Educational Psychology Review (2025) 37:115
https://doi.org/10.1007/s10648-025-10075-7

META-ANALYSIS

The Effect of Prequestions on Learning: A Multilevel Meta-Analysis

Quentin W. King-Shepard[1,2] · Julia Walker[1,2] · Timothy J. Nokes-Malach[1,2] · Shana K. Carpenter[3] · Scott H. Fraundorf[1,2]

Received: 24 February 2025 / Accepted: 11 September 2025

Abstract

For over sixty years, researchers have explored how prequestions—questions answered before an educational activity—affect learning. Although much prior work has shown an advantage of prequestions on the prequestioned content, the effects of prequestions on non-prequestioned content have resulted in a mix of positive, negative, and null effects. To better understand these findings, we conducted a multilevel meta-analysis on two questions that have been central to the prequestion literature: 1) do prequestions promote the learning of prequestioned content and does this benefit extend to non-prequestioned content? and 2) under what specific conditions are prequestions more or less effective? We found evidence that prequestions facilitated the learning of information specific to the initial prequestions asked ($g = .66$). Additionally, we found no evidence of a general learning benefit of prequestions for other, non-prequestioned, information present within the educational activity ($g = .01$). The specific learning benefits were robust and found across variations in the learning event, participant sample, and assessment conditions. Further, we found that the experimental condition moderated the strength of the effect such that conditions that provided feedback in addition to the prequestions led to better specific learning than receiving prequestions alone. We discuss our findings in relation to the hypothesized learning mechanisms, implications for education, and future research.

Keywords Prequestions · Meta-analysis · Mechanisms · Learning · Education

✉ Quentin W. King-Shepard
quk1@pitt.edu

1 Department of Psychology, University of Pittsburgh, Pittsburgh, USA

2 Learning Research & Development Center, University of Pittsburgh, 3420 Forbes Avenue, Pittsburgh, PA 15260, USA

3 School of Psychological Science, Oregon State University, Corvallis, USA

Published online: 10 December 2025

The Effect of Prequestions on Learning: A Multilevel Meta-Analysis

A frequently-employed pedagogical method is administering a pretest to students at the beginning of a course or unit before they begin learning the course material. This can serve multiple purposes. It can provide an initial assessment of students' existing knowledge, which can then be used to tailor subsequent instruction to address areas of poor performance, and it can provide an opportunity to measure learning by comparing pre- to post-test results (Simkins & Allen, 2000; Szafran, 1986). While *prequestions* – questions asked before engaging with subsequent learning resources – have traditionally been viewed as a means of assessing prior knowledge, research in psychological science has explored whether answering prequestions can enhance students' encoding and retention of new information.

Research over the past sixty years has examined the effects of prequestions across a wide range of content areas, contexts, and question types (for a comprehensive narrative review of prequestions and other pre-instruction testing methods, see Pan & Carpenter, 2023). The results of these studies have included many positive (Peeck, 1970; Haimowitz, 1972; Bull & Dizney, 1973; Manchester, 1984; Pressley et al., 1990; Richland et al., 2009; Little & Bjork, 2011), some negative (Frase et al., 1970; Duell, 1974; Sagaria & Di Vesta, 1978), and a few null findings (Felker & Dapra, 1975; Frase, 1968; Memory, 1981; Miyagi, 1995).

Given the variation of implementation and mixed findings, we conducted a meta-analytic review of the prequestion literature to better understand the effect of prequestions on learning. Specifically, we seek to address two key questions: 1) Do prequestions promote learning specific to prequestioned content or does this learning benefit extend to non-prequestioned content as well? and 2) Under what specific conditions are prequestions more or less effective? (i.e., are there specific prequestion types, content areas, or other variables that influence how likely prequestions are to benefit learning?).

Prequestions: Distinguishing Terminology and Purpose

We define *prequestions* as questions introduced *before* engaging with a subsequent learning resource or placed prior to material that contains the answers or solutions to those prequestions. Over the years, the terminology used to refer to such questions has varied. Researchers have referred to questions given during a learning activity using the terms *adjunct questions*, *interspersed questions*, *embedded questions*, *learning goals*, and *advanced organizers*. All of these could potentially refer to prequestions, but they also sometimes refer to questions given *after* the material. Therefore, we use the term *prequestions* to distinguish our focus on questions before a learning resource from the larger category of questions that could come before, during, or after learning content (Hamaker, 1986; Rickards, 1976). For a more comprehensive review of the adjunct question literature, we guide readers to Hamaker (1986; see also Lyday, 1983, for a meta-analytic review).

The prequestion effect has sometimes been used interchangeably with the term *pretesting*, a commonly used term that often refers to testing that may occur at the beginning of study or educational instruction (e.g., diagnostic pretests, baseline measurements, skill assessments). Pretesting is typically employed with the express purpose of assessing a learner's current knowledge to assist in identifying knowledge gaps that can guide subsequent teaching (Kim & Wilson, 2010). This purpose as an assessment tool, however, diverges from the purpose of implementing prequestions to stimulate learning at the beginning of a learning activity. Therefore, while prequestions—under the broad terminology of pretesting—can be used as an assessment tool, we focus here on the distinct use and theoretical foundation of prequestions as a learning tool.

In sum, while some of the articles in our record search included studies using the terms *adjunct questions* and *pretesting*, our analysis solely included studies that fell under our operational definition of prequestions, and for consistency we use the term *prequestioning* throughout the present work.

Experimental Studies of Prequestions on Learning

In a typical prequestion experiment, participants are randomly assigned to a prequestion or control condition (see Fig. 1 for an illustration). Participants in the prequestion condition typically begin the study by answering prequestion(s)without receiving feedback.[1] The prequestion asks about some specific piece of to-be-learned information that is presented during the learning event. In contrast, those in the control condition typically move straight to the learning event. During this learning event, participants in both conditions are given the same learning resource or activity (e.g., reading a text, watching a video, listening to a lecture) to learn about the target topic. After the learning event, both conditions are then given a post-test that is based on information presented in that learning event.

The post-test typically includes two types of questions: (a) *repeated questions*, which are questions that assess the same information as the prequestions (most often instantiated with verbatim wording and format to the prequestions) and (b) *new question*s, which are questions that assess information that the prequestions did not assess but can also be found in the learning resource. These two different question types can give us insight into what is learned from prequestions. If participants given prequestions show improvement on repeated questions, they demonstrate a very specific form of learning that is connected directly to the prequestioned material. Alternatively, if participants show an improvement on new questions, this result suggests that prequestions confer a more general learning benefit for additional content from the learning resource that is not specific to the prequestions.

[1] Feedback is not typically given in experiments on prequestion design because giving feedback reduces the ability to isolate the prequestion effect. For instance, if there is a positive learning benefit observed, it would be unclear if the learning benefit occurred from the prequestions themselves or simply from being given the answer in the feedback. Further, giving feedback may encourage learners to simply remember the correct answer rather than engage with the to-be-learned content.

For instance, in a study conducted by Carpenter and Toftness (2017), participants were randomly assigned into a prequestion or control group. After the learning event had finished, both groups completed a 12-question final test. For the prequestioned group, 6 of the questions were identical to the prequestions they had been tested on prior to the learning event (repeated questions) while the other 6 questions had not been seen before but pertained to information that was presented in the learning event (new questions). Researchers observed that the participants who received prequestions performed better on both repeated and new questions compared to those in the control group, who did not receive prequestions.

Hypothesized Learning Mechanisms

Why might prequestions promote subsequent learning? Here, we briefly review four hypothesized mechanisms that have been discussed in the literature.

Attention Several researchers have hypothesized that prequestions influence a learner's attention by orienting them towards the questioned information within a subsequent learning event (Rothkopf & Billington, 1979; Lewis & Mensink, 2012; Carpenter & Toftness, 2017; Carpenter et al., 2018; Pan & Sana, 2021; Bostan & Ozcelik, 2024). However, the benefit of focused attention may be a double-edged sword, as prequestions have the potential to also direct attention away from non-prequestioned information in the learning resource (Rothkopf & Billington, 1979; Lewis & Mensink, 2012; Bostan & Ozcelik, 2024). For instance, Rothkopf and Billington (1979) found that people who received prequestions would re-read and process more carefully the material directly related to those questions but would more quickly read through the material that was not related to the prequestions.

Elaboration Elaboration is the process of connecting new information to existing knowledge through inference generation. This process of engaging elaboration facilitates deeper encoding of information by having learners connect new information to prior knowledge, which in turn leads to better comprehension and retention (Endres et al., 2017; McDaniel & Donnelly, 1996). Elaboration could also potentially explain the benefit of prequestions. To the extent that answering prequestions involves some degree of prior knowledge, prequestions may serve to connect this prior knowledge to new information in the learning event. For instance, learners may answer a prequestion, respond incorrectly, then interact with a learning resource in which the correct answer can be found, allowing the chance to reconcile that prior knowledge with the information being learned.

Curiosity Curiosity is a powerful intrinsic motivator that can drive learners to seek out new information and retain that information effectively (Loewenstein, 1994; Hidi et al., 2004; Kang et al., 2009). Research investigating the effects of curiosity on memory suggests that curiosity can stem from a gap in one's knowledge between what one knows and what one wants to know (Loewenstein, 1994). It has been hypothesized that prequestions enhance a learner's intrinsic motivation by piquing

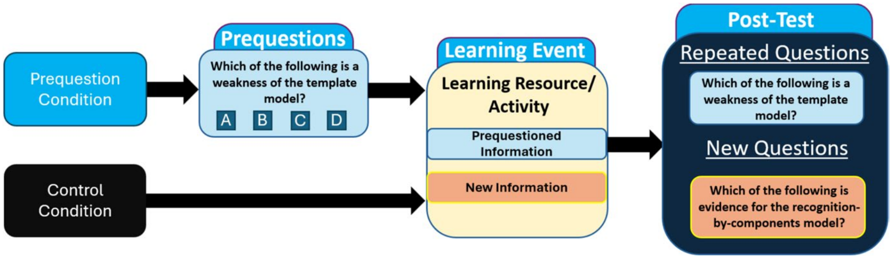

Fig. 1 A general experimental design used in prequestion research. *Note.* This figure was adapted from Carpenter et al. (2023)

their curiosity about the upcoming material. The uncertainty involved in generating an answer to a prequestion may create an interest or curiosity about that particular answer and encourage more focused or enriched encoding of the subsequent learning material. Indeed, Bull and Dizney (1973) found that students who received prequestions designed to arouse curiosity performed better at remembering the answers to those questions later, compared to students who did not receive prequestions.[2]

Metacognition We define *metacognition* as the awareness and regulation of one's own thoughts (Kuhn & Dean, 2004). Prior experimental research on metacognition has examined the impact of interventions that aim to promote metacognitive awareness and/or regulation skills, such as planning, monitoring, and evaluation (Lai, 2011). Engaging these processes as a learner, or being given an intervention designed to promote their use, can enhance learning and transfer (Dunlosky & Metcalfe, 2008; Thiede, 1999; Thiede et al., 2003; Zepeda et al., 2015). It has been theorized that prequestions may act as a metacognitive check for what learners do and do not know. By answering prequestions, learners may engage in metacognitive monitoring and assess their own knowledge of the to-be-learned material. In doing so, learners become more aware of knowledge gaps that can be addressed when interacting with the subsequent learning resource.

Summary of Mechanisms While we review these mechanisms to give insight into the potential cognitive and motivational processes involved in answering prequestions, the current work does not aim to isolate the effect of one mechanism or another. This is due to the likelihood that these mechanisms may not operate purely independently, but could instead be highly interactive such that multiple mechanisms could be driving the effects of prequestions simultaneously or in tandem. For instance, an increase in a learner's curiosity would likely increase their attention towards specific information (whether that information is prequestioned or not). Alternatively, when a learner becomes metacognitively aware of their knowledge gaps, they may become

[2] However, this study did not include a comparison with prequestions that were specifically designed not to arouse curiosity, which would provide a more direct test of curiosity's role in the prequestion effect.

more curious about the learning material, which could then lead to more focused attention. Because these mechanisms are not mutually exclusive, it may be difficult to disentangle the exact mechanism that contributes to a specific outcome without an experiment precisely designed for that purpose, which does not characterize most of the studies in our meta-analysis. The current work thus focuses instead on empirical questions and analyses of potential moderators that have emerged from the growing literature on prequestions. However, we return to some of these theoretical mechanisms in the discussion as possible explanations for some of our meta-analytic findings and fruitful directions for future work. Now that we have reviewed the potential learning mechanisms underlying prequestions, we next discuss potential variables that may moderate the effects of prequestions.

Potential Moderators of the Prequestion Effect

To better understand the variables contributing to the prequestion effect and to determine whether there are particular conditions under which prequestions have stronger or weaker effects, we examined potential moderators, such as type of prequestions, the type of knowledge tested, mode of information delivery, and others that we review below.

