable difficulties are factors which make the learning process more difficult (i.e., less fluent) but are nonetheless "desirable" because they motivate learners to process study materials more elaboratively and thoroughly, producing superior learning gains (Bjork & Bjork, 2011; Bjork et al., 2014). Testing (i.e., retrieving information from memory), as an example of desirable difficulty, may therefore overwhelm students and induce anxiety (O'Neil Jr et al., 1969). Moreover, in educational settings, individuals' test performance is typically under evaluation from teachers, peers, and care-givers, and students frequently worry about performing poorly in assessments (Wenzel & Reinhard, 2021). Fear of failure or anticipation of frustration at not achieving desired learning goals can induce anxiety (Cardozo et al., 2020; Downing et al., 2020).

Even though practitioners frequently express the concern that excessive tests create stress and worry among students, there are also many reasons to believe that practice tests (quizzes) may alleviate TA. For instance, it is well-known that tests motivate students to study harder (Yang et al., 2017a), encourage them to read the assigned textbook materials to prepare for the lecture (Heiner et al., 2014), reduce mind wandering while learning (Szpunar et al., 2013), and increase class attendance (Schrank, 2016). These beneficial effects of practice tests may make students more prepared for tests and reduce their worry about poor test performance, therefore alleviating TA (Brown & Tallon, 2015; Yusefzadeh et al., 2019). Furthermore, tests may inform students about the formats and contents of future assessments, hence reducing uncertainty (i.e., uncertainty about how and what content will later be assessed) and mitigating anxiety (Jerrell & Betty, 2005).

Some empirical studies did observe that practice tests (or quizzes) can reduce TA. For instance, in a quasi-experimental study conducted by Piroozmanesh and Imanipour (2018), two classes of nursing undergraduates took a coronary care course, with the experimental class taking class quizzes across the semester, whereas the control class did not take these quizzes. For both classes, students' TA was measured at the beginning of the semester (pretest) and one week before the final exam (posttest). The results showed that although there was minimal difference in TA during the pretest between the two classes, students in the experimental class were much less anxious before the final exam than those in the control class. In another classroom study, Brown and Tallon (2015) observed that pre-lecture quizzes (i.e., quizzes including five multiple-choice questions relating to to-be-taught materials before a lecture) significantly reduced students' TA before exams.

In the laboratory study described previously, Szpunar et al. (2013) obtained consistent findings. Specifically, after both the test and the no-test group completed the interim test on Segment 4, both groups were told that they would take a cumulative test on all four segments and were instructed to report how anxious they were about the cumulative test. Consistent with the findings from Piroozmanesh and Imanipour (2018) and Brown and Tallon (2015), Szpunar et al. (2013) observed that participants in the test group were much less anxious than those in the no-test group.

Besides these empirical studies, a set of questionnaire studies measured students' beliefs about the effect of practice tests on TA. For instance, Agarwal et al. (2014) asked 1,404 middle and high school students to answer a question about "*whether clicker quizzes made you more or less nervous for unit tests*". They were provided with three response options: "*more*", "*less*", and "*about the same*". Most (72%) students reported that clicker quizzes made them less nervous about unit tests, and only 6% chose the option “*more*”, with the remaining 22% choosing “*about the same*”. These findings reveal that a majority of adolescent students believe that class quizzes appreciably alleviate TA.

Although all the aforementioned findings support the claim that practice tests reduce TA, it has to be highlighted that other studies observed minimal influence of tests on TA. For instance, in an Educational Psychology course, Denny et al. (1964) explored the effect of testing on TA. Students in the testing condition took a quiz after each weekly lecture, whereas those in the review condition were provided a narrative review sheet of the lecture content to restudy. Before the unit exam, all students reported their levels of TA. At variance with the studies discussed above, Denny and colleagues observed no treatment (study method) effect on students’ TA. In another laboratory study, Yang et al. (2020) employed the same multi-section learning procedure as Szpunar et al. (2013) to explore the effect of practice tests on TA. With a large sample size (i.e., over 1,000 participants), Yang et al. observed Bayesian evidence ($BF_{10}$ = 0.07) supporting the null hypothesis (i.e., no influence of practice tests on TA). Similar null findings have also been documented by other studies (e.g., Elskamp, 2020; Haleem & Saeed, 2021; McKenzie & Henry, 1979; Wong & Lim, 2021).

Overall, practitioners tend to believe that practice tests are anxiety-provoking. However, many studies have demonstrated the exact opposite pattern by showing that tests reduce TA. At the same time, other studies showed minimal influence.

**The current review**

Although dozens of literature and meta-analytic reviews have assessed the effects of testing on learning and academic performance (e.g., Adesope et al., 2017; Carpenter et al., 2022; Chan et al., 2018; Pan & Rickard, 2018; Shanks et al., 2023; Yang et al., 2021; Yang, Potts, et al., 2018), to our knowledge no reviews (either narrative or meta-analytic) have been conducted to evaluate the research base of the effect of practice tests on TA. The current review aims to fill this important gap. Furthermore, as discussed above, previous findings about the effect of tests on TA are inconsistent, which is unhelpful for practitioners and may make instructors hesitate to implement quizzes in their courses. Hence, it is important to conduct a meta-analysis to integrate existing results to clarify whether practice tests on average alleviate or provoke TA.

Another underexplored (or even unexplored) question concerns which factors moderate the effect of practice tests on TA. This research question is of considerable importance for at least two reasons. First, it is important to inform practitioners under which conditions and what kinds of tests reduce (or provoke) TA. Second, exploring potential moderators can also sharpen our understanding about why previous findings are divergent. Hence, the second goal of the current review is to identify potential moderators of the effect. In total, five potential moderators were evaluated here, including research setting, measurement type, practice test format, practice test performance, and publication status.

***Research setting***: Many of the included studies were conducted in the real classroom (e.g., Denny et al., 1964; Piroozmanesh & Imanipour, 2018), with several others performed in laboratory settings (e.g., Szpunar et al., 2013; Yang et al., 2020). Classroom and laboratory studies differ in many aspects (for detailed comparison, see Yang et al., 2021). For instance, in laboratory studies, participants are typically less motivated to perform the learning task well because their test performance generally has minimal influence on task compensation (Kang & Pashler, 2014). By contrast, in classroom studies, students are motivated to achieve high grades and their test performance is frequently under evaluation by teachers and parents (Wenzel & Reinhard, 2021). Given these critical divergences between laboratory and classroom research, it is important to assess whether research setting moderates the effect of practice tests on TA.

***Measurement type***: In many of the included studies, TA was measured by standardized questionnaires or scales, composed of dozens of measurement items. For instance, Powell (2002) used the Phobos inventory to measure students' anxiety about mathematics tests (Ferguson, 1986), and Haleem and Saeed (2021) adopted the Test Anxiety Inventory (Spielberger, 1980) to quantify students' TA in a Science Education course. By contrast, in some other studies, TA was measured by a single or multiple questions developed by the researchers themselves (e.g., Brown & Tallon, 2015; Dustin, 1971; Szpunar et al., 2013; Yang et al., 2020). For instance, before taking a cumulative test on all segments, Szpunar et al. (2013) asked participants to report how anxious they were about the cumulative test on a scale ranging from 1 to 7.

It is well-known that standardized measures are generally more reliable and less susceptible to measurement error than non-standardized ones. Hence, it is important to explore if the effect of practice tests on TA is moderated by measurement type.

***Practice format***: The included studies employed different formats of practice tests or quizzes, such as free recall (Yang et al., 2020), short answer (Szpunar et al., 2013), multiple-choice (Mayers, 2004), and so on. It is well-established that recall tests, by comparison with recognition tests, tend to be more challenging and are associated with poorer test performance (McDaniel et al., 2007; Sax &

Collet, 1968; Yang et al., 2021). Accordingly, we hypothesized that difficult recall tests might be less effective in reducing TA than easy recognition tests (Higham et al., 2022).