While many design variations exist within the prequestion literature, we found that these potential moderating variables generally fall into five categories of variations associated with 1) the prequestions themselves (type of prequestions, presentation of prequestions, placement of prequestions, and number of prequestions), 2) the learning event (mode of information delivery, duration of the learning event, amount of text material, nature of the control conditions, and content covered), 3) the assessment conditions (post-test format and retention interval), 4) participant sample (age of tested population and prior knowledge), and 5) the experimental setting (laboratory or classroom studies). We further review how these potential moderators have been variably assessed in prior investigations and, where appropriate, how the broader learning literature supports how these variables may moderate the effect of prequestions.

Variations of the Prequestions

Type of Prequestion Throughout the prequestion literature, researchers have constructed different types of prequestions (e.g., factual and conceptual) to test differences in performance outcomes. Different types of prequestions could have different effects on learning. For example, Hausman and Rhodes (2018) provided prequestioned groups with either factually- or conceptually-based prequestions. The answers to factual prequestions could be found directly in the text that students read whereas the answers to conceptual prequestions were not directly stated anywhere in the text and required participants to make inferences from the information in the text. Relative to a control group that received no prequestions, those who received factual prequestions performed significantly better on repeated factual questions, but not on new factual questions. In contrast, those who received conceptual prequestions did not perform better on either conceptual repeated or conceptual new questions relative to the control group.

Presentation of Prequestions Prequestion studies vary substantially in how prequestions are presented and in how participants are instructed to interact with prequestions. Typically, learners read prequestions, are instructed to answer them, are given no feedback, and then start the learning event. However, there are cases where learners hear prequestions rather than see them (Skiba, 1975; Wiseman, 1982) and cases where they see prequestions, but are not asked to generate a response (Lackman, 1970; Peck, 1970; Rickards, 1972; Benya, 1980; Shanahan, 1986; Pressley et al., 1990; Richland et al., 2009; Little, 2011, see experiment 7). Additionally, researchers have also investigated how supplemental materials and activities (e.g., note-taking, access to all prequestions during the learning event, giving feedback, bolding, italicizing, highlighting text) might help complement receiving prequestions (Lackman, 1970; Sinnott & Alderman, 1977; Kirschner & Brink, 1979; Benya, 1980; Shanahan, 1986; Park, 1993; Lenz et al., 2007; Kealy et al., 2003; Little & Bjork, 2016, see Experiments 1–2; James & Storms, 2019, see Experiment 4; Janelli, 2019; St. Hilaire et al., 2019, see Experiment 2; St. Hilaire & Carpenter, 2020, see Experiments 2–4; Pan & Sana, 2021, see Experiments 3–4).

Prequestion Placement Prequestion placement refers to the position of prequestions within a given learning event. The most typical approach is to give all prequestions to participants prior to the learning event (Benya, 1980; Peeck, 1970; Pressley et al., 1990; Richland et al., 2009; St. Hilaire & Carpenter, 2020). However, some earlier research varied the position of prequestions such that sometimes they were interspersed throughout the learning event (Dickerson, 1987; Feil, 1977; Hausman & Rhodes, 2018; Kirschner & Brink, 1979; Park, 1993; Sana et al., 2021). Prequestions interspersed within the learning material may provide clearer expectations for learners in terms of the questions' proximity to their answers. For instance, when prequestions are embedded throughout learning material, learners encounter each question immediately before its corresponding solution or answer. This in turn creates a tight question–answer loop that may better facilitate encoding of information. Contrastingly, receiving all prequestions before engaging with the learning material results in greater temporal, proximal, and cognitive distance between questions and their solutions, thereby potentially making it more difficult for learners to connect the two during encoding.

Number of Prequestions The number of prequestions that are received by learners can vary, and there is not a known number of prequestions that optimize performance (though see discussion in Carpenter et al., 2018). Indeed, researchers have given as few as one prequestion (Lenz et al., 2007; Geller et al., 2017; Carpenter et al., 2018) to as many as sixty-four (Taylor, 1996). James and Storm (2019) investigated whether the number of prequestions contributed to their effect on learning. In one experiment, they found that participants who received more prequestions (eight) compared to those who received fewer prequestions (two) performed significantly worse on repeated questions but similarly on new questions.[3]

Variations within the Learning Event

Mode of Information Delivery The mode of information delivery refers to how information is presented during the learning event. Researchers observing the effects of prequestions on performance have used either text (Richland et.al., 2009; James & Storm, 2019; Sana et.al., 2021) or audiovisual materials, such as recorded videos or lectures (Kirschner & Brink, 1979; Toftness et.al., 2018; James & Storm, 2019 see Experiment 5; Lenz et.al., 2007; Carpenter et.al., 2018). Prequestions have been shown to benefit learning of both text and audiovisual materials (Pan & Carpenter, 2023). Although research has not often directly compared the effects of prequestions between text and video materials, James and Storm (2019, Experiment 5) compared the effects of prequestions across these two modes of delivery. They found that the benefits of prequestions on repeated questions was greater for text passages than for videos, but the mode of delivery did not affect new question performance.

Duration of Learning and amount of Text The duration of learning refers to how long the learning event lasted. Some prior work suggests that prequestions may be most effective when the duration of the learning event is shorter rather than longer (James & Storm, 2019; for discussion, see Toftness et al., 2018). For example, Carpenter and Toftness (2017) gave Students short, 2–3-min videos and found that students who received prequestions significantly outperformed those who did not, showing a benefit of prequestions on both repeated and new questions. Subsequent work with longer videos and live lectures found that prequestions benefited performance on repeated questions, but not on new questions (Carpenter et al., 2018; Toftness et al., 2018).

Another variable related to the duration of learning for studies using text-based learning materials is the amount of text, which can be quantified as the number of words in the text. We expect these two variables to be positively correlated with one another (See Table 7 in Appendix). We use both measures because some studies report one variable but not the other; for example, some older studies report the word count of the text but not the time duration of the learning event (Hillman,1972; Rowls, 1975; Ruff, 1975), while newer studies often report both.

Control Conditions Much research compares the performance of a prequestioned group to a non-prequestioned group that receives everything the prequestion group does except for prequestions. However, some studies have employed a second type of control group in which the non-prequestioned group receives a resource or activity that the prequestion group does not. In contrast to a standard (or non-active) control, these active control conditions may have participants receiving extra reading

[3] A follow-up experiment (Experiment 2) did not replicate these results, though they trended in the same direction.

time (e.g., Peeck, 1970), unrelated prequestions (e.g., Pan et al., 2020; Taylor, 1996), or learning objectives (e.g., Sana et al., 2020). Typically, researchers use these active control conditions to control potential confounding variables, such as time on task or time spent with the learning material (Peeck, 1970; Sana et al., 2021).

Content Covered Content covered refers to the subject matter or topic of the learning material (e.g., science, historical fiction, etc.). While prequestions have been tested across a wide array of content areas, we know of no current work that has experimentally tested the potential moderating role of content on learning outcomes. We had no strong reasoning about how the various factors associated with content may interact with prequestions. Thus, we include this moderator to understand how generalizable the effect of prequestions is across different topics.

Variations across Assessment Conditions

Post-Test Format Post-test questions have appeared in the prequestion literature in a variety of formats, such as multiple choice, short answer, fill-in-the-blank, and others. The broader literature on human memory (e.g., Craik et al., 1983) traditionally categorizes these types of questions into types of tests such as (a) *recognition,* in which a learner judges a presented stimulus (e.g., multiple choice, true/false), (b) *cued recall*, in which a learner receives partial information and must retrieve the rest (e.g., fill-in-the-blank), and (c) *free recall*, in which a learner must retrieve all of the information (e.g., essay tests, short answer).

Retention Interval Retention interval refers to the time that passes, if any, between the learning event and the posttest measuring performance. Prior work has shown that the benefits of prequestions on repeated questions can be observed when the posttest occurs immediately (e.g., Smith, 1976; Davidson ,1978; Khoynejad, 1980; McBeady,1982; Richland et al., 2009), is delayed after a distractor (e.g., James & Storm, 2019; St. Hilaire & Carpenter, 2020), and is delayed after several days (e.g., Carpenter et al., 2018; Pan & Sana, 2021; Richland et al., 2009). However, few studies have provided evidence that the benefits of prequestions on new questions consistently occur across different retention intervals (Little & Bjork, 2016; Carpenter & Toftness, 2017; St. Hilaire et.al., 2019; Pan et al., 2020).

Variations across the Participant Sample

Age of Tested Population[4] Age of tested population refers to the age of the learners who completed the study. While no prequestion research that we know of has directly investigated age differences, the broader learning literature would suggest that differences in both learning and performance may arise from age-related differences (Jenkins & Hoyer, 2000; Touron et al., 2001) and may be mediated by variables like skill usage, processing speed, and reasoning (Hoyer

[4] Referred to as Tested Population in the preregistration.

et al., 2003; Rogers et al., 2000; Verhaeghen & Salthouse, 1997). To the degree that children have yet to develop learning strategies or skills compared to older populations, it is possible that prequestions are more beneficial for children due to the additional learning scaffolding that they provide. However, a number of other variables differ across studies involving different age groups (e.g., researchers are likely to adjust their materials to be age appropriate in complexity and duration). Therefore, we explored the potential moderating effect of age group in consideration of how generalizable the effects of prequestions may be across these age ranges.

Prior Knowledge Prior knowledge refers to participants' pre-existing knowledge about the material that they learn in a study. Typically, prequestion researchers control for prior knowledge by creating learning resources that participants could not have knowledge about (e.g., facts about a fictious country; Goldberg, 1980; Rickards et al., 1976) or by surveying participants as to whether they had prior knowledge about the topic such that participants who answered positively were excluded from subsequent analyses (Toftness et al., 2018). However, there are cases in which prior knowledge is not controlled, such as classroom-based studies in which the learning material for the study overlaps with the material students are learning for their class (McDaniel et al., 2011).

Variations across Experimental Setting

Finally, we explored the potential moderating effects of experimental setting, which refers to the space in which participants completed the study. While many prequestion studies have been conducted in a laboratory setting (Karjala, 1984; Little & Bjork, 2016; Carpenter & Toftness, 2017; St. Hilaire, 2017; Hausman & Rhodes, 2018; Pan et al., 2020; Sana et al., 2021), there have also been studies exploring prequestions in the classroom (Hollen, 1970; Richmond, 1973; Patrick, 1976; Benya, 1980; Memory,1983; McDaniel et al., 2011; Carpenter et al., 2018; De Lima & Jaeger, 2020) or online (Janelli, 2019). Because there is no prior work that investigates the differing effects of prequestions in different environmental settings, we include this potential moderator to contribute to our understanding of how the effect of prequestions generalizes across ecological contexts.

Present Study

In the present study, we employ a meta-analysis to synthesize and evaluate findings within the prequestion literature. In doing so, we evaluate (1) Do prequestions promote learning specific to prequestioned content or does this learning benefit extend to non-prequestioned content as well? and (2) Under which specific conditions are prequestions more or less effective?

A recent meta-analysis of the prequestion literature was conducted by St. Hilaire et al. (2024), and while our overall analysis also examines prequestions, there are a few key differences between our analyses. First, the current analysis uses a multi-level meta-analytic approach that controls for both within- and across-study variations, which allows us to include and assess data nested within individual studies. This allows us to maintain calculated effect size independence even when multiple experimental conditions are compared against a singular control condition without having to aggregate experimental condition data as St. Hilaire et al. did. The current approach thus allows for greater precision in calculating effect sizes while accounting for additional sources of variance both within and across studies.

Second, while both meta-analyses investigated some of the same moderators (presentation of prequestions, prequestion placement, mode of information delivery, etc.), the specific levels or categories of those moderators differed across studies in how those categories were coded, moderators had additional levels not shared across analyses, or moderators were categorical rather than continuous. Further, each analysis also explored unique moderators. While St. Hilaire et al. (2024) explored the relation between tested and non-tested material, research design, pre-study phase timing, study phase timing, and match between prequestions and retention questions, we examined the content covered, duration of learning, amount of text material, and experimental setting. Thus, the current approach examines additional variables related to the prequestions themselves, as well as the experimental conditions that may contribute to the effects of prequestions, that have not been explored before. Where our findings diverge from those of St. Hilaire et al., we provide a description of outcome discrepancies in the discussion, as well as a short-hand table summarizing these discrepancies in our Appendix (See Table 7).

Third, while our Boolean search chains differed, the initial records collected in both meta-analyses were roughly similar (2,993 in the current analysis, and 2,137 in St. Hilaire et al., 2024). This would most likely be due to our literature searches occurring the same year (2021) and use of shared databases. However, our screening process resulted in some different records being included in the overall analysis (55 records as compared to 76 records in St. Hilaire et al.) as St. Hilaire and colleagues included abstract books from select cognitive, learning, and educational research conferences,[5] as well as conducted forward and backward literature searches, which were not included in the current work. Overall, these differences between our approaches resulted in 35 records shared across analyses (roughly 64% of total records) and 20 records (35%) unique to the current analyses.[6]

[5] Five records included in their analysis were reported to be from conferences while 4 were from unpublished data.

[6] 17% of records included in the St. Hilaire et al. (2024) analysis did not pass inclusion criteria for the current meta-analyses.

Methods

Transparency and Openness

We adhered to the PRISMA 2020 guidelines for systematic reviews (Page et al., 2021). Our analysis plan, data, and scripts are shared and can be assessed at (https://osf.io/cgbw8/?view_only=d1e145f5f8724362bb93c7ec9f5d5992).

Preregistration

The current work was preregistered on the Open Science Framework (OSF) prior to data collection. Since the preregistration, additional potential moderating variables were identified and added; these are noted below. Additionally, we changed the label "post-question" from the preregistration plan to "repeated question" in the current document because we regard the latter term as a clearer descriptor of the assessment and its relation to the prequestion content. We report how we collected, assessed, and determined eligible records for inclusion into the current meta-analytic models as well as all manipulations and measures below.

Collection, Screening, Eligibility, and Extraction

Collection Relevant articles were identified on July 19th, 2021, by conducting an online search using the following Boolean search chain "Prequest*" OR "Pre-quest*" OR "Adjunct-quest*" OR "Adjunct quest*" OR ("pretest" AND "learning" AND "retrieval practice") in ProQuest and Eric academic search engines. Within the ProQuest search engine, we selected APA PsycInfo, APA Psychnet, Psychology Database, and ProQuest Thesis and Dissertations. 3,116 records (journal articles, theses, dissertations, and books) were identified. Before the screening processes began, 213 duplicates were identified, and were merged or removed, leaving 2,993 records to be screened.

Screening Screening of records occurred in two stages, abstract and full-text level screening. Abstracts were screened based on three inclusion criteria:

1. They described one or more original empirical study. Literature reviews were excluded.
2. They described the use of prequestions as an experimental variable and not as a pre-test for assessing knowledge.[7]

[7] If adjunct questions were used but question placement was not specified, the records were forwarded to full-text screening.

3. They were written in English. Records with abstracts that were only in another language were excluded.

Records without an abstract were automatically moved to full-text screening. The first 300 abstracts were screened in batches of 100 by the first two authors to assess interrater reliability. Cohen's Kappa was calculated for each batch, and all disagreements between raters were discussed and resolved. Once a Cohen's kappa of .90 or above was achieved, the remaining abstracts were screened individually. The final Cohen's kappa was .988, which indicated an almost perfect interrater reliability according to the criteria of Landis and Koch (1977). The same criteria were used during full-text screening along with two additional criteria:

4. Text documentation had to specify question placement/location if not specified in the abstract.
5. Outcome measures needed to be performance on repeated questions and/or new questions.[8]

The first 25 records were screened by the first two authors and showed almost perfect interrater reliability (Cohen's kappa = .905). Given this level of reliability in our screening procedures, the remaining texts were then screened by the lead author alone. After screening procedures had been completed, a total of 2,895 records were excluded. This left 98 records to be evaluated for eligibility into the meta-analysis.

Eligibility Prior to data extraction, we assessed whether records reported the statistics required to conduct the meta-analysis. Specifically, the results needed to include statistics that permitted the calculation of Hedges' *g* scores (e.g., *n*, *M*/*SD* or *SE*, *t*, *F*). If these statistics were missing but the raw data were reported, the desired statistics were calculated based on those data and included in the analysis. If these statistics were not reported or were incomplete (e.g., reporting means without variation) and the data were not available, the authors were contacted to request missing or unreported data. Due to lack of required statistical information, forty-one records were excluded (12 due to unreported statistics, 16 due to unusable statistics, 3 duplicate records, and 10 due to absence of a control condition), leaving a total of 57 records for inclusion in the meta-analysis.[9]

Extraction If a record contained multiple experiments or experimental groups of interest, then multiple comparisons could be extracted from a single record. For instance, if an experiment had three groups (a prequestion group, a prequestion supplemental group, and a control group), then two comparisons could be extracted (prequestion vs. control, prequestion supplemental vs. control). Alternatively, if a record had two experiments of interest, each with a single experimental and control

[8] The terminology changed over the years. For example, some descriptions of repeated questions used the label "relevant aligned", and descriptions of new questions used the labels "non-relevant" or "incidental."

[9] While 7 of these records were excluded based on the present criteria, they were included in the St. Hilaire et al. (2024) analysis.

group, then two comparisons could be extracted. Two hundred and six comparisons were extracted from the initial 57 records that were to be included in the meta-analysis. See Fig. 2 below for a detailed PRISMA diagram of the above procedures.

Coding

Initial coding of all records was conducted by the lead author, with coding encompassing the extraction of moderating variables (e.g., type of prequestion, duration of learning, retention interval), reported statistics for effect size calculations (e.g., *n*, *M*/*SD* or *SE*, *t*, *F*), and descriptive information about each record (e.g., record type, author, publication status). Following this, 39 comparisons (20% of the data) were randomly selected to be independently coded by the third author. This second round of coding only extracted moderating information and truncated statistics (i.e., total *N* and sample sizes for experimental and control conditions). Calculated kappas from the second round of extraction varied substantially across the potential moderators, ranging from .18 (slight agreement) to 1.00 (perfect agreement). In particular, initial agreement was low for four specific moderator variables: duration of learning, type of prequestion, number of prequestions, and presentation of prequestions. Given this low level of agreement, coders resolved disagreements through discussion and re-evaluated coding criteria for rule clarification. Once completed, a third round of coding saw the remaining 80% of the data re-coded for moderators that had seen low agreement in the second round. Specifically, all of these remaining 80% of records were recoded by the first author, 40% of the data was independently coded by a research assistant, and the other 40% of the data was independently coded by another research assistant. Disagreements in coding were resolved through discussion.

During these rounds of coding and agreement checks, it was revealed that two studies included in the analysis did not make proper comparisons between a prequestion and control condition. This left a total of 55 records (204 comparisons) to be included in our meta-analytic models.[10]

Assessing Publication Bias

Since there is no definitive operation to assess and correct publication bias, it is recommended to use multiple methods to estimate its effects (Kepes et al., 2023). Within the current overall models, we assess publication bias using three methods: 1) PET-PEESE, 2) Visual inspection of funnel plots, and 3) Assessing publication status (published or unpublished) as a potential moderator.

Statistical Models and Effect Size Calculations

The primary effect size measure was the standardized mean difference (i.e., Hedges' *g*). To calculate effect sizes, we extracted sample sizes, means, and standard deviations from studies that compared performance across groups (e.g., prequestion vs. control). Where

[10] Figure 2 above was updated to reflect the accurate number of records included in our meta-analyses.

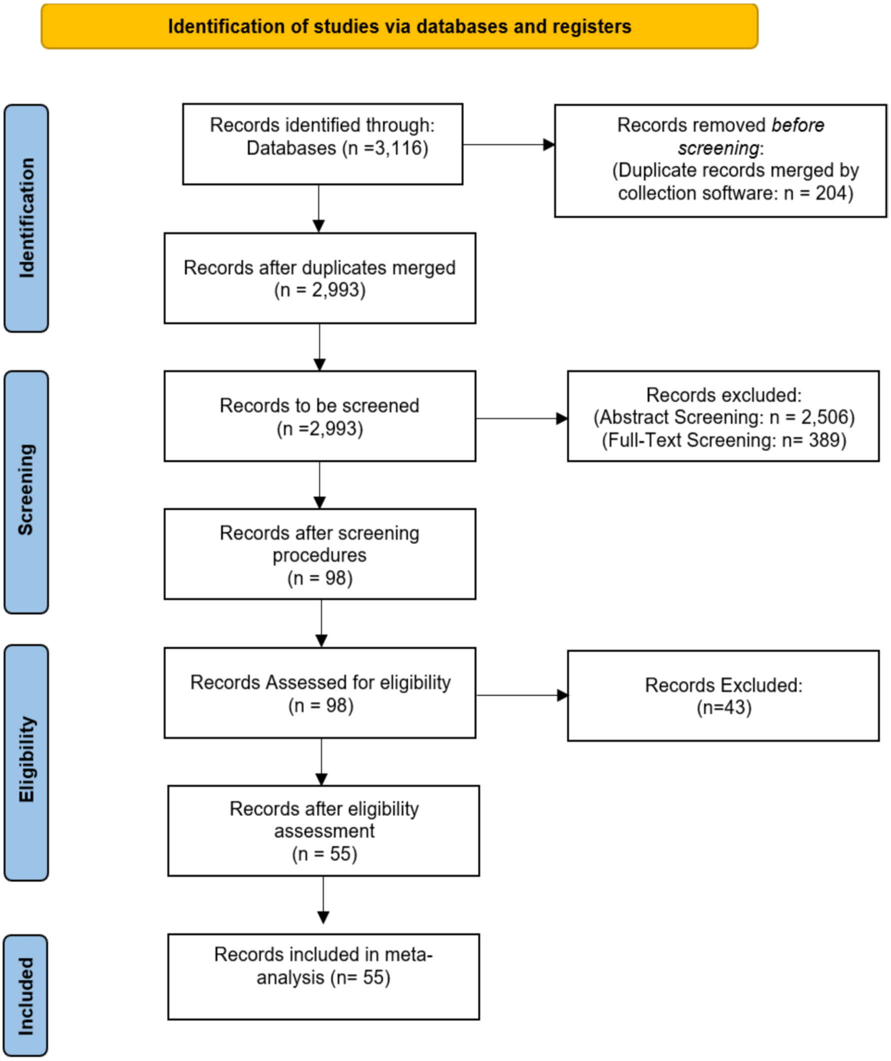

Fig. 2 PRISMA flow diagram of record identification, screening, eligibility, and inclusion

necessary, effect sizes were computed from other reported statistics (e.g., F or t). As multiple comparisons could be nested in one study (e.g., multiple experimental groups compared to one control) and thereby contribute multiple effect sizes, a multi-level model was used to account for likely effect size dependence. In doing so, Level 1 of our model pools Hedges' g scores calculated from our comparisons (k). Level 2 then nests these effect sizes within studies from which they were calculated, accounting for when one study contributes multiple effect sizes. We refer to this level as the within-study cluster level. Lastly, at Level 3, these effect sizes are assessed across these study clusters to determine the overall true Hedges' g score. This level is referred to as the between-study cluster level. Within-study cluster heterogeneity ($\tau^2_{level\ 2}$) on Level 2 and between-study cluster heterogeneity ($\tau^2_{level\ 3}$) on Level 3 were evaluated by their respective τ^2 scores. Additionally, we computed I^2 at each level to quantify the proportion of variation that can be attributed to within- or between-study cluster heterogeneity. All analyses were conducted using the *metafor* (Viechtbauer, 2010) and *robumeta* (Fisher & Tipton, 2015) package in R statistical software. The restricted maximum likelihood method was used to estimate the parameters of our meta-analytic models. Subgroup analyses were then conducted to evaluate the influences of moderator variables on the prequestion effect.

Potential Moderating Factors

Variation of the Prequestions

Type of Prequestion We classified types of prequestions under two general categories: factual and conceptual. We coded *factual* prequestions as questions that can be answered from verbatim information explicitly stated in the learning materials. These typically can be answered with a word, phrase, or sentence that comes directly from material within the learning resource. We coded *conceptual* prequestions as any question that cannot be answered with verbatim information from the learning resource and instead requires the generation of new words, phrases, or sentences to articulate the answer.

Presentation of Prequestions[11] We coded five levels in which the presentation of a prequestion could be categorized: prequestion alone, prequestion minus, prequestions optional, prequestions plus feedback, and prequestion plus supplemental activity. We refer to prequestioned groups that were instructed to read and answer prequestions without being given any supplemental material as *prequestions alone*. This is the most common prequestion condition that is tested in the literature. By comparison, we coded a condition as *prequestions optional* if the act of answering prequestions was optional for participants who received them, and as *prequestions minus* if prequestions were given but participants were specifically instructed *not* to generate an answer to them. If a prequestion group received feedback in which participants were presented the solution to prequestions regardless of answering correctly or incorrectly, then they were categorized as *prequestions plus feedback*. Lastly, if a prequestion group received an additional learning aid or material along with prequestions, the condition was coded as *prequestion plus supplemental activity*. While there were several such supplemental activities,[12] the frequency of each individual category was below the threshold ($k < 10$) to make reasonable individual comparisons.

[11] This variable had been referred to as “experimental conditions” in the preregistration.

[12] Supplemental activities included participants taking notes, participants discussing as a group, participants having prequestions available during the learning event, participants being instructed to write prequestions from memory before or after the learning event, participants being able to refer to the learning passages while answering prequestions, participants having access to a content map during the learning event, having key information within the learning content bolded, and participants receiving study objectives in addition to prequestions.

Prequestion Placement We classified prequestions given prior to the learning event as *prequestions before learning*. Prequestions that were given during the learning event but prior to targeted material were categorized as *prequestions interspersed*. For example, participants might answer prequestions related to content in paragraph one, then read paragraph one, answer prequestions related to content in paragraph two, then read paragraph two, and so on. Lastly, prequestions that appeared both prior to the learning event *and* just prior to targeted material within the learning event were categorized as *prequestions before and interspersed.*

Number of Prequestions The number of prequestions refers to the amount of prequestions participants received prior to interacting with learning material. In cases where prequestions were interspersed throughout the material, we counted only the prequestions presented before the specific to-be-learned section. For example, Carpenter and Toftness (2017) gave participants two prequestions prior to each of three video segments. Thus, the number of prequestions would be two. Alternatively, St. Hilaire and Carpenter (2020) gave participants 15 prequestions prior to interacting with any learning material. Thus, the number of prequestions would be 15.