***Practice test performance***: To foreshadow, the results concerning the moderating effect of practice test format (i.e., practice test difficulty) were inconclusive (see below for details). A possible reason is that the number of included effects for recall and recognition tests might be too small to provide sufficient statistical power to detect a significant difference. To further explore the potential moderating role of test difficulty, we conducted another meta-regression analysis in which we regressed the effect sizes onto test performance (i.e., performance in the practice tests). The logic of taking practice test performance as an index of test difficulty is straightforward: the easier the practice test, the superior the test performance.

As discussed above, we hypothesized that easy practice tests should be more effective in reducing TA than difficult ones. Put differently, we predicted that practice performance negatively predicts the effect sizes of practice tests on TA.

***Publication status***: Most of the included effects were extracted from published journal articles (e.g., Szpunar et al., 2013; Yang et al., 2020), with several others coming from unpublished dissertations (e.g., Elskamp, 2020). It is well-known that studies with non-significant findings (i.e., small effect sizes) generally face a higher publication barrier than those with statistically significant findings (i.e., large effect sizes) (Franco et al., 2014). Hence, if the included studies suffered from publication bias, the magnitude of published effects should be larger than that of unpublished ones.

# Method

## Literature search

To obtain a comprehensive set of eligible studies, we conducted a systematic search in 28 electronic databases, including Web of Science, ProQuest (composed of 26 databases, including PsychArticles, PsychInfo, Psychology Database, Education Database, ProQuest Dissertations & Theses Global Database, Ebook Central, Business Market Research Collection, and others), and Google Scholar. The search terms were [(practice test* OR quiz* OR exam* OR retrieval practice) AND (test anxiety OR test-anxiety)]. Note that Google Scholar returned over 2,950,000 results, and we only screened the first 1,000. In addition, the reference lists and Google Scholar citations of the included studies and those of Agarwal et al. (2014) were manually screened to identify additional studies. The literature search was performed on 10th August, 2022.

**Inclusion and Exclusion Criteria**

1. Only experimental studies which directly measured the effect of practice tests (quizzes) on TA were included. Questionnaire studies which measured students' beliefs about the effect of practice tests on TA were excluded (e.g., Agarwal et al., 2014), because it is well-known that individuals' beliefs may be susceptible to a variety of illusions and sometimes do not align with reality (Rhodes & Castel, 2008, 2009; Yang, Huang, et al., 2018; Yang et al., 2017b).
2. Two studies explored whether individuals perceive practice testing as more stressful than restudying (Hinze & Rapp, 2014; Wenzel & Reinhard, 2021). These two studies were excluded from the meta-analysis because there were many fundamental divergences between them and the included studies.[1] These two studies are reviewed in the General Discussion.

[1] For instance, both of these studies were conducted in a laboratory, only administered a single practice test, and measured how stressful participants perceived the practice test to be. By contrast, most (i.e., 22 out of 25) of the effects included in the meta-analysis

3. Studies investigating the effect of test modality (e.g., computer-based *vs*. paper-based) on TA were excluded because there was no control comparison condition against which practice tests were compared (e.g., Kolagari et al., 2018).
4. Only studies reporting sufficient data for effect size calculation were included.
5. Only studies written in English or Chinese were considered because of the authors' language proficiency.
6. Both laboratory and classroom studies were included. Below we assess whether research setting (laboratory *vs*. classroom) moderates the effect of practice tests on TA.

The screening procedure and results are depicted by a flowchart in Figure 1. In total, 24 studies were identified as eligible for the meta-analysis. From these studies, $k$ = 25 effects based on data from 3,374 participants were extracted. The characteristics of the included studies are reported in Table 1.

**Moderator coding**

Besides assessing the main effect of practice tests on TA, another goal of the current meta-analysis is to explore potential moderators of the effect, which has been largely overlooked previously. In total, five potential moderators were coded here.

***Research setting (laboratory vs. classroom)***: Most (22 out of 25) of the included effects were extracted from classroom studies (i.e., quasi-experiments), which explored the effect of class quizzes

---

were conducted in the classroom and all of them implemented multiple practice tests (e.g., administering a quiz each week across a whole semester) and measured participants' TA about upcoming tests or exams. It should be noted that one of the included effects also measured anxiety experienced during practice tests (Higham et al., 2022). However, the study by Higham et al. (2022) differed from those by Hinze and Rapp (2014) and Wenzel and Reinhard (2021) in many key aspects. For instance, Higham et al.'s study was conducted in an Introduction to Psychology course across a semester, by comparison with the brief lab experiments conducted by Hinze and Rapp (2014) and Wenzel and Reinhard (2021). Higham et al. administered three practice tests (i.e., successive tests) on each weekly lecture, by comparison with the single practice test implemented by Hinze and Rapp (2014) and Wenzel and Reinhard (2021). We further review the study by Higham et al. (2022) in the General Discussion. It should be highlighted that the results do not change if Higham et al.'s (2022) results are excluded from the meta-analyses (see below for details).

on TA (e.g., Denny et al., 1964; Piroozmanesh & Imanipour, 2018). These effects were assigned to a classroom category. The other three effects were extracted from laboratory studies (e.g., Szpunar et al., 2013; Yang et al., 2020), and were allocated to a laboratory category.

***Measurement type (standardized vs. non-standardized measures)***: Eighteen effects measured TA using standardized questionnaires or scales (e.g., Powell, 2002) and were coded into a standardized measure category. By contrast, the other seven effects measured TA via a single or multiple questions developed by the researchers themselves (e.g., Brown & Tallon, 2015; Dustin, 1971; Szpunar et al., 2013; Yang et al., 2020), and hence were coded into a non-standardized measure category.

***Practice test format (recall vs. recognition vs. mixed)***: According to the practice test format in each study, the effects were divided into three categories: $k$ = 4 for recall tests (e.g., free recall, short answer), $k$ = 7 for recognition tests (e.g., multiple-choice), and $k$ = 5 for mixed tests (i.e., mixed recall and recognition tests).[2]

***Practice test performance***: Among the 25 included effects, 18 did not provide practice test performance results and were therefore excluded from the meta-regression analysis, leaving final data from only five effects (Higham et al., 2022; Szpunar et al., 2013; Wong & Lim, 2021; Yang et al., 2020).

***Publication status***: Among the included effects, 21 were extracted from published journal articles (e.g., Szpunar et al., 2013; Yang et al., 2020), with the other four extracted from unpublished dissertations (e.g., Elskamp, 2020).

[2] For the other nine effects, the studies did not provide sufficient information to identify their test formats, and hence these were excluded from the sub-group analysis.

**Data extraction and analysis method**

The first two authors independently performed data extraction and moderator coding. All divergences were settled through discussion.

Among the 25 included effects, 12 adopted a pretest-posttest control group design in which TA in both the experimental and control groups was measured before and after the intervention phase (see the above discussion of Piroozmanesh & Imanipour, 2018, as an example). Among these 12 effects, 10 reported pretest means, posttest means and pretest *SD*s of TA in both groups. For these 10 effects, their effect sizes (i.e., Cohen's *d*s) were calculated based on the mean pre-post change in the experimental group minus the mean pre-post change in the control group, divided by the pooled pretest *SD*, as recommended by Morris (2007). The other two studies employing a pretest-posttest control group design (Denny et al., 1964; Dustin, 1971)[3] did not provide detailed pretest TA data and hence we calculated Cohen's *d*s based on the posttest TA data (e.g., mean, *SD*, *SE*, or *t*).

For another 13 effects, eight employed between-subjects designs. For instance, Szpunar et al. (2013) asked the test and restudy groups to report their TA before taking a cumulative test. The other five effects utilized within-subjects designs. For example, Cardozo et al. (2020) had half the students attend a traditional lecture without practice tests and measured their TA before the summative assessment on the lecture. Next, these students attended another lecture with practice tests, and then measured their TA before the summative assessment on that lecture. The order of the lectures with and without practice tests was reversed for the other half students. For these 13 effects involving

[3] Although Denny et al. (1964) and Dustin (1971) did not report detailed pretest TA data, both of them explicitly reported that there was minimal difference (e.g., $t < 1$) in pretest TA between the experimental and control groups.

either between- or within-subjects designs, Cohen's *d*s were calculated using other reported statistics (e.g., *t*, *SE*, *N*) according to the formulae provided by Borenstein et al. (2009).[4]

To mitigate potential bias from effects with small sample sizes, all Cohen's *d*s were then transformed into Hedges' *g*s using the formula provided by Borenstein et al. (2009). There were 25 effects extracted from 24 studies and all effects came from independent groups of participants, which means that there was no statistical dependency among the effects. Hence, all meta-analyses were performed using random-effects models via the R *metafor* package (Pastor & Lazowski, 2018; Van Den Noortgate & Onghena, 2003).

As mentioned above, the meta-analysis assessed five potential moderators of the effect of practice tests on TA. For some moderators (e.g., practice test format, practice test performance), some studies did not provide sufficient information to identify their characteristics which made it impossible to include all moderators in a multivariate meta-regression analysis. Hence, we conducted univariate meta-regression analyses to assess the moderating effect of each factor. Another reason for conducting univariate meta-regression analyses is that the number ($k$ = 25) of included effects was too small to run a multivariate meta-regression analysis to simultaneously assess the moderating effects of five factors. To foreshadow, most (four out of five) of the assessed moderators did not show a statistically detectable moderating effect, which should allay any concern about their potential confounding effects in univariate analyses.

## Results

[4] There was minimal heterogeneity among the effects involving pretest-posttest control group, between-subjects, and within-subjects designs, $Q(2) = 1.38$, $p = .50$, $I^2 = 81.1\%$, $\tau^2 = 0.36$. Hence, below we do not further discuss experimental design issues.

The Results section is structured as follows. First, we describe the total effect of practice tests on TA. Then we report results from meta-regression analyses to assess potential moderators of the effect. Finally, we report results from three analyses to determine whether the included studies suffered from publication bias.

**Effect of practice tests on TA**

A random-effects meta-analysis found that the total effect of practice tests on TA is Hedges' $g$ = -0.52, 95% CI = [-0.69, -0.36], $p < .001$, and the median effect size is $g$ = -0.53.[5] Most (88.0%; 22 out of 25) of the included studies found evidence that practice tests reduced TA, with only a minority (12.0%) showing the converse pattern. The proportion showing that practice tests reduced TA was substantially greater than the proportion showing the converse pattern, $\chi^2(1) = 25.92$, $p < .001$.

Although these results indicate that practice tests reduced TA to a medium extent, the heterogeneity among the included effects is substantial, $Q(24) = 146.69$, $p < .001$, $I^2$ = 83.6%, $\tau^2$ = 0.13, indicating the necessity to explore potential moderators.

**Potential moderators**

Table 2 summarizes the results of categorical moderator analyses.

***Research setting***: A random-effects meta-regression analysis found that research setting (laboratory *vs*. classroom) did not significantly moderate the effect of practice tests on TA, $Q(1)$ = 0.87, $p = .35$, $I^2$ = 80.3%, $\tau^2$ = 0.12. Practice tests significantly reduced TA in the classroom, $g$ = -0.55, 95% CI = [-0.72, -0.38], $p < .001$, whereas the results from laboratory research were statistically inconclusive, $g$ = -0.31, 95% CI = [-0.78, 0.15], $p = .18$.

[5] Only one of the included effects assessed anxiety experienced during practice tests (Higham et al., 2022). After excluding this effect, the results again showed a robust negative effect on TA, $g$ = -0.54, 95% CI = [-0.71, -0.37], $p < .001$.

***Measurement type***: There was no statistically detectable heterogeneity between effects measured via standardized and non-standardized measures, $Q(1) = 0.02$, $p = .90$, $I^2 = 82.6\%$, $\tau^2 = 0.14$. Importantly, there was a reliable negative effect of practice tests on TA for both standardized, $g = -0.53$, 95% CI = [-0.73, -0.33], $p < .001$, and non-standardized measures, $g = -0.50$, 95% CI = [-0.82, -0.19], $p = .002$. These results suggest that the effect of practice tests on TA tends not to be related to measurement type (i.e., measurement reliability).

***Practice test format***: Although the results showed no significant moderating effect of practice test format, $Q(2) = 2.26$, $p = .32$, $I^2 = 80.4\%$, $\tau^2 = 0.10$, only recognition, $g = -0.59$, 95% CI = [-0.87, -0.31], $p < .001$, and mixed, $g = -0.35$, 95% CI = [-0.68, -0.02], $p = .04$, tests significantly reduced TA, whereas the results for recall tests were not statistically significant, $g = -0.27$, 95% CI = [-0.62, 0.08], $p = .13$. Further analyses showed no significant difference in the effect sizes between recall and recognition tests, $Q(1) = 1.15$, $p = .28$. In brief, only recognition and mixed tests showed significant negative effects on TA, and recognition tests were numerically (but not significantly) more effective in mitigating TA than recall tests.

***Practice test performance***: As shown in Figure 2, there was a negative relationship between practice test performance and the effect sizes, $b = -1.68$, 95% CI = [-3.06, -0.31], $p = .02$, indicating that every increase of 10% in practice test performance reduces TA by $g = -0.17$.[6] Thus, easy practice tests are more effective in reducing TA than difficult ones. However, we warn readers to interpret these results cautiously because the number of included effects in this meta-regression analysis was small.

[6] After excluding the effect of Higham et al. (2022) from this meta-regression analysis, the negative relationship persisted, $b = -2.22$, 95% CI = [-3.35, -1.09], $p < .001$.

**Publication bias detection**

Three methods were implemented to assess if the included effects were contaminated by publication bias. The first method is to test the asymmetry of the funnel plot (Egger et al., 1997). As shown in Figure 3, the funnel plot was not significantly asymmetric, $Z = -1.01$, $p = .31$, indicating minimal publication bias. The second method is to assess whether publication status moderates the observed effect sizes. There was no significant difference between published, $g = -0.50$, 95% CI = [-0.68, -0.32], $p < .001$, and unpublished, $g = -0.63$, 95% CI = [-1.06, -0.20], $p = .004$, effects, $Q(1) = 0.29$, $p = .59$, $I^2 = 83.7\%$, $\tau^2 = 0.13$, further mitigating worry about publication bias.

Finally, we assessed publication bias using Robust Bayesian Meta-Analysis (RoBMA; Bartoš et al., 2022; Maier et al., 2022). RoBMA utilizes Bayesian model-averaging to combine estimates from 36 models — including PET-PEESE (Stanley & Doucouliagos, 2014) and selection models (Vevea & Hedges, 1995) — both with and without publication bias. Each model is fit to the data, and then the estimated effect is computed by weighting each of the estimated effects according to its likelihood. RoBMA also computes Bayes factors to quantify the evidence for the existence or absence of an effect, heterogeneity, and publication bias. Recent research demonstrated that RoBMA outperforms other methods in detecting publication bias (Bartoš et al., 2022; Maier et al., 2022).

RoBMA was conducted via JASP (Version 0.16.2), with all parameters set as default values. The results revealed extremely strong evidence for the effect of practice tests on TA, $BF_{10} = 25,856$, together with extremely strong evidence for residual heterogeneity, $BF_{10} > 10,000$. Although the Bayesian evidence for publication bias in the included studies was anecdotal, the evidence is about $BF_{01} = 1.74$ times more consistent with there being no publication bias than there being publication bias. The mean estimated effect (i.e., the corrected effect) was $g$ = -0.50, 95% CI = [-0.69, -0.32], with a median effect size of $g$ = -0.50. Thus, the corrected effect ($g$ = -0.50) was similar to the uncorrected one ($g$ = -0.52), again mitigating worry about publication bias.