We categorized the number of prequestions, based on a quartile split, into four levels: small, medium, large, and very large. A *small number* of prequestions was fewer than 2 prequestions ($M=1.06$, $Mdn=1$), a *medium number* between 2–4 prequestions ($M=2.4$, $Mdn=2$), a *large number* between 5–12 prequestions ($M=8.84$, $Mdn=10$), and a *very large number* between 13–64 prequestions ($M=21.14$, $Mdn=20$).

Variation within the Learning Event

Mode of Information Delivery[13] In the initial extraction of records, we coded for the way in which information was delivered in the study (e.g., lecture, video, text, radio, and tape-recording format), which we subsequently grouped into two categories: text and audiovisual. Text-based delivery was defined as information that contains purely text-based elements (e.g., paragraphs from a textbook or hyper-text). Audiovisual-based delivery was defined as any mode of information delivery containing audio information (e.g., tape-recording, narrated videos) but could also include hybrid formats that contained text (e.g., videos or in-class lectures).[14]

Duration of Learning[15] The duration of learning refers to the amount of time that participants interacted with the learning material. Similar to the number of prequestions, this moderator encompassed the duration of time participants spent on the learning

[13] This variable had been referred to as "stimuli format" in the preregistration.

[14] This category was created because other reported formats had k counts too low for comparison (i.e., $k=5$ for oral lecture and $k=8$ for tape-recording).

[15] This variable had been referred to as "length of learning event" in the preregistration.

material immediately following prequestions, regardless of whether prequestions were all given before interacting with learning material or interspersed throughout the learning material. Returning to our example, Carpenter and Toftness (2017) gave participants two prequestions prior to each of three video segments lasting about 2.5 min each. Thus, the duration of time spent on the learning material immediately following prequestions was 2.5 min. Alternatively, St. Hilaire and Carpenter (2020) gave participants 15 prequestions prior to interacting with learning material for 31 min. Thus, the duration of time spent on the learning material immediately following prequestions was 31 min.

We classified the duration of the learning event into four ordinal categories based on a quartile split of the reported learning durations. A *short learning event* ranged from 1–5 min ($M = 1.73$, $Mdn = 2$), a *medium learning event* ranged from 6–8 min ($M = 7.07$, $Mdn = 7.33$), a *long learning event* ranged from 9–16 min ($M = 12.13$, $Mdn = 12$), and an *extended learning event* ranged from 17–100 min ($M = 30.8$, $Mdn = 30$).[16]

Amount of Text Material The amount of text material refers to the number of words participants read after receiving prequestions. Similar to the duration of learning, this moderator encompassed the number of words participants read during the learning event immediately following prequestions to account for both instances in which prequestions were given all at once before interacting with learning material, or those in which prequestions were interspersed throughout the learning material. For instance, Hausman and Rhodes (2018) gave participants three prequestions prior to individual text segments that were on average 710 words long. Therefore, the amount of text material would be 710 words. Alternatively, St. Hilaire et al. (2019) had participants in the prequestioned conditions receive all prequestions before interacting with a singular text segment that was 675 words. Therefore, the amount of text material would be 675 words.

For studies using text-based materials, we further classified the amount of text based on a quartile split into four categories. A *small amount* of text constituted any word count between 1–358 words ($M = 232.81$, $Mdn = 250$), a *medium amount* of text was any word count between 359–710 words ($M = 594.98$, $Mdn = 648$), a *large amount* of text was any word count between 711–1,696 words ($M = 955.35$, $Mdn = 925$), and a *very large amount* of text was any word count between 1,697–15,000 words ($M = 3{,}366$; $Mdn = 2{,}230$).[17]

Control Conditions We categorized control conditions into two levels: standard and active. We defined a *standard* control group as a group that receives everything that

[16] In cases in which a range of time was reported (i.e., 5–7.5 min per video or 4 min per passage) the maximum amount of time that could elapse was used in the quartile split and descriptive statistic calculations.

[17] Records indicating other units of word count (e.g., pages, paragraphs) were estimated based on average word count of 150 words per paragraph or 500 words per page. Researchers that reported word length as a unit of passages were excluded unless word count was specified or could be obtained through given tested material ($k = 11$).

the experimental group receives (e.g., the same learning event and post-test), but does not receive prequestions nor any other type of activity prior to the learning event. We defined an *active* control group as a group that receives some kind of additional activity prior to or during the learning event. For example, in Pan et al. (2020), students in the prequestion group received prequestions about the material that they were about to learn, whereas an active control group received prequestions about a topic completely unrelated to what they were about to learn. Further, we coded for specific active control groups that were either *related* to the learning resource content (e.g., more study time, learning objectives) or *unrelated* to the learning resource content (e.g., prequestions for a different topic).

Content Covered We classified the content covered into three broad categories: STEM, non-STEM expository, and non-STEM narrative. We defined STEM content as any information related to science, technology, engineering, and mathematics disciplines (e.g., biology, communication theory, rocketry, vehicle systems, etc.) and all other content as non-STEM. Non-STEM material was further classified into either expository if it provided fact-like[18] information to describe or explain a specific topic (e.g., history, foreign language comprehension) or as narrative if it was structured as a story (e.g., fictional literature, historical fiction, mystery fiction).[19]

Variations across Assessment Conditions

Post-Test Format We categorized these formats as *free recall* (e.g., short answer tests), *cued recall* (e.g., fill-in-the blank tests), *recognition* (e.g., multiple-choice tests), or *varied.* Our varied format captured post-tests that included more than one form of test question (e.g., multiple choice and fill-in the blank, or free recall and yes/no questions).

Retention Interval[20] We coded *immediate testing* as any post-test measuring performance just following the learning phase. A post-test that occurred the same day but after the completion of a distractor task was coded as *testing after a distractor task*. Lastly, we coded a post-test that occurred 1 day or more after the learning event as *delayed testing*.

Variation across the Participant Sample

Age of Tested Population We coded for three categories: children, teenagers, and adults. We defined *children* as students from kindergarten to sixth grade, *teenagers* from seventh to twelfth grade, and *adults* from college and beyond.

[18] We state "fact-like" because some text provided "facts" regarding fictional events.

[19] There were some cases in which the stimuli covered both STEM and Non-STEM material. However, the number of such instances ($k < 10$) were too small to make proper comparisons within our new question performance sub-model, and thus was dropped..

[20] This variable had been referred to as "time of post-test" in the preregistration.

Prior Knowledge We classified studies as prior knowledge *controlled* if the researchers used methods to minimize, reduce, or eliminate the potential effects of prior knowledge. These methods included creating fictional learning content (e.g., facts about a fake country, Goldberg, 1980; Manchester, 1984; Rickards, 1972), specific participant selection[21] (Boker, 1974; Bull & Dizney, 1973; Peeck, 1970), content creation that was expected to reduce familiarity with the to-be-learned content[22] (Chimezie, 1970; Rauscher, 1978; Chobot, 1984; Dickerson, 1987), or experimental criteria that excluded participant data from subsequent analyses (Richland et al., 2009; Carpenter & Toftness, 2017; Toftness et al., 2018; Sana et al., 2020; St. Hilaire & Carpenter, 2020). Alternatively, we categorized records as *uncontrolled* if the researchers explicitly reported being unable to control for or implement techniques to reduce prior knowledge. For example, studies conducted in schools may not have been able to control for prior knowledge due to the need for the learning resource to be aligned with course content.[23]

Variation across Experimental Setting

We classified the experimental setting on two general levels: laboratory and school. Further, we coded whether school-based experiments used learning materials that were either *related* (e.g., study materials aligned with class content) or *not related* to the school's educational curriculum (e.g., a study occurred at a school but learning materials in the study were not aligned with class content).

Results

Figure 3 displays the number of records included in our analysis, spanning the past 50 years. We found that the effect of prequestions was prominently investigated from 1970 to 1989, and from 2010 to 2021 has begun to have a resurgence. From the 55 records that were included in our analysis, data from 12,003 participants[24] were extracted. We interpret effects as significant if $p < .05$ and as marginal if $.05 < p < .10$.

[21] Selecting participants who would have general familiarity but not specific knowledge (Bull & Dizney), only subjects who had never been to Greece were allowed to enlist due to content being about Greek rural life (Peeck), students who majored in fields concerned with biology and historical geology were not permitted to participate in the study due to learning content being focused in those areas (Boker).

[22] Content used here was not fictitious in nature but chosen/created based on how little participants are likely to be familiar with the content. Fictitious content, however, is created with the express purpose that no participant could have familiarity with the content because it is imagined.

[23] Records were categorized as *undefined* when researchers did not explicitly report using a technique nor state their reasoning for not controlling for prior knowledge. This level was removed from subsequent analysis as there could be no reasonable interpretations made regarding a category for which it could not be determined whether participants had prior knowledge of the learning content.

[24] 6,704 participants were in experimental groups; 5,299 participants were in control groups.

Publication Bias

Adjustments within the *robumeta* package were made to fit linear regression models that estimate study precision (PET) and adjust for potential small-study bias (PEESE). If the regression intercepts are significantly different from 0, then the PEESE model should be used to assess publication bias; otherwise, the PET model should be used.

For repeated questions, the regression intercepts for both the PET and PEESE analysis were significantly different from 0, so the PEESE test was used to assess publication bias and estimate the intercept of the true effect size (g=.68, 95% CI [.53,.83], p< .001). The results of our PEESE test suggest that publication bias is likely to exist in our model of performance on repeated question performance. For new questions, the regression intercepts for both the PET and PEESE analysis did not significantly differ from 0, indicating that the PET test should be used to assess publication bias and estimate the intercept of the true effect size (g= .03, CI [−0.16, 0.22], p= .74). This result indicates that publication bias is not likely to exist in our model of new question performance.

Funnel plots were generated to visually assess publication bias. If bias is present, then data points on the funnel plots will appear asymmetrical, while if bias is absent then data points will appear symmetrical. The funnel plot for repeated question performance (Fig. 4) appears asymmetrical while the funnel plot for new question performance (Fig. 5) appears symmetrical. This visual inspection may indicate that papers displaying an effect of prequestions on repeated question performance are published more often than papers that do not display an effect.

In the meta-analysis itself, the magnitude of the prequestion effect on repeated question performance did not significantly differ when comparisons were extracted from a published (g= .77, k=76) or unpublished source (g= .56, k=105), F (1, 179)=2.07, p= .15. This means that publication status is not a significant moderator of the relationship between prequestions and repeated question performance. Similarly, there was not a significant effect of publication status on new question performance, F(1, 173)=2.13, p= .14. The magnitude of the prequestion effect on new question performance did not significantly differ between published sources (g= .10, k=68) and unpublished sources (g=−0.05, k=107).

In summary, our assessment of publication bias employed three methods: 1) PET-PEESE, 2) visual inspection of funnel plots, and 3) assessing publication status (published or unpublished) as a potential moderator. Our PET-PEESE models and funnel plots suggest that publication bias is present for repeated question performance, however, treating publication bias as a moderator did not provide evidence for publication bias. Additionally, all of our analyses suggest that publication bias is not likely to exist for new question performance.

The presence of publication bias in two out of three measures raises the concern that there is potential overestimation of effect sizes found in our repeated question performance outcomes. When interpreting the results of our overall repeated question model, interpretations may need to be made using the adjusted true effect size calculated by our PEESE model. The adjusted true effect of prequestions on repeated question performance was g= .68, 95% CI [.53, .83], p< .001. That is,

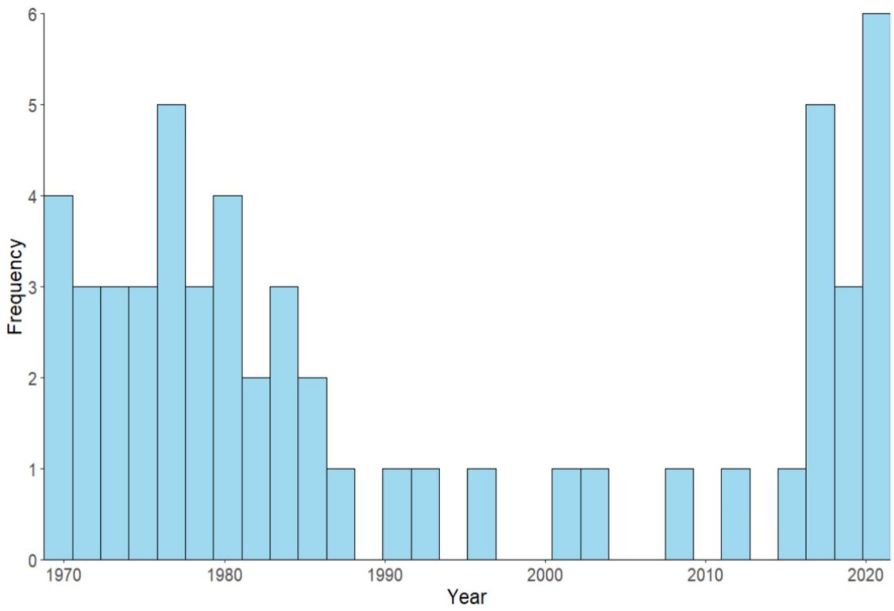

Fig. 3 The frequency of prequestion studies conducted from 1970–2021**.** *Note*. Our search was conducted in July 2021, so the graph only includes records available from January to July 2021

across comparisons, prequestions benefited performance on repeated questions, with a medium-to-large effect size according to the criteria of Cohen (1988).