In summary, three methods were utilized to determine if the included effects suffered from publication bias. All of them converged on the inference that there was little risk of publication bias among the included effects, and the documented effect of practice tests on TA is reliable.

## General Discussion

Although practice testing has been repeatedly established as an efficient tool to facilitate learning, students and instructors frequently hold a legitimate concern that practice tests or quizzes may provoke TA, which might be one reason why practice testing has not been widely applied to aid learning in educational settings. However, there are many reasons to believe that practice tests may reduce TA (see below for detailed discussion). The current review integrated results from 24 studies to determine the effect of practice tests on TA. The results showed extremely strong Bayesian evidence ($BF_{10}$ > 25,000) that practice tests decrease TA to a medium extent ($g$ = -0.52). Furthermore, results from three publication bias detection analyses jointly indicate that the included studies suffered minimally from publication bias.

In the classroom, almost every curriculum requires a summative evaluation, and preparation for such evaluations is generally an integral part of the education process. Given that summative assessments are almost inevitable, the question of how to reduce TA about these assessments is of major interest to researchers, teachers, students, and other stake-holders. The principal implication from the current review is that practice tests (quizzes) can be utilized as a practical approach to alleviate TA in students faced with challenging exams. Besides the benefits in reducing TA, numerous studies and meta-analyses have confirmed that practice testing also consolidates long-term retention of studied information, facilitates subsequent learning of new information, and fosters knowledge transfer (Carpenter, 2012; Pastötter & Bäuml, 2014; Roediger & Karpicke, 2006a; Shanks et al., 2023; Yang et al., 2021; Yang, Potts, et al., 2018). These beneficial effects of practice tests on reducing TA and enhancing learning jointly make a strong recommendation for instructors to incorporate quizzes into their curriculum.[7]

Even though dozens of studies have explored the effect of practice tests on TA, most of them were primarily conducted to assess the effect of tests on learning, with the influence on TA as a subsidiary research topic (e.g., Brown & Tallon, 2015; Dustin, 1971; Fulkerson & Martin, 1981; Molin et al., 2021; Wong & Lim, 2021). Hence, unsurprisingly, previous studies did not investigate the mechanism(s) through which practice tests reduce TA and most did not attempt to provide any explanation of this effect (e.g., Denny et al., 1964; Dustin, 1971; Haleem & Saeed, 2021), but with a few exceptions (e.g., Agarwal et al., 2014; Cardozo et al., 2020).

Here, we propose two theoretical explanations of this effect: (1) enhanced learning and (2) reduced uncertainty. The enhanced learning hypothesis asserts that practice tests decrease TA through boosting attainment. Specifically, tests motivate students to study harder and prepare better for future assessments (Brown & Tallon, 2015), which in turn mitigate worry about poor performance and decrease TA (Dustin, 1971; Kilickaya, 2017). Consistent with this explanation,

---

[7] Instructors may have other negative evaluations about class quizzes. For instance, they may be concerned that quizzes squeeze time available for other teaching activities or may increase the attainment gap between students with good and poor learning abilities. Readers can consult recent reviews by Yang et al. (2021, 2023) which provide responses to several concerns that instructors may have about class quizzes.

Brown and Tallon (2015) observed that, by comparison with students in a no pre-test class, those in a pre-test class reported that they complied better with assigned readings before the lecture and felt more prepared for and were less anxious about course exams. Along the same lines, Yang et al. (2017a) demonstrated that practice tests sustained participants' study effort across a prolonged learning session; Wong and Lim (2021) showed that inserting interim tests into a learning session reduced mind wandering; Depaolo and Wilkinson (2014) observed that regular quizzes increased class attendance; and Connor-Greene (2000) found that class quizzes encouraged students to study and review course materials regularly rather than depending on last-minute cramming for exam preparation.

It has also been observed that quizzes help students to identify what they have already mastered and what they have not yet understood, which then allows them to effectively regulate study resource allocation to further encode unmastered concepts (Prashanti & Ramnarayan, 2019). Additionally, quizzes help teachers to monitor students' progress, identify their difficulties, and modify teaching activities to best meet students' learning needs (Cardozo et al., 2020). These findings jointly point to the inference that practice tests can reduce TA through enhancing attainment (for related discussion, see Agarwal et al., 2014; Brown & Tallon, 2015; McKenzie & Henry, 1979; Yang et al., 2020).

The reduced uncertainty hypothesis, in contrast, proposes that practice tests inform students about what content will be later assessed (i.e., test content), in which format it will be evaluated (i.e., test format), and how difficult the future exam will be (i.e., test difficulty). These informative functions of practice tests may reduce students' uncertainty about upcoming exams and alleviate their TA (Cardozo et al., 2020; Downing et al., 2020). Accordingly, it is reasonable to predict that good practice performance should enhance students' confidence of performing well on future exams and reduce their anxiety about exam performance (Bangert-Drowns et al., 1991; Hinze & Rapp, 2014; von der Embse et al., 2018). Hence, easy practice tests (generally associated with good practice performance) should be more effective in reducing TA than difficult ones. Consistent with this prediction, the present meta-analysis did observe a significant relationship between practice test difficulty (indexed by practice test performance) and the effect size of practice tests on TA. Furthermore, the meta-analysis demonstrated that easy recognition tests produced a significant alleviating effect on TA whereas the effect of difficult recall tests on TA was not statistically significant (though in the same direction). These findings are also consistent with the reduced uncertainty theory's prediction that easy practice tests are better able to mitigate TA.

A tempting inference is that instructors should frequently implement easy tests and avoid difficult ones. However, we again warn readers to interpret these findings as suggestive rather than conclusive because (1) there were only a few effects available to assess the relationship between practice test difficulty and the effect size of practice tests on TA, and (2) the difference between the effects of recognition and recall tests was not statistically significant. Further tests on this hypothesis are called for. Moreover, it should be highlighted that the retrieval effort theory of test-enhanced learning asserts that practice tests consolidate long-term memory through inducing desirable difficulty (Bjork et al., 2014; Rowland, 2014) and predicts that difficult tests should be more effective in facilitating learning than easy ones. Indeed, Rowland's (2014) meta-analysis demonstrated that the enhancing effects of cued recall ($g$ = 0.72) and free recall tests ($g$ = 0.81) are much larger than that of recognition tests ($g$ = 0.36) (but also see Adesope et al., 2017; Yang et al., 2021).

It seems there is a tradeoff between the alleviating effect of practice tests on TA and their enhancing effect on learning: easy practice tests tend to be more effective in reducing TA but less effective in enhancing learning. One technique, successive relearning (for a review, see Rawson & Dunlosky, 2022), may maximize both kinds of benefits. Successive relearning involves repeatedly taking practice tests on study materials with corrective feedback until successful retrieval is achieved repeatedly over long retention intervals. Through utilizing successive relearning, instructors can still implement difficult quizzes and repeat challenging questions with feedback over spaced intervals. Across successive relearning, students will not only gain robust learning but also end up with very little anxiety (Higham et al., 2022).

Before administering class quizzes, instructors need to realize that practice tests (quizzes) themselves may be more stressful than other learning activities (e.g., restudying), even though they reduce students' TA about summative exams. Indeed, two laboratory studies have explored whether learners perceive practice testing as more stressful than restudying (Hinze & Rapp, 2014; Wenzel & Reinhard, 2021). Wenzel and Reinhard (2021) asked two (test *vs*. restudy) groups of participants to read mathematical/statistical information. After initial study, the review phase began, in which the test group took a test on the studied information while the restudy group reread the information. Next, both groups reported how stressful the review phase was. The results showed that the test group experienced numerically (but not significantly) greater stress during the review phase than the restudy group. Consistent findings were observed in the study by Hinze and Rapp (2014) in which two (test *vs*. restudy) groups of participants read four biology texts. Then, the test group completed a free recall test on each text while the restudy group reread the texts. Next, both groups rated the level of pressure they experienced during the rereading/testing phase. The results showed that even though participants were pre-informed that their task performance had no impact on their task bonus, the test group still reported a significantly higher level of pressure than the restudy group.