Research Question 1: Overall Effects

Table 1 displays statistics for the meta-analyses of repeated and new question performance. The pooled effect size of the true effect of prequestions on repeated question performance was g=.66, CI [0.52, 0.81], (df = 45, k = 181), p < .0001. This shows that, across comparisons, prequestions benefited performance on repeated questions, with a medium-to-large effect size.

In contrast, the pooled effect size of prequestions for new question performance was g = 0.01, with a 95% CI [−0.09, 0.12] (df = 45, k = 175), p=.80. This result shows that, across comparisons, there is no effect of prequestions on new question performance.

For repeated questions, the estimated variances were $\tau^2_{\text{level 2}} = 0.01$ CI [0.00, 0.06] and $\tau^2_{\text{level 3}} = 0.14$ CI [0.07, 0.27], showing that the between-cluster heterogeneity was significantly different from 0 while our within-cluster heterogeneity was not. Further, $I^2_{\text{level 1}} = 61.05\%$ of the total variation can be attributed to between-comparison heterogeneity, $I^2_{\text{level 2}} = 3.26\%$ to within-cluster heterogeneity, and $I^2_{\text{level 3}} = 35.68\%$ to between-cluster heterogeneity. These scores indicate that a moderatelevel of heterogeneity is attributed to variability in calculated effect sizes extracted from all studies,

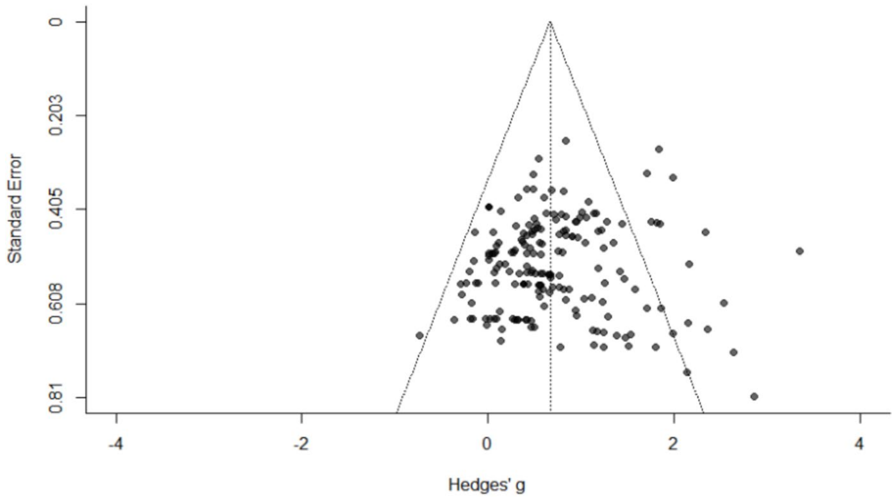

Fig. 4 Funnel plot of standard error by Hedges' g for the repeated question multilevel model

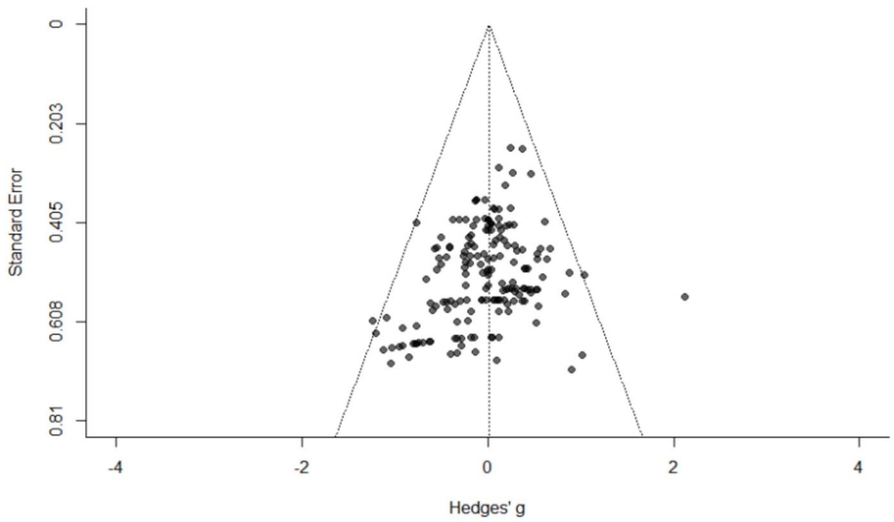

Fig. 5 Funnel plot of standard error by Hedges' g for the new question multilevel model

virtually no observed heterogeneity in calculated effect sizes extracted from the same study, and lastly a moderate level of heterogeneity between studies.

For new questions, the estimated variances were $\tau^2_{\text{level 2}}=0.00$ CI [0.00, 0.01] and $\tau^2_{\text{level 3}}=0.05$ CI [0.01, 0.12], showing that the between-cluster heterogeneity was significantly different from 0 while the within-study cluster heterogeneity was not. This is further reflected in $I^2_{\text{level 1}}=82\%$ of the total variation being attributed to between-comparison heterogeneity, $I^2_{\text{level 2}}=0\%$ to within-cluster heterogeneity,

Table 1 Repeated and New Question Multi-level Model Statistics

Models		*k*	*g*	SE	CI
Repeated Question	Random Effects Intercept	181	.66*	.07	(0.52, 0.81)
	Variance Components				
	Second Level	179	.01	-	(0.00, 0.06)
	Third Level	46	.14	-	(0.07, 0.27)
New Question	Random Effects Intercept	175	.01	.05	(−0.09, 0.12)
	Variance Components				
	Second Level	175	.00	-	(0.00, 0.01)
	Third Level	46	.05	-	(0.01, 0.12)

*$p < .05$

and $I^2_{\text{level 3}} = 17.93\%$ to between-cluster heterogeneity. These scores indicate that a substantial level of heterogeneity is attributed to variability in calculated effect sizes extracted from all studies,y no observed heterogeneity in calculated effect sizes extracted from the same study, and lastly a significant but low level of heterogeneity between studies.

The significant τ^2 and moderate I^2 statistics for both our repeated and new question performance models lend further support for our use of random-effects models, which account for this heterogeneity. They also indicate that the size of each effect may be influenced by moderating characteristics, which we analyze below.

Research Question 2: Moderator Analyses

The results of our moderator analyses for repeated and new question performance are reported in Tables 2, 3, 4, 5 and 6. We report the number of comparisons within moderator levels (k) and the estimated effect at each level (g). Additionally, we report the sub-analytic model's F-statistic as a measure of significant differences between levels and the amount of within-cluster heterogeneity ($\tau^2_{level\ 2}$) and between-cluster heterogeneity ($\tau^2_{\text{level 3}}$) as well as their respective confidence intervals.[25]

Moderators Related to the Prequestions

Type of Prequestion The magnitude of the prequestion effect on repeated question performance did not significantly differ depending on whether prequestions were considered factual ($g = 0.71$, $k = 148$), conceptual ($g = 0.45$, $k = 18$), or consisted

[25] The number of comparisons in our quartile split moderator variables (i.e., number of prequestions, duration of learning, amount of text material) varied from moderator to moderator across levels. This was due to some comparisons sharing the same values near the quartile mark (e.g., many comparisons only gave 1 prequestion). Additionally, when filtering for the specific models (repeated or new) these levels could further deviate from their original split count. This is why comparison counts (k) are not exactly 25% of the total number of comparisons at each level.

Table 2 Moderator Analysis for Potential Moderators Related to Prequestions

Model	Moderators	Level	*k*	*g*	*F*	*p*	$\tau^2_{level\ 2*}$	$\tau^2_{level\ 3*}$
	Type of Prequestion				1.70	.18	0.01 [0.00-0.07]	0.14 [0.07-0.28]
		Factual	148	0.71				
		Conceptual	18	0.45				
		Both	10	0.38				
Repeated Question								
	Presentation of Prequestions				6.20	<.001	0.00 [0.00-0.03]	0.14 [0.07-0.26]
		Prequestions Alone	110	0.60				
		Prequestions Minus	17	0.65				
		Prequestions Plus SA*	24	0.84				
		Prequestions Plus FB**	26	1.25				
	Prequestion Placement				0.01	.91	0.01 [0.00-0.06]	0.14 [0.07-0.27]
		Prequestions Before Learning	88	0.65				
		Prequestions Interspersed	91	0.67				
	Number of Prequestions				.13	.94	0.01 [0.00-0.06]	0.15 [0.08-0.30]
		Small	44	0.68				
		Medium	44	0.71				
		Large	48	0.65				
		Very Large	40	0.60				
New Question	Type of Prequestion				5.39	<.01	0.00 [0.00-0.01]	0.03 [0.00-0.09]
		Factual	137	0.002				

Table 2 (continued)

Model	Moderators	Level	*k*	*g*	*F*	*p*	$\tau^2_{level\ 2*}$	$\tau^2_{level\ 3*}$
		Conceptual	18	0.34				
		Both	10	−0.40				
	Presentation of Prequestions				0.22	.88	0.00 [0.00-0.01]	0.05 [0.01-0.13]
		Prequestions Alone	114	0.03				
		Prequestions Minus	15	−0.06				
		Prequestions Plus SA*	17	0.05				
		Prequestions Plus FB**	25	−0.04				
	Prequestion Placement				1.14	.28	0.00 [0.00-0.01]	0.04 [0.01-0.12]
		Prequestions Before Learning	85	0.06				
		Prequestions Interspersed	89	−0.04				
	Number of Prequestions				.81	.48	0.00 [0.00-0.01]	0.04 [0.01-0.12]
		Small	38	−0.13				
		Medium	45	0.002				
		Large	46	0.09				
		Very Large	41	0.001				

*SA** = Supplemental Activity, *FB*** = Feedback

Table 3 Moderator Analysis for Potential Moderators Related to the Learning Event

Model	Moderator	Level	*k*	*g*	*F*	*p*	$\tau^2_{\text{level 2*}}$	$\tau^2_{\text{level 3*}}$
Repeated Question	Mode of Information Delivery				0.02	.88	0.01 [0.00–0.06]	0.14 [0.07–0.28]
		Text	155	0.67				
		Audiovisual	26	0.64				
	Duration of Learning				0.07	.97	0.00 [0.00–0.03]	0.06 [0.01–0.20]
		Short	20	0.67				
		Medium	27	0.62				
		Long	15	0.66				
		Extended	18	0.71				
	Amount of Text				.59	.62	0.00 [0.00–0.02]	0.15 [0.07–0.32]
		Small	38	0.57				
		Medium	28	0.54				
		Large	36	0.58				
		Very Large	35	0.79				
	Control Conditions				0.41	.66	0.01 [0.00–0.06]	0.15 [0.07–0.28]
		Standard	147	0.68				
		Active-Related	23	0.54				
		Active-Unrelated	11	0.75				

Table 3 (continued)

Model	Moderator	Level	*k*	*g*	*F*	*p*	$\tau^2_{\text{level 2*}}$	$\tau^2_{\text{level 3*}}$
	Content Covered				0.61	.60	0.01 [0.00–0.06]	0.15 [0.07–0.30]
		STEM	47	0.72				
		Non-STEM exp.	97	0.66				
		Non-STEM narr.	30	0.62				
		STEM & Non-STEM	10	0.40				
New Question	Mode of Information Delivery				0.73	.39	0.00 [0.00–0.01]	0.04 [0.01–0.12]
		Text	153	–0.005				
		Audiovisual	22	0.10				
	Duration of Learning				0.08	.96	0.00 [0.00–0.03]	0.05 [0.00–0.25]
		Short	14	0.18				
		Medium	23	0.11				
		Long	10	0.04				
		Extended	20	0.10				
	Amount of Text				0.14	.92	0.00 [0.00–0.01]	0.03 [0.00–0.13]
		Small	23	–0.03				
		Medium	39	0.05				
		Large	32	–0.02				
		Very Large	39	–0.03				

Table 3 (continued)

Model	Moderator	Level	*k*	*g*	*F*	*p*	$\tau^2_{level\ 2*}$	$\tau^2_{level\ 3*}$
	Control Conditions				2.68	.07	0.00 [0.00–0.01]	0.04 [0.01–0.12]
		Standard	145	0.01				
		Active-Related	18	–0.23				
		Active-Unrelated	12	0.29				
	Content Covered				1.46	.23	0.00 [0.00–0.01]	0.05 [0.01–0.13]
		STEM	52	0.12				
		Non-STEM expo.	90	–0.05				
		Non-STEM narr.	27	–0.08				

Table 4 Moderator Analysis for Potential Moderators Related to the Assessment Conditions

Model	Moderators	Level	*k*	*g*	*F*	*p*	$\tau^2_{level\ 2*}$	$\tau^2_{level\ 3*}$
Repeated Question	Post-Test Format				0.42	.65	0.01 [0.00–0.07]	0.15 [0.07–0.29]
		Free-Recall	33	0.73				
		Cued-Recall	45	0.58				
		Recognition	100	0.65				
	Retention Interval				2.06	.13	0.01 [0.00–0.06]	0.14 [0.07–0.27]
		Immediate	103	0.64				
		Delayed after Distractor	45	0.58				
		Delayed 1 Day +	33	0.87				
New Question	Post-Test Format				.06	.93	0.00 [0.00–0.01]	0.05 [0.01–0.13]
		Free-Recall	24	–0.03				
		Cued-Recall	40	–0.00				
		Recognition	109	0.01				
	Retention Interval				1.21	.30	0.00 [0.00–0.01]	0.04 [0.00–0.11]
		Immediate	106	–0.01				
		Delayed after Distractor	41	0.11				
		Delayed 1 Day +	28	–0.06				

Table 5 Moderator Analysis for Potential Moderators Related to the Participant

Model	Moderator	Level	k	g	F	p	$\tau^2_{level\ 2*}$	$\tau^2_{level\ 3*}$
Repeated-Question	Age of Tested Population				0.69	.50	0.01 [0.00–0.07]	0.14 [0.07–0.28]
		Children	19	0.52				
		Teens	27	0.51				
		Adults	135	0.71				
	Prior Knowledge				0.36	.54	0.10 [0.01–0.26]	0.10 [0.01–0.33]
		Controlled	46	0.66				
		Uncontrolled	21	0.80				
New-Question	Age of Tested Population				0.14	.86	0.00 [0.00–0.01]	0.05 [0.01–0.13]
		Children	24	0.001				
		Teens	28	0.08				
		Adults	123	0.002				
	Prior Knowledge				0.71	0.40	0.00 [0.00–0.02]	0.08 [0.00–0.29]
		Controlled	47	–0.00				
		Uncontrolled	17	0.16				

Table 6 Moderator Analysis for Potential Moderators Related to the Experimental Setting

Model	Moderators	Level	*k*	*g*	*F*	*p*	$\tau^2_{\text{level 2*}}$	$\tau^2_{\text{level 3*}}$
Repeated Question	Experimental Setting$_a$				0.36	.54	0.01 [0.00–0.07]	0.15 [0.07–0.28]
		Laboratory	87	0.71				
		School	93	0.62				
	Experimental Setting$_b$				0.67	.41	0.00 [0.00–0.10]	0.22 [0.09–0.58]
		Related	18	0.46				
		Not-Related	75	0.68				
New Question	Experimental Setting$_a$	Laboratory	82	0.08	2.09	.14	0.00 [0.00–0.01]	0.04 [0.01–0.12]
		School	92	–0.06				
	Experimental Setting$_b$	Related	13	–0.03	0.06	.80	0.00 [0.00–0.02]	0.04 [0.00–0.14]
		Not-Related	79	–0.07				

of a mix of factual and conceptual questions ($g=0.38$, k = 10), $F(2, 173)=1.70$, $p=.18$. However, the magnitude of the prequestion effect on new question performance did significantly differ depending on the type of prequestions, $F(2, 166)=5.39$, $p<.01$. When prequestions contained a mix of factual and conceptual questions, the prequestion effect was *negative* ($g=-0.40$, $k=10$) and significantly less than when prequestions were only factual ($g=0.002$, $k=141$; $t(166)=2.12$, $p<.05$) or only conceptual ($g=0.34$, $k=18$; $t(166)=3.23$, $p<.01$). Further, the magnitude of the prequestion effect was greater for conceptual prequestions ($g=0.33$) than factual ($g=0.003$), $t(157)=2.23$, $p<.05$).

Presentation of Prequestions The magnitude of the prequestion effect on repeated question performance significantly differed depending on how participants received and were instructed to interact with the prequestions, $F(3, 173)=6.20$, $p<.001$.[26] More specifically, the magnitude of the prequestion effect on repeated questions was significantly greater when prequestions were given with feedback (prequestion plus feedback, $g=1.25$, $k=26$, $t(173)=7.64$, $p<.001$), compared to when prequestions were answered without feedback (prequestions alone, $g=.60$, $k=110$). Further, there was a marginal difference between when prequestions were given with a supplemental activity (prequestions plus supplemental activity, $g=0.84$, $k=24$) compared to when prequestions were given alone, $t(173)=1.79$, $p=.07$.[27] However, the magnitude of the effect of prequestions on repeated questions did not significantly differ when prequestions were given in typical fashion compared to when prequestions were presented without requiring an answer (prequestions minus, $g=.65$, $k=17$; $t(173)=.36$, $p=.71$).

The magnitude of the prequestion effect on new question performance did not significantly differ between those comparisons in which prequestions were given alone ($g=0.03$, $k=114$), when prequestions were presented but not answered (prequestions minus, $g=-0.06$, $k=15$), when feedback was given after answering prequestions (prequestions plus feedback, $g=-0.04$, $k=25$), or when prequestions were given with a supplemental activity (prequestions plus supplemental activity, $g=0.05$, $k=17$); $F(3, 167)=0.22$, $p=.88$.

Prequestion Placement The magnitude of the prequestion effect on repeated question performance did not significantly differ when prequestions were all placed before learning ($g=0.65$, $k=88$) or were interspersed within the learning event but before targeted information ($g=0.67$, $k=91$) $F(1, 177)=0.01$, $p=.91$. Further, the magnitude of the prequestion effect on new questions did not significantly differ when all prequestions were placed before learning ($g=0.06$, $k=85$)

[26] Prequestions alone was set as the reference level for our sub-analytic models as it is the most typical experimental group assessed within the literature.

[27] The comparison level of "prequestion optional" (i.e., a level in which the act of answering prequestions was optional for participants who received them) was dropped in both our repeated ($k=4$) and new ($k=4$) models due to having fewer than 10 comparisons.

or if they were interspersed within the learning event but before targeted information ($g = -0.04$, $k = 89$) $F(1, 172) = 1.14$, $p = .28$.[28]

Number of Prequestions The magnitude of the prequestion effect on repeated question performance did not differ according to whether the dosage of prequestions was small ($g = 0.68$, $k = 44$), medium ($g = 0.71$, $k = 44$), large ($g = 0.65$, $k = 48$), or very large ($g = 0.60$, $k = 40$), $F(3, 172) = 0.13$, $p = .94$. Similarly, the magnitude of the prequestion effect on new question performance did not differ according to whether the dosage of prequestions was small ($g = -0.13$, $k = 38$), medium ($g = 0.00$, $k = 45$), large ($g = 0.09$, $k = 46$), or very large ($g = 0.01$, $k = 41$) $F(3, 166) = 0.81$, $p = .48$.

Moderators Related to the Learning Event

Mode of Information Delivery The magnitude of the prequestion effect on repeated question performance did not significantly differ based on whether participants interacted with text ($g = 0.67$; $k = 155$) or audiovisual material ($g = 0.64$; $k = 26$) during the learning event, $F(1, 179) = 0.02$, $p = .88$. Similarly, the magnitude of the prequestion effect on new questions did not significantly differ based on whether participants interacted with text ($g = -0.005$; $k = 153$) or audiovisual material ($g = 0.10$; $k = 22$) during the learning event, $F(1, 173) = 0.73$, $p = .39$.

Duration of Learning The magnitude of the prequestion effect on repeated question performance did not significantly differ according to whether the duration of the learning event was short ($g = 0.67$; $k = 20$), medium ($g = 0.62$; $k = 27$), long ($g = 0.66$; $k = 15$), or extended ($g = 0.71$; k = 18) $F(3, 76) = 0.07$, $p =. 97$. Similarly, the magnitude of the prequestion effect on new question performance did not differ according to whether the duration of the learning event was short ($g = 0.18$; $k = 14$), medium ($g = 0.11$; $k = 23$), long ($g = 0.04$; $k = 10$), or extended ($g = 0.10$; $k = 20$), $F(3, 63) = 0.08$, $p =. 96$.

Amount of Text Material The magnitude of the prequestion effect on repeated question performance did not differ according to whether the word count of text material was considered small ($g = 0.57$, $k = 38$), medium ($g = 0.54$, $k = 28$), large ($g = 0.58$, $k = 36$), or very large ($g = 0.79$, $k = 35$), $F(3, 133) = .59$, $p = .62$. The magnitude of the prequestion effect on new question performance also did not differ according to whether the word count of text material was small ($g = -0.03$, $k = 23$), medium ($g = 0.05$, $k = 39$), large ($g = -0.02$, $k = 32$), or very large ($g = -0.03$, $k = 39$), $F(3, 129) = .14$, $p =. 92$.

Control Conditions The magnitude of the prequestion effect on repeated question performance did not significantly differ whether comparisons consisted of a standard

[28] The comparison level of "prequestion before and interspersed" (i.e. a level in which prequestions were all placed before and interspersed throughout the learning resource; specifically, before key paragraphs) was dropped due to only having a single comparison ($k = 1$) in the model of new questions and fewer than ten comparisons ($k = 2$) in the model of repeated questions.

($g=0.68$, $k=147$), active-related ($g=0.54$, $k=23$), or active-unrelated control group ($g=.75$, $k=11$); $F(2, 178)=0.41$, $p=.66$. Similarly, the magnitude of the prequestion effect on new question performance did not significantly differ whether comparisons consisted of a standard ($g=0.01$, $k=145$), active-related ($g=-0.23$, $k=18$), or active-unrelated control group ($g=.29$, $k=12$); $F(2, 172)=2.68$, $p=.07$.

Content Covered The magnitude of the prequestion effect on repeated question performance did not significantly differ whether the content covered had elements of STEM and non-STEM content ($g=.40$, $k=10$) compared to non-STEM expository-based ($g=0.66$; $k=87$), STEM-based ($g=0.72$; $k=55$) or non-STEM narrative-based ($g=0.62$; $k=30$), $F(3, 177)=0.61$, $p=.60$. Similarly, the magnitude of the prequestion effect on new question performance did not differ according to whether the content was STEM-based ($g=0.12$; $k=60$), non-STEM expository-based ($g=-0.05$; $k=80$), or non-STEM narrative-based ($g=-0.08$; $k=27$), $F(2, 166)=1.46$, $p=.23$.[29]

Moderators Related to Assessment Conditions

Post-Test Format The magnitude of the prequestion effect on repeated question performance did not significantly differ between post-test questions that were free recall ($g=0.73$, $k=33$), cued-recall ($g=0.58$, $k=45$), or recognition ($g=0.65$, $k=100$; $F(2, 175)=0.42$, $p=.65$. Similarly, the magnitude of the prequestion effect on new question performance did not significantly differ between post-test questions that were free recall ($g=-0.03$, $k=24$), cued-recall ($g=-0.006$, $k=40$), or recognition ($g=0.01$, $k=109$) $F(2, 170)=0.06$, $p=.93$.[30]

Retention Interval The magnitude of the prequestion effect on repeated question performance did not significantly differ when post-testing occurred immediately ($g=0.64$, $k=103$), was delayed after a distractor task ($g=0.58$, $k=45$) or was delayed by 1 day or more ($g=0.87$, $k=33$), $F(2, 178)=2.06$, $p=.13$. Similarly, the magnitude of the prequestion effect on new question performance did not significantly differ when post-testing occurred immediately ($g=-0.01$, $k=106$), was delayed after a distractor task ($g=0.11$, $k=41$), or was delayed by 1 day or more ($g=-0.06$, $k=28$), $F(2, 172)=1.21$, $p=.30$.

Moderators Related to the Participant Sample

Age of Tested Population The magnitude of the prequestion effect on repeated question performance did not significantly differ according to whether participants were

[29] Comparisons containing content that were categorized as having both STEM and Non-STEM elements were dropped from the new question sub-analysis ($k=9$) due to a low comparison count.

[30] The comparison level of "multiple retrieval" (i.e., a level in which multiple forms of retrieval existed on the post-tests) was dropped in our repeated question ($k=2$) and new question ($k=1$) model due to having fewer than 10 comparisons.

children ($g = 0.52$; $k = 19$), teenagers ($g = 0.51$; $k = 27$), or adults ($g = 0.71$; $k = 135$), $F(2, 178) = 0.69$, $p = .50$. Similarly, the magnitude of the prequestion effect on new question performance did not significantly differ between children ($g = 0.001$; $k = 24$), teenagers ($g = 0.08$; $k = 28$), and adults ($g = 0.002$; $k = 123$), $F(2, 172) = 0.14$, $p = .86$.

Prior Knowledge The magnitude of the prequestion effect on repeated question performance did not significantly differ if researchers controlled prior knowledge of the content learned ($g = 0.66$; $k = 46$) or did not control prior knowledge ($g = 0.80$; $k = 21$), $F(1, 65) = .36$, $p = . 54$. Similarly, the magnitude of the prequestion effect on new question performance did not significantly differ whether researchers controlled prior knowledge ($g = -0.00$; $k = 47$) or did not control prior knowledge ($g = 0.16$; $k = 17$), $F(1, 62) = .71$, $p = .40$.