We conducted an exploratory mini meta-analysis to integrate the results of Hinze and Rapp (2014) and Wenzel and Reinhard (2021). The results showed that practice testing was more stressful than restudying, $g$ = 0.60, 95% CI = [0.15, 1.05], $p$ = .009. This finding implies that when instructors administer quizzes, they also need to appreciate that quizzing is a stressful learning activity. In addition, they should seek to find techniques to make quizzes less stressful and more enjoyable. Previous studies have provided several methods to achieve this goal. For instance, Khanna (2015) demonstrated that students perceived ungraded quizzes as less stressful than graded ones, and reported that they enjoyed taking ungraded quizzes in the course.

Davis and Winston (2022) found that 89% of students reported positive experiences when taking quizzes allowing multiple answer attempts. Similarly, Sullivan (2017) found that over 90% of students agreed with the statement “*the option to retake a quiz makes me feel secure and reduces anxiety*”. Consistent findings were also documented by Higham et al. (2022). In this study, after each weekly lecture, students were asked to either restudy class knowledge statements three times or take three practice tests on these statements, with the same questions presented in each test. After each review session, students reported their levels of anxiety during the review phase. Consistent with Hinze and Rapp (2014) and Wenzel and Reinhard (2021), Higham et al. (2022) found that students reported higher levels of anxiety in the test than in the restudy condition during the first review session, but reported about the same level of anxiety during the second review session. Critically, during the third review session, students were much less anxious in the test than in the restudy condition. Such findings are consistent with the conjecture that allowing students to take quizzes with multiple attempts can effectively reduce TA. Higham et al. (2022) also found that practice test performance substantially increased across the first ($M$ = 24%), second ($M$ = 68%), and third ($M$ = 82%) tests, coinciding with the decreasing effect on TA. These findings imply that repeated tests gradually become easier to perform and that as test performance steadily increases across test cycles, anxiety experienced during the practice tests gradually reduces.

Besides the aforementioned approaches, another method for mitigating quiz stress is gamification. Har (2022) demonstrated that students enjoyed gamified quizzes for their fun, engaging, and entertaining features.[8] In summary, instructors are recommended to administer ungraded and gamified quizzes, and also consider allowing students to retake quizzes multiple times, in order to alleviate test stress.

**Limitations and future research directions**

It should be acknowledged that the current meta-analysis suffers from several limitations, and there are many important questions awaiting future research. For instance, even though we searched 28 electronic databases to ensure the comprehensiveness of the meta-analysis, it is still uncertain whether all relevant studies have been identified and included. Another limitation is that the number ($k$ = 25) of included effects was small, yielding low statistical power for moderator analyses. Indeed,

[8] Several platforms or techniques are available to deliver gamified quizzes, such as Kahoot, Quizlet, Edupuzzle, Quizizz, and clicker response system.

among the five moderating factors assessed here, only practice test performance significantly moderated the effect of practice tests on TA, with the moderating effects of the other four factors not statistically significant. As discussed above, it is premature to draw any firm conclusions based on the non-significant findings observed here. We hence reiterate our warning to readers not to weigh the moderator results too heavily. To our knowledge, none of the previous studies has directly explored any of these potential moderators. Future research is encouraged to do so as a way to sharpen our understanding about the mechanisms underlying the effect and also provide valuable educational guidance about how to administer practice tests in the most effective way to reduce TA.

A potentially important moderator not discussed so far is the provision of corrective feedback in the practice tests. Many laboratory studies of the testing effect withhold feedback, for the simple reason that researchers are often interested in investigating the effects of pure retrieval, unconfounded by new learning that can occur when feedback is provided. In the classroom, in contrast, it would be perverse to withhold feedback in a practice test, the very purpose of which is to help students learn, and indeed feedback does tend to enhance the efficacy of practice tests (see Shanks et al., 2023; Yang et al., 2021). We coded feedback provision in the studies included in the meta-analysis and found no significant difference between the impact on TA of practice tests with and without feedback. However, this comparison has low statistical power because many of the studies did not clearly report whether they provided feedback or not and therefore had to be excluded.

Another important but unexplored question is about individual differences in the effect of practice tests on TA. As noted by Agarwal et al. (2014), although class quizzes reduced TA for a majority (72%) of students, there remained 28% of students in their study who reported that quizzes did not influence TA or even increased it. Undoubtedly, practice tests do not produce an equivalent effect on TA for all students and there are very likely to be individual differences in this effect. Future research should seek to clarify what kind of individual differences (e.g., self-efficacy, beliefs about failure) predict the sign and magnitude of the effect, and determine for whom practice tests appreciably lessen or increase TA.

Overall, research on the effect of practice tests on TA is still in its infancy, and many critical questions concerning this effect (e.g., underlying mechanisms, moderating factors, boundary conditions, and individual differences) are underexplored or even unexplored. Future research on these questions is encouraged.

## Concluding Remarks

Practice tests appreciably reduce TA to a medium extent, but practitioners should be aware that practice tests themselves may be more stressful than other study activities (e.g., restudying). It remains largely unknown through which mechanisms practice tests alleviate TA, which factor(s) moderate this effect, and what kinds of individual differences predict its sign and magnitude. All these questions await future research. Nevertheless the available evidence, as summarized in the present meta-analysis, is sufficient to support the conclusions that (1) there is no reason to believe that the robust benefits of practice tests on learning are offset by a tendency to increase TA, and (2) that indeed practice tests have the capacity to alleviate TA.

## References

*References marked with an asterisk indicate studies included in the meta-analysis.*

Act, E. S. S. (2015). Every student succeeds act. *Public Law*, 114-195.

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*, 659-701. doi:10.3102/0034654316689306

Agarwal, P. K., D'Antonio, L., Roediger, H. L., 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 and Cognition, 3*, 131-139. doi:https://doi.org/10.1016/j.jarmac.2014.07.002

Akinsola, E. F., & Nwajei, A. D. (2013). Test anxiety, depression and academic performance: Assessment and management using relaxation and cognitive restructuring techniques. *Psychology, 4*, 18-24. doi:10.4236/psych.2013.46A1003

Ali, M. S., & Mohsin, M. N. (2013). Relationship of test anxiety with students' achievement in science. *International Journal of Educational Science and Research, 3*, 99-106.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research, 85*, 89-99. doi:10.1080/00220671.1991.10702818

Bartoš, F., Maier, M., Quintana, D. S., & Wagenmakers, E.-J. (2022). Adjusting for publication bias in JASP and R: Selection models, PET-PEESE, and robust bayesian meta-analysis. *Advances in Methods and Practices in Psychological Science, 5*, 25152459221109259. doi:10.1177/25152459221109259

Bates, L. E. (2019). *Mastery quizzing: Assessing a novel testing technique in the classroom and the laboratory.* (Ph.D.). Colorado State University, Ann Arbor. Retrieved from https://search.proquest.com/docview/2378111550?accountid=8554

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth.