Moderators Related to Experimental Setting

The magnitude of the prequestion effect on repeated question performance did not differ according to whether the experimental procedures were conducted in a laboratory setting ($g = 0.71$; $k = 87$) or a school setting ($g = 0.62$; $k = 93$), $F(1,178) = 0.36$, $p = .54$. Further, this result stayed consistent whether the studies conducted in schools were aligned with or related to class content ($g = .46$; $k = 18$) or not related to class content ($g = .68$; $k = 75$) $F(1, 91) = .67$, $p = .41$. Similarly, the magnitude of the prequestion effect on new question performance did not significantly differ according to whether the experimental procedures were conducted in a laboratory setting ($g = 0.08$; $k = 82$) or a school setting ($g = -0.06$; $k = 92$), $F(1, 172) = 2.09$, $p = .14$. Further, this result stayed consistent whether the studies conducted in schools were aligned with or related to class content ($g = -0.03$; $k = 13$) or not related to class content ($g = -0.07$; $k = 79$) $F(1, 90) = .06$, $p = .80$.

Discussion

Research over the past sixty years has investigated the effects of prequestions on learning across many situations that vary in the learning and test materials, procedures, populations, and comparison conditions. This variation has resulted in a broad array of studies that have shown learning benefits in some situations and for some measures but not for others. Given the variation of approaches and outcomes, we conducted a meta-analysis to help answer the following two questions: 1) Do prequestions promote learning specific to prequestioned content, or does this learning benefit extend to non-prequestioned content as well? 2) Under what specific conditions are prequestions more or less effective? In the following sections, we discuss the answers to these questions in light of our results, the implications for educational practice, the limitations of the current work, and directions for future research.

Research Question 1: Overall Effects of Prequestions

We found that prequestions facilitated learning information that is specific to the prequestions (repeated question performance, g=.66), but not other information from the learning resource (new question performance, g=.01). Although there was not an overall benefit of prequestions on new question performance, there was also no cost relative to a control condition. These results align with those of St. Hilaire and colleagues (2024) showing significant benefits of prequestions on repeated questions but not on new questions.

Taken alone, the benefit of prequestions on repeated questions could be consistent with any or all four of the learning mechanisms reviewed in the introduction including attention, elaboration, curiosity, and metacognition. Contrastingly, the null effect of prequestions on new question performance is inconsistent with explanations based on curiosity and metacognition at the general topic level. Additionally, the null effect on new questions suggests that if the mechanism underlying the prequestion effect is orienting attention to the prequestioned information, such orienting does not appear to impede the learning of nonprequestioned information.

The benefit of prequestions for repeated but not new questions is consistent with the hypothesis that the encoding of the initial prequestions is very specific to the particular question(s) and not at the more general level (e.g., at the level of the topic area). This specific encoding then could lead to learners attending to that particular content in the learning resource, elaborating upon that content when encountered, closing a specific metacognitive gap, and/or increasing curiosity to answer a specific question. This interpretation is consistent with theories of cognition that propose that new information is often encoded with a preference for the specific (Medin & Ross, 2014; Nosofsky, 1991). However, it is inconsistent with accounts that place the locus of the prequestion effect in general content learning mechanisms, such as generally increased attention, elaboration, metacognition, or curiosity.

Research Question 2: Under what Conditions are Prequestions More or Less Effective?

There was little observed heterogeneity among effect sizes extracted from the same study. However, both our repeated and new question meta-analytic models found that significant variation existed among extracted effect sizes across study clusters, meaning that it was likely that some moderators could be influencing the magnitude of the prequestion effect on repeated and new questions. Therefore, we conducted sub-analyses to investigate our potential moderators.

Despite this heterogeneity, we found that the effect of prequestions on repeated questions was robust across variations in the learning event, participant sample, and assessment conditions. Only one variable, presentation of prequestions, significantly moderated the magnitude of the prequestion effect on repeated questions. Across the respective levels of all other potential moderators (i.e., mode of information delivery, content covered, duration of learning, amount of text material, age of tested population, prior knowledge, retention interval, control condition, prequestion placement, post-test format, type of prequestion, number of prequestions, and experimental setting) the magnitude of the prequestion effect did not differ and the benefit of prequestions on repeated questions was observed.

However, the same cannot be stated for the effect of prequestions on new questions. We found little evidence for a general effect of prequestions—as measured by new question performance—with similar null results across levels of potential moderators. For the effect of prequestion on new questions, the only significant moderator was the type of prequestion; all other potential moderators were found to be nonsignificant, though the moderator of control conditions was marginal (p=.07). While we discuss the variation across prequestion type below, these interpretations should be contextualized by the fact that the effect was relatively small (g< .35).

Moderators of the Prequestion Effect on Repeated Questions

Presentation of Prequestions. Compared to prequestions alone, prequestions were most effective when given with feedback (g=1.25, p< .01), marginally more effective when given with a supplemental activity (e.g., note-taking, group discussion, learning objectives; g=.84, p= .07), and no different when viewing but not answering prequestions (g=.65, p= .71). This finding may not be surprising, as feedback is well documented to significantly improve the encoding of correct information (Cai et al., 2023; Wisniewski et al., 2020). However, we pause to make straightforward interpretations about the effect of prequestions plus feedback in this instance. In the typical prequestion design, learners do not receive immediate feedback because if given, feedback takes away the ability to make interpretable results between the effects of prequestions or feedback. For instance, if there is a positive learning benefit observed, it may not be clear if the prequestions promoted learning or if learners just memorized the given answer from the feedback. Further, giving feedback may encourage learners to simply remember the correct answer rather than engage with the to-be-learned content. In this instance, rather than learners being oriented to engage with learning material, either to search for the solution or to confirm their answer, they again simply memorize the given solution.

Additionally, the marginal difference of prequestions plus a supplemental activity suggests that receiving two learning aids that work congruently could potentially engage multiple cognitive processes more than one alone. For instance, prequestions alongside notetaking may create a stronger approach to learning information from subsequent material in that prequestions would assist students in identifying what information they should be searching for while note-taking would allow them to systematically record and organize new information.

Interestingly, we did *not* find that the prequestion minus condition produced a smaller effect compared to prequestions alone; indeed, the prequestion effect was numerically larger in the prequestion minus condition than the prequestion alone condition. This is despite the fact that, compared to a prequestions alone condition, learners randomly assigned to a prequestion minus condition do not interact with prequestion in the same manner in that learners only receive and are made aware of prequestions but are instructed not to answer them.

This finding helps clarify the mechanism that underlies the prequestion effect. If prequestions primarily direct participants' attention to the specific information in those questions, then we would not expect differences between prequestions alone and prequestions minus because both conditions include the same question content. From this perspective, whether or not the participant answered the questions would not impact the later search of that content in the learning resource. Alternatively, if prequestions serve as a metacognitive check, then we would expect that attempting to generate an answer to a prequestion is an important step in helping a learner become aware of their knowledge gaps. Therefore, we would expect the prequestion minus condition to be minimally effective because participants were not required to give an explicit response to a question. Nevertheless, it is possible that participants in the prequestion minus conditions may have covertly answered the questions even though they were not explicitly asked for a response, thereby showing similar effects to the participants who did explicitly answer the questions in the prequestions alone condition. Research has shown that covert retrieval of information can enhance learning (Yu et al., 2025).

One possible way to differentiate between these accounts would be to compare performance between those who receive prequestions and those who receive learning objectives. Although both of these devices could serve to direct attention, learning objectives in the form of a statement would not allow learners to generate answers, therefore eliminating the possibility of covertly serving as a metacognitive check. Sana et al. (2020, Experiment 2) performed this very comparison: participants were randomly assigned to receive either learning objectives (e.g., "In the first passage, you will learn about where the mirror neurons are located."), multiple-choice prequestions converted from those objectives (e.g., "Where are the mirror neurons located?"), or a study-fact control (e.g., "In the first passage, you will learn that mirror neurons are located in the ventral premotor cortex."). Participants in the prequestion condition performed significantly better on a final test compared to those in the learning objective or study-fact control conditions, who performed similarly. The result of Sana et al.'s (Experiment 2) study provides support for the metacognitive-based mechanism in that it suggests the act of answering prequestions does contribute to increased later performance.

In sum, our results appear to best align with expectations for an attention-based explanation of the prequestion effect; however, evidence from situations where participants are unable to covertly answer questions also suggests a possible metacognitive mechanism. To further explore the roles of these mechanisms, we suggest that future work could include direct measures of metacognition (e.g., a judgment of learning prompt) and/or attention measures (e.g., using eye-tracking to observe where participants are spending more or less time within the learning resource or specific passages containing either prequestioned or nonprequestioned information).

Moderators of the General Effect on New Questions

Type of Prequestion If prequestions serve to focus a learner's attention on subsequent prequestioned information, then we would expect that factually-based prequestions may contribute to a stronger prequestion benefit on repeated questions than on new questions, in that factual prequestions would orient learners to search for information explicitly stated in the learning resource. Alternatively, conceptually-based prequestions, which require a learner to make inferences from information in the learning resource, may be less likely to benefit performance on repeated questions because the information is not directly stated and thus harder to find in the learning resource. Instead, conceptually-based prequestions may be more likely to benefit performance on new questions as they direct learners to encode information at a broader level, guiding them to search for coherence and understanding to effectively generate inferences from the learning material.

We found that the magnitude of the prequestion effect on new questions was larger if learners received either factual ($g = 0.002$) or conceptual prequestions ($g = 0.34$) compared to learners who received a mix of both question types ($g = -0.40$). Several hypotheses could explain this finding. It could be that receiving a mix of both question types may be confusing to learners as they may not know which type of information to focus on. Alternatively, learners may simply focus on factual questions because searching for correct solutions in the learning material may be easier than inferring them. Lastly, it might simply be that when receiving both types of questions, learners perform well on factual questions but not so well on conceptual questions.

Additionally, we found that the magnitude of the prequestion effect on new questions was significantly larger if learners received conceptual prequestions compared to factual prequestions. Though we suspected that conceptual prequestions would not benefit new question performance because discovering answers from the learning resource would be harder, it appears that this generative process supported new question performance. Perhaps requiring participants to generate an answer from the learning material pushes an individual to encode information more generally rather than searching for a specific bit of information. To test these potential hypotheses, we suggest that future work could directly compare performance between participants who receive factual, conceptual, or a mix of both question types.

Control Conditions We found that the magnitude of the prequestion effect on new questions was marginally larger if those in the control condition received an activity that was unrelated to the learning resource (active-unrelated, $g = 0.29$) compared to participants in a standard control condition (standard control, $g = 0.01$) or participants who received an activity related to the learning resource (active-related, $g = -0.23$). On one hand this finding may not be too surprising, as we suspect that completing some other activity related to the learning resource (e.g., having more study time, being given learning objectives) would help participants learn and thereby reduce the outcome differences between prequestion vs. control groups. On the other hand, completing an activity that is unrelated to the learning resource (e.g., answering prequestions on a different topic) would serve to widen those outcome differences.

Discrepancies Between Outcomes

Our results diverge somewhat from those of St. Hilaire and colleagues (2024), who found significant moderating effects of prequestion placement and mode of information delivery on new question performance, and significant moderating effects of age of tested population and post-test question format on repeated question performance. Specifically, they found that the benefits of prequestions on new questions were stronger when prequestions were placed all at the beginning of a learning event instead of interspersed throughout, and when information was learned through audiovisual format instead of text format. Further, the benefit of prequestions on repeated questions was stronger for adult learners compared to children and when the post-test question format was short answer instead of multiple-choice.

We observed no significant effects of these moderators in the current analyses. Though the exact reasons for the discrepancies are unknown, there are several factors that could contribute to these outcome differences. One likely contributing factor could be the exact records included in the meta-analytic comparisons across the two studies. While the current analyses included 55 records compared to the 76 records included in the St. Hilaire et al. (2024) analyses, across the overall analyses (new, repeated) the current study included 84 additional comparisons. This would be due to our multi-level meta-analytic approach allowing for the inclusion of multiple comparisons from a single record containing multiple experimental conditions of interest within an experiment while maintaining effect size independence and accounting for variance in effect sizes both across and within studies, which further allowed for greater precision in calculated effect sizes. For instance, in the new question performance sub-analytic model for prequestion placement[31] which had direct mapping of levels across moderators (prequestions before learning = massed, prequestions interspersed = interleaved), the current analyses included 37 (prequestion before learning) and 45 (prequestions interspersed) additional comparisons within the moderator analysis.

Another contributing factor could be the differences between how we coded and defined these shared moderators. For instance, St. Hilaire et al. (2024) coded multiple-choice or short answer as the categories for their retention test format moderator. Within our post-test format moderator, we had initially similarly captured these formats as categories (i.e., multiple-choice and short answer) as well as more (e.g., cued-recall ($k = 16$), free-recall ($k = 17$), free response ($k = 2$), fill in the blank ($k = 8$), yes/no ($k = 4$), open-ended ($k = 1$), close-ended ($k = 1$), word anagram ($k = 3$)). However, so that we need not exclude categories for having low comparison counts ($k < 10$) we decided to categorize levels based on the broader type of test (i.e., recognition, free-recall, cued-recall) rather than specifically the question type within the test. Our broader categorization based on cognitive processes rather than specific question formats may have assessed different aspects of the prequestion effect than St. Hilaire et al.'s (2024) more narrowly defined categories.