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition, 3*, 165-170. doi:10.1016/j.jarmac.2014.03.002

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

*Brown, M. J., & Tallon, J. (2015). The effects of pre-lecture quizzes on test anxiety and performance in a statistics course. *Education, 135*, 346-350. Retrieved from https://www.ingentaconnect.com/content/prin/ed/2015/00000135/00000003/art00011

Bücker, S., Nuraydin, S., Simonsmeier, B. A., Schneider, M., & Luhmann, M. (2018). Subjective well-being and academic achievement: A meta-analysis. *Journal of Research in Personality, 74*, 83-94. doi:https://doi.org/10.1016/j.jrp.2018.02.007

*Cardozo, L. T., Azevedo, M. A. R. d., Carvalho, M. S. M., Costa, R., Lima, P. O. d., & Marcondes, F. K. (2020). Effect of an active learning methodology combined with formative assessments on performance, test anxiety, and stress of university students. *Advances in Physiology Education, 44*, 744-751. doi:10.1152/advan.00075.2020

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*, 279-283. doi:10.1177/0963721412452728

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology, 1*, 496-511. doi:10.1038/s44159-022-00089-1

Cassady, J. C. (2004). The influence of cognitive test anxiety across the learning–testing cycle. *Learning and Instruction, 14*, 569-592. doi:https://doi.org/10.1016/j.learninstruc.2004.09.002

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

Connor-Greene, P. A. (2000). Assessing and promoting student learning: Blurring the line between teaching and testing. *Teaching of Psychology, 27*, 84-88. doi:10.1207/S15328023TOP2702_01

Davis, D., & Winston, C. (2022). *Decreasing student stress through multi-attempt digital engineering assessments with rotating questions.* Paper presented at the 2022 ASEE Annual Conference & Exposition.

*Denny, T., Paterson, J., & Feldhusen, J. (1964). Anxiety and achievement as functions of daily testing. *Journal of Educational Measurement, 1*, 143-147. Retrieved from http://www.jstor.org/stable/1433684

Depaolo, C., & Wilkinson, K. (2014). Recurrent online quizzes: Ubiquitous tools for promoting student presence, participation and performance. *Interdisciplinary Journal of e-Skills and Lifelong Learning, 10*, 075-091. doi:10.28945/2057

Downing, V. R., Cooper, K. M., Cala, J. M., Gin, L. E., & Brownell, S. E. (2020). Fear of negative evaluation and student anxiety in community college active-learning science courses. *CBE—Life Sciences Education, 19*, ar20. doi:10.1187/cbe.19-09-0186

*Dustin, D. S. (1971). Some effects of exam frequency. *The Psychological Record, 21*, 409-414. doi:10.1007/BF03394033

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*, 629-634. doi:10.1136/bmj.315.7109.629

*Elskamp, F. (2020). The effects of quizzing in recorded lectures on test-anxiety and delayed learning outcomes. (Master's thesis). University of Twente. Retrieved from https://essay.utwente.nl/80610/

Ferguson, R. D. (1986). Abstraction anxiety: A factor of mathematics anxiety. *Journal for Research in Mathematics Education, 17*, 145-150. doi:https://doi.org/10.5951/jresematheduc.17.2.0145

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*, 1502. doi: https://doi.org/10.1126/science.1255484

*Fulkerson, F. E., & Martin, G. (1981). Effects of exam frequency on student performance, evaluations of instructor, and test anxiety. *Teaching of Psychology, 8*, 90-93. doi:10.1207/s15328023top0802_7

*Haleem, B., & Saeed, M. (2021). Effect of frequent testing on students' test anxiety at university level. *Perennial Journal of History, 2*, 163-171. doi:10.52700/pjh.v2i2.69

Har, F. (2022). English language learning in response to the COVID-19 pandemic: Hong Kong English as a Second Language students' perceptions of Badaboom! *Frontiers in Education, 7*. doi:10.3389/feduc.2022.966059

Heiner, C. E., Banet, A. I., & Wieman, C. (2014). Preparing students for class: How to get 80% of students reading the textbook before class. *American Journal of Physics, 82*, 989-996. doi:10.1119/1.4895008

Heitmann, S., Grund, A., Berthold, K., Fries, S., & Roelle, J. (2018). Testing is more desirable when it is adaptive and still desirable when compared to note-taking. *Frontiers in Psychology, 9*, 2596-2596. doi:10.3389/fpsyg.2018.02596

*Higham, P. A., Zengel, B., Bartlett, L. K., & Hadwin, J. A. (2022). The benefits of successive relearning on multiple learning outcomes. *Journal of Educational Psychology*, *114*, 928-944. doi:10.1037/edu0000693

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology, 28*(4), 597-606. doi:10.1002/acp.3032

Jerrell, C. C., & Betty, E. G. (2005). The effects of online formative and summative assessment on test anxiety and performance. *The Journal of Technology, Learning and Assessment, 4*. Retrieved from https://ejournals.bc.edu/index.php/jtla/article/view/1648

Kang, S. H. K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition, 3*, 183-188. doi:https://doi.org/10.1016/j.jarmac.2014.05.006

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

Khanna, M. M. (2015). Ungraded pop quizzes: Test-enhanced learning without all the anxiety. *Teaching of Psychology, 42*, 174-178. doi:10.1177/0098628315573144

Kilickaya, F. (2017). The effects of pre-lecture online quizzes on language students' perceived preparation and academic performance. *PASAA: Journal of Language Teaching and Learning, 53*, 59-84. Retrieved from http://www.culi.chula.ac.th/publicationsonline/home_p1.php

Kolagari, S., Modanloo, M., Rahmati, R., Sabzi, Z., & Ataee, A. J. (2018). The effect of computer-based tests on nursing students' test anxiety: A quasi-experimental study. *Acta Informatica Medica, 26*, 115-118. doi:10.5455/aim.2018.26.115-118

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47*, 674-682. doi:10.1111/medu.12141

Leeming, F., C. (2002). The Exam-A-Day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210-212. doi:10.1207/S15328023TOP2903_06

Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2022). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, Advance online publication. doi:10.1037/met0000405

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *The Journal of Abnormal and Social Psychology, 47*, 166-173. doi:10.1037/h0062855

*Mayers, M. F. (2004). *Relationships among test anxiety, state anxiety, and academic performance as a function of academic assessment frequency.* (Ph.D.). University of Connecticut, Ann Arbor. Retrieved from https://www.proquest.com/dissertations-theses/relationships-among-test-anxiety-state-academic/docview/305211848/se-2?accountid=8554

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

McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology, 21*, 89-101. doi:10.1080/01443410020019867

*McKenzie, G. R., & Henry, M. (1979). Effects of testlike events on on-task behavior, test anxiety, and achievement in a classroom rule-learning task. *Journal of Educational Psychology, 71*, 370-374. doi:10.1037/0022-0663.71.3.370

*Molin, F., Cabus, S., Haelermans, C., & Groot, W. (2021). Toward reducing anxiety and increasing performance in physics education: Evidence from a randomized experiment. *Research in Science Education, 51*, 233-249. doi:10.1007/s11165-019-9845-9

*Moradi, S., Maraghi, E., Babaahmadi, A., & Younespour, S. (2021). Application of pop quiz method in teaching biostatistics to postgraduate midwifery students and its effect on their statistics anxiety, test anxiety and academic achievement: A quasi experimental study with control group. *Journal of Biostatistics and Epidemiology*, *7*, 181-188. doi:10.18502/jbe.v7i2.6736

Morris, S. B. (2007). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364-386. doi:10.1177/1094428106291059

*Naseem, A. (2021). Effect of quizzes on anxiety and performance in mathematics at middle level. *Bulletin of Education and Research, 43*, 59-75. Retrieved from https://eric.ed.gov/?id=EJ1320239

O'Neil Jr, H. F., Spielberger, C. D., & Hansen, D. N. (1969). Effects of state anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology, 60*, 343-350. doi:10.1037/h0028323

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*, 710-756. doi:10.1037/bul0000151

Pan, S. C., Sana, F., Schmitt, A. G., & Bjork, E. L. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition, 9*, 542-554. doi:https://doi.org/10.1016/j.jarmac.2020.07.004

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

*Piroozmanesh, A., & Imanipour, M. (2018). The effect of formative assessment on test anxiety of nursing students. *Journal of Medical Education Development*, *10*, 18-26. doi:10.29252/edcj.10.28.18

*Powell, N. L. (2002). *A study of writing to learn, frequent quizzes, and learning styles in a developmental algebra course.* (Ph.D.). Auburn University, Ann Arbor. Retrieved from https://www.proquest.com/dissertations-theses/study-writing-learn-frequent-quizzes-learning/docview/304799893/se-2?accountid=8554

Prashanti, E., & Ramnarayan, K. (2019). Ten maxims of formative assessment. *Advances in Physiology Education, 43*(2), 99-102. doi:10.1152/advan.00173.2018

Rawson, K. A., & Dunlosky, J. (2022). Successive relearning: An underexplored but potent technique for obtaining and maintaining knowledge. *Current Directions in Psychological Science, 31*, 362-368. doi:10.1177/09637214221100484

Rhodes, M. G., & Castel, A. D.. Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*, 615-625. doi:10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*, 550-554. doi:10.3758/PBR.16.3.550

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*, 20-27. doi:10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 17*, 249-255. doi:https://doi.org/10.1111/j.1745-6916.2006.00012.x

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

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

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. doi:10.1037/xap0000134

*Sanaeifar, S. H., & Nafari, F. N. (2018). The effects of formative and dynamic assessments of reading comprehensions on intermediate EFL learners' test anxiety. *Theory and Practice in Language Studies, 8*, 533-540. doi:10.17507/tpls.0805.12

Sax, G., & Collet, L. S. (1968). An empirical comparison of the effects of recall and multiple-choice tests on student achievement. *Journal of Educational Measurement, 5*, 169-173. Retrieved from http://www.jstor.org/stable/1434415

Schrank, Z. (2016). An assessment of student perceptions and responses to frequent low-stakes testing in introductory sociology classes. *Teaching Sociology, 44*, 118-127. doi:10.1177/0092055X15624745

*Seeley, E. L., Andrade, M., & Miller, R. M. (2018). Exam anxiety: Using paired adaptive tests to reduce stress in business classes. e-*Journal of Business Education and Scholarship of Teaching*, *12*, 1-13.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*, 27-41. doi:10.1080/08917779108248762

Shanks, D. R., Don, H. J., Boustani, S., & Yang, C. (2023). Test-enhanced learning. In *Oxford Research Encyclopedia of Psychology*: Oxford University Press.

Spada, M. M., & Moneta, G. B. (2014). Metacognitive and motivational predictors of surface approach to studying and academic examination performance. *Educational Psychology, 34*, 512-523. doi:10.1080/01443410.2013.814196

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

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *5*, 60-78. doi:10.1002/jrsm.1095

Steele, D. (2011). Does test-taking help students learn? *New York Times (Letter to the editor)*, p. A24.

Steinmayr, R., Crede, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.01994

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*, 106-123. doi: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. doi:10.1073/pnas.1221764110

Tahoon, R. J. P.-E. R. R. (2021). Effects of test anxiety, distance education on general anxiety and life satisfaction of University students. *Psycho-Educational Research Reviews*. *10*,107-17. doi:https://journals.lapub.co.uk/index.php/perr/article/view/1601

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*, 419-435. https://doi.org/10.1007/BF02294384

*Ugodulunwa, C., & Okolo, P. U. (2015). *Effects of formative assessment on mathematics test anxiety and performance of senior secondary school students in Jos, Nigeria. Journal of Research and Method in Education*, *5*, 38-47. doi: 10.9790/7388-05223847

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483-493. doi:https://doi.org/10.1016/j.jad.2017.11.048

Wenzel, K., & Reinhard, M.-A. (2021). Does the end justify the means? Learning tests lead to more negative evaluations and to more stress experiences. *Learning and Motivation, 73*, 101706. doi:10.1016/j.lmot.2020.101706

*Wilson, M. (2018). The Impact of formative computer-based testing on learners' anxiety and performance on the american registry for diagnostic medical sonography examination. *Journal of Diagnostic Medical Sonography, 34*, 425-433. doi:10.1177/8756479318799338

*Wong, S. S. H., & Lim, S. W. H. (2021). A mind-wandering account of the testing effect: Does context variation matter? *Psychonomic Bulletin & Review. 29,* 220-229. doi:10.3758/s13423-021-01989-8

*Yamin, S. B. (1988). *Frequency of testing and its effects on achievement, test anxiety and attitudes toward science of students at University Technology of Malaysia.* (Ph.D.). Oregon State University, Retrieved from http://hdl.handle.net/1957/28982

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. doi:https://doi.org/10.1016/j.jml.2017.11.005

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

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

Yang, C., Potts, R., & Shanks, D. R. (2017b). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1073-1092. doi:10.1037/xlm0000363

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. doi:10.1038/s41539-018-0024-y

Yang, C., Shanks, D. R., Zhao, W., Fan, T., & Luo, L. (2023). Frequent quizzing accelerates classroom learning. In C. E. Overson, M. C. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In their own words: What scholars and teachers want you to know about why and how to apply the science of learning in your academic setting* (pp. 190-199). Society for the Teaching of Psychology.

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

Yusefzadeh, H., Iranagh, J. A., & Nabilou, B. (2019). The effect of study preparation on test anxiety and performance: A quasi-experimental study. *Advances in Medical Education and Practice, 10*, 245-251. doi:10.2147/amep.S192053

Zeidner, M. (1998). Test anxiety: The state of the art. Plenum Press, New York.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in Education* (pp. 165-184). Burlington: Academic Press.

Zernike, K. (2015). Obama administration calls for limits on testing in schools. *The New York Times*. Retrieved from https://www.nytimes.com/2015/10/25/us/obama-administration-calls-for-limits-on-testing-in-schools.html

***Table 1.*** *Characteristics of studies included in the meta-analysis*

| Study, Experiment | Research setting | Measurement type | Practice test format | Publication status | Hedges' *g* |
|---|---|---|---|---|---|
| Denny et al. (1964), Exp 1 | Classroom | Standardized | Recognition | Published | 0.23 |
| Dustin (1971), Exp 1 | Classroom | Non-standardized | Recognition | Published | -0.98 |
| McKenzie and Henry (1979), Exp 1 | Classroom | Non-standardized | Recognition | Published | -0.46 |
| Fulkerson and Glen (1981), Exp 1 | Classroom | Standardized | NA | Published | -0.39 |
| Yamin (1988), Exp 1 | Classroom | Standardized | Recognition | Unpublished | -0.94 |
| Murphy et al. (1994), Exp 1 | Classroom | Non-standardized | Mixed | Published | -0.53 |

| | | | | | |
|---|---|---|---|---|---|
| Powell (2002), Exp 1 | Classroom | Standardized | NA | Unpublished | -0.60 |
| Mayers (2003), Exp 1 | Classroom | Standardized | Recognition | Unpublished | -0.80 |
| Szpunar et al. (2013), Exp 2 | Laboratory | Non-standardized | Recall | Published | -1.30 |
| Ugodulunwa et al. (2015), Exp 1 | Classroom | Standardized | NA | Published | -1.67 |
| Brown et al. (2015), Exp 1 | Classroom | Non-standardized | Recognition | Published | -0.53 |
| Piroozmanesh (2018), Exp 1 | Classroom | Standardized | NA | Published | -0.84 |
| Wilson et al. (2018), Exp 1 | Classroom | Standardized | NA | Published | -0.81 |

| | | | | | |
|---|---|---|---|---|---|
| Sanaeifar et al. (2018), Exp 1 (Formative) | Classroom | Standardized | NA | Published | -0.19 |
| Sanaeifar et al. (2018), Exp 1 (Dynamic) | Classroom | Standardized | NA | Published | -0.58 |
| Seeley et al. (2018), Exp 1 | Classroom | Standardized | Mixed | Published | -0.38 |
| Cardozo et al. (2020), Exp 1 | Classroom | Standardized | Mixed | Published | -0.71 |
| Elskamp (2020), Exp 1 | Classroom | Standardized | Mixed | Unpublished | 0.04 |
| Yang et al. (2020), Exp 1 | Laboratory | Non-standardized | Recall | Published | -0.00 |
| Higham et al. (2022), Exp 1 | Classroom | Standardized | Recall | Published | -0.22 |

| | | | | | |
|---|---|---|---|---|---|
| Moradi et al. (2021), Exp 1 | Classroom | Standardized | NA | Published | -0.72 |
| Sarah and Stephen (2021), Exp 1 | Laboratory | Non-standardized | Recall | Published | -0.13 |
| Haleem et al. (2021), Exp 1 | Classroom | Standardized | Mixed | Published | 0.04 |
| Molin et al. (2021), Exp 1 | Classroom | Standardized | Recognition | Published | -0.37 |
| Naseem (2021), Exp 1 | Classroom | Standardized | NA | Published | -0.54 |

*Note*: NA = not available.

***Table 2.*** *Categorical moderator analysis results*

| Categorical moderators | $k$ | $g$ | 95% CI | $Q_B$ | $p$ | $I^2$ | $\tau^2$ |
|---|---|---|---|---|---|---|---|
| Research setting | | | | 0.87 | .35 | 80.3% | 0.12 |
| Classroom | 22 | -0.55 | [-0.77, -0.38] | | < .001 | | |
| Laboratory | 3 | -0.31 | [-0.78, 0.15] | | .18 | | |
| Measurement type | | | | 0.02 | .90 | 82.6% | 0.14 |
| Non-standard | 7 | -0.50 | [-0.82, -0.19] | | .002 | | |
| Standard | 18 | -0.53 | [-0.73, -0.33] | | < .001 | | |
| Practice test format | | | | 2.26 | .32 | 80.4% | 0.10 |
| Mixed | 5 | -0.35 | [-0.68, -0.02] | | .04 | | |
| Recall | 4 | -0.27 | [-0.62, 0.08] | | .13 | | |
| Recognition | 7 | -0.59 | [-0.87, -0.31] | | < .001 | | |
| Publication status | | | | 0.29 | .59 | 83.7% | 0.13 |
| Published | 21 | -0.50 | [-0.68, -0.32] | | < .001 | | |
| Unpublished | 4 | -0.63 | [-1.06, -0.20] | | .004 | | |

***Figure 1***. *Screening results of the meta-analysis*

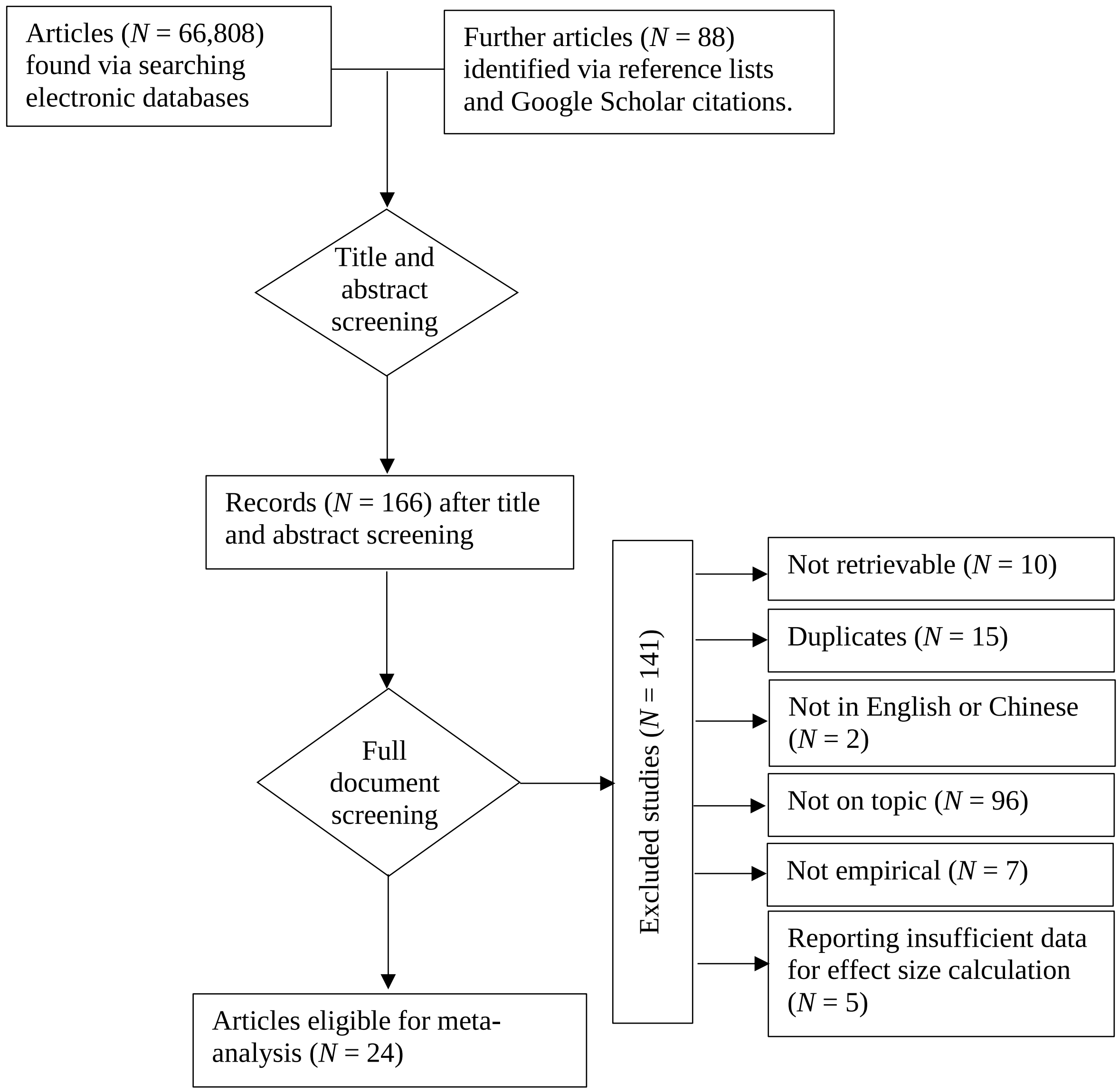
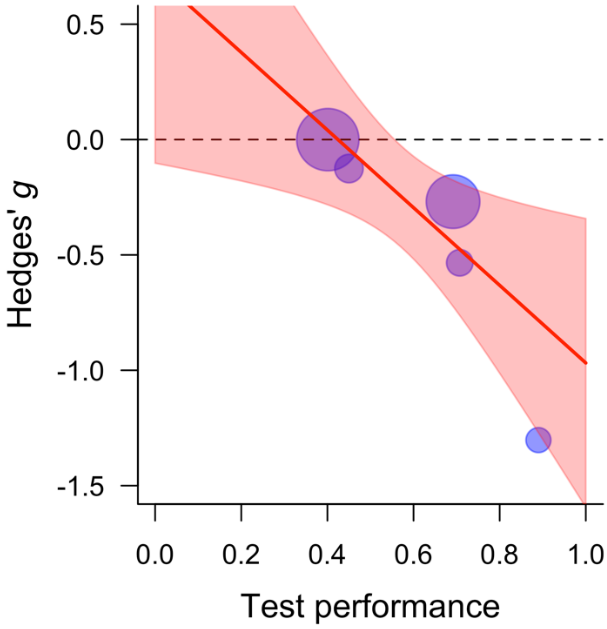


***Figure 2.*** *Relationship between practice test performance and the effect sizes of practice tests on TA*




*Note*: Bubble size represents the relative weight of the effects. The pink colored region represents the 95% CI.

***Figure 3.*** *Relationship between effect size and standard error*