[31] This would be the interleaved study moderator analysis within St. Hilaire et al. (2024) (see Table 9 for numerical comparisons).

Lastly, while some of our shared moderators had similarly defined categories, the current work included additional levels not shared across analyses. For instance, St. Hilaire et al.'s (2024) participant age moderator categorized levels as grade school children or adults, whereas our age of participant moderator categorized three age groups: children (k-6th), teens (7th-12th), or adults (college aged-beyond).[32]

Addressing Unique Moderators

There were unique moderators not shared across analyses. For instance, while St. Hilaire et al. (2024) reported the publication year, research design, match between pre-questions and retention questions, pre-study phase timing, study phase timing, reading speed, and relation between tested and non-tested material, we examined the content covered, duration of learning, amount of text material, and experimental setting. There were several reasons why the variables unique to the St. Hilaire et al. (2024) analyses were not pursued in the current work. First, our initial data collection captured the publication year, type of research design, and match between prequestions and repeated questions. However, as we had no theoretical reasoning as to how these moderators could influence the effect of prequestions on repeated or new question performance, they were ultimately dropped from subsequent moderator analyses.

Second, although we had unique moderators that accounted for time spent with prequestions or learning materials, both the current moderators and St. Hilaire et al. (2024) moderators appear to be aimed at trying to assess timing effects on different materials. That is, both analyses were focused on capturing similar issues, but we operationally defined these moderators somewhat differently.

Lastly, we did not include moderators that specifically assessed whether the duration of learning was experimenter-paced or self-paced (e.g., pre-study phase timing, study phase timing, reading speed) as it was not a feature that appeared when we were considering moderators.

In summary, we believe that it should not be a concern that results between analyses do not align perfectly. It is quite possible that some of the moderator analyses explored by the current meta-analysis and St. Hilaire et al. (2024) meta-analysis represent areas of the prequestion literature where the effects may be somewhat inconsistent. Having more data to contribute to the literature is beneficial as it helps us understand these effects more, and it highlights the areas within this literature that may be the most important to follow up on.

Implications for Practice

Our findings suggest that prequestions can be an effective learning tool for specific information related to the prequestion content and can be easily implemented with

[32] A subsequent sub-analysis aggregating comparisons categorized as children and teens into one level (repeated model $g = .51$, $k = 46$; new model $g = .04$, $k = 52$) found a similar non-significant result when compared to adults for the repeated, $F(1, 179) = 1.43$, $p = 0.23$, and new question model, $F(1, 173) = .12$, $p = 0.72$.

relatively low cost to instructors. Indeed, our meta-analysis showed that even a small dosage of prequestions, within STEM or non-STEM content areas, over short or longer periods of learning, can result in learning benefits on information specific to the prequestions. This suggests that, across diverse educational contexts, educators could integrate prequestions into lesson plans to promote more effective learning experiences. Moreover, if prequestions are paired alongside a supplemental activity or aid, such as note-taking or learning goals, these benefits increase even more.

Although we found no benefits of prequestions for learning nonprequestioned information from the learning resource, there was also no detriment of prequestions either. This result is relevant to practice in that it assuages the concern that prequestions would lead learners to focus only on the prequestioned information at the expense of the nonprequestioned information.

The absence of a prequestion effect on nonprequestioned information is consistent with the literature on the transfer of learning, in that near transfer (i.e., "transfer between very similar contexts") is more likely to occur than far transfer (i.e., "transfer between contexts that, on appearance, seem remote and alien to one another"; Perkins & Salomon, 1992). One could view the relationship between prequestions and repeated questions as a case of near transfer (answering questions with identical structure at different times) and performance on new questions as a case of far transfer (learners applying encoded knowledge to answer less directly related questions).

A possible direction that future researchers could explore is how to facilitate the learning of nonprequestioned information through methods known to promote far transfer in other circumstances (e.g., presenting learners with contrasting cases or requiring them to engage in self-explanation of their answers; see Alfieri et al., 2013; Richey & Nokes-Malach, 2015 for reviews). Future studies could present prequestions that ask learners to compare and contrast or could have learners elaborate their answers to prequestions to generate more initial abstract encoding. Such work may establish ways to yield a more general effect from prequestions; however, for now, it must suffice to say that prequestions are an educational tool that boosts the learning of prequestioned information without being a detriment to nonprequestioned information.

Limitations

Despite the overall and sub-analytic findings, our study had limitations that warrant consideration in interpreting our results. First, publication bias may have influenced our findings for our repeated question model. Regarding the model of performance on repeated questions, we conducted three tests of publication bias, two of which (PET-PEESE, funnel plot) revealed that bias was likely to exist while the third (publication status) was non-significant but numerically consistent with the other tests. These results suggest that publication bias is likely to exist in our model of repeated question performance.

However, the adjusted overall true effect size ($g = .68$) displays only a small difference between our repeated model's computed effect size ($g = .66$), and indeed the adjusted effect size was *larger*. Despite different approaches to modeling the data, our PEESE and three-level model converged on similar estimates of the overall true effect size. This could imply that our three-level model structure is a good fit for our data and is sensitive enough to capture the overall true effect size despite publication bias. Further, we found no evidence of publication bias in our model of performance on new questions.

Second, we did not analyze the learning resource to ascertain the relation between the different types of questions and how the answers are represented in the learning resource. For instance, Little and Bjork (2016) designed incorrect answer choices to multiple-choice prequestions to be the correct answer choices to new questions. Here, one could argue that the effect of prequestions on new questions could be moderated by the relatively close, far, or non-existent relation between prequestioned and nonprequestioned material. In the current work, we were unable to make clean assessments of this variable due to the insufficient number of studies that specifically report this level of relatedness. However, St. Hilaire et al. (2024) assessed the relation between prequestioned and nonprequestioned material and found that it moderated the effects of prequestions on new question performance with a relatively small number of comparisons ($k = 15$; $p = .02$), suggesting that it may be a viable factor to explore more fully in future studies and meta-analyses.

Third, we did not analyze the learning resource to assess the influence of information location on the relationship between prequestions and learning. Recently, Sana and Carpenter (2023) found that the placement of information within the learning resource can differentially influence the effect of prequestions on new questions. When prequestioned information was located at the beginning of the learning resource, with nonprequestioned information following, prequestions benefited only repeated questions and not new questions. However, when prequestioned information was located at the end of the learning resource, with nonprequestioned information located at the beginning, prequestions significantly benefited both repeated and new questions. These recent findings support the role of attention in the prequestion effect and suggest that learning resources can be constructed in a way to draw attention to both prequestioned and nonprequestioned information to maximize overall learning.

Conclusion

Our multi-level meta-analysis supported a specific benefit of prequestions on repeated information; however, we did not find evidence of a general benefit of prequestions on encoding new information that was not targeted by the prequestions. This suggests that the initial encoding of the prequestions is specific to the prequestions; there is little evidence for abstraction. Further, we found that the effect of prequestions on repeated questions was quite robust in that only one examined variable (presentation of prequestions) moderated the relationship between prequestions and repeated question performance. Similarly, we found only one variable (type of prequestion) that moderated the effect of prequestions on new question performance. Though much work has been conducted to assess the effects of prequestions, we advocate for research efforts to now advance our understanding of *when* and *how* prequestions could be used to promote the learning of nonprequestioned information.

Additionally, we considered how significant moderators may speak to the theorized underlying cognitive mechanisms of the prequestion effect on repeated and new question performance. However, due to the robustness of the prequestion effect on repeated question performance and the null effect on new question performance, we observed few specific moderators that could definitively elucidate or provide evidence for a single driving mechanism or knowledge representation. Thus, we urge future work to focus on further testing these underlying mechanism(s) to determine which ones are contributing to the particular learning effects of prequestions.

Lastly, our findings underscore the pedagogical value of prequestions as a simple yet powerful tool to promote the learning of prequestioned information across a diverse range of educational contexts. Teachers can readily use and adapt prequestions to help promote student learning of a variety of topics.

Appendix

Correlations & Outcome Discrepancies

We conducted correlation analyses of our moderating factors to investigate whether our factors were related to each other. Several factors were found to have significant associations. Tables 7 and 8 display Spearman's rho and Cramer v statistics of these correlations respectively. Further, we created a short-hand table summarizing discrepancies between the present sub meta-analytic models and those created by St. Hilaire et al. (2024). Table 9 therefore displays where shared moderators diverged or converged across these analyses

Table 7 Ordinal Correlation Table of Moderating Factors

Variable	1	2	3	4
1. Age of Tested Population	—			
2. Amount of Text Material	.173*	—		
3. Duration of Learning	-.368*	.572*	—	
4. Number of Prequestions	.014	.537*	.687*	—

p <.05*

Table 8 Nominal Correlation Table of Moderating Factors

Variable	1	2	3	4	5	6	7	8	9	10
1. Experimental Setting$_a$	—									
2. Content Covered	.407*	—								
3. Prior Knowledge	0.52	.162	—							
4. Mode of Information Delivery	.129	.282*	.107	—						
5. Presentation of Prequestions	.189	.229*	.342*	.166	—					
6. Prequestion Placement	.449*	.300*	.502*	.252*	.262*	—				
7. Control Condition	.314*	.264*	.148	.087	.115	.158	—			
8. Type of Prequestion	.121	.109	.219	.021	.197*	.165	.079	—		
9. Retention Interval	.592*	.281*	.221	.185*	.202*	.244*	.198*	.205*	—	
10. Post-Test Format	.487*	.206*	.141	.114	.225*	.347*	.255*	.237*	.297*	—

p <.05*

Table 9 Shared Moderators Across Meta-analytic models

Present Moderators	St. Hilaire Moderators	St. Hilaire Levels	Present Levels	Levels Definition(M/NM)*	Outcome (M/NM)	New Level
Type of prequestion	Prequestion Knowledge Type	1) Factual, 2) Conceptual	1) Factual 2) Conceptual 3) Both	M	Repeated Question Performance (M) New Question Performance (NM)	Both
Presentation of prequestions	Prequestion Guess	1) Guess 2) Read-Only	1) Prequestions alone 2) Prequestions minus 3) Prequestions optional 4) Prequestions plus feedback 5) Prequestions plus supplemental activity	M	Repeated Question Performance (M) New Question Performance (M)	Prequestions Optional Prequestions plus feedback Prequestions plus supplemental activity
Prequestion Placement	Interleaved study	1) Massed 2) Interleaved	1) Prequestions before learning 2) Prequestions interspersed 3) Prequestions before and interspersed	M	Repeated Question Performance (M) New Question Performance (NM)	Prequestions before and interspersed
Number of Prequestions	Number of Prequestions	Continuous	1) Small 2) Medium 3) Large 4) Very Large	NM	Repeated Question Performance (M) New Question Performance (M)	Different variable types
Mode of information Delivery	Type of Learning Materials	1)Text 2) Audiovisual	1) Text 2) Audiovisual	M	Repeated Question Performance (M) New Question Performance (NM)	N/A

Table 9 (continued)

Present Moderators	St. Hilaire Moderators	St. Hilaire Levels	Present Levels	Levels Definition(M/NM)*	Outcome (M/NM)	New Level
Retention Interval	Retention interval – between study phase and retrieval phase	Continuous	1) Immediate 2) Delayed after distractor 3) Delayed 1-Day+	NM	Repeated Question Performance (M) New Question Performance (M)	Different variable types
Age of Tested Population	Participant Age	1) Adults 2) Grade school children	1) Children 2) Teens 3) Adults	M	Repeated Question Performance (NM) New Question Performance (M)	Teens
Control Conditions	Control Condition Task	1) Mock Pretest 2) Extended study time 3)Study-only	1) Standard 2) Active-related 3) Active-unrelated	NM	Repeated Question Performance (M) New Question Performance (M)	N/A
Post-Test Question Format	Retention Test Format	1) Multiple-choice 2) Short Answer	1) Free-Recall 2) Cued-Recall 3) Recognition 4) Multiple-Retrieval	NM	Repeated Question Performance (NM) New Question Performance (M)	Cued Recall Multiple Retrieval
Prior Knowledge	Exclusion of accurate prequestion guesses	1) Included 2) Excluded 3) Undefined	1) Controlled 2) Uncontrolled	NM	Repeated Question Performance (M) New Question Performance (M)	N/A

**M*=Matched, *NM*=Not Matched; **Denotes the direct mapping of outcomes between repeated questions performance to repeated questions effect sub-analytic models and new question performance to new question effect sub-analytic models; New level denotes a category in the present study not being shared across analyses. Additionally, this column denotes that definitions for these new categories did not match existing categories across analyses

Acknowledgements We thank Keya Dixit and Emily Khuu for assistance coding some of the moderator variables as well as Dr. Benjamin Rottman for his detailed feedback on earlier iterations of this work.

Funding This research is based upon work supported by the James S. McDonnell Foundation 21 st Century Science Initiative in Understanding Human Cognition, Collaborative Grant No. 220020483.

Data availability All data for this project, along with relevant code, have been made publicly available on the Open Science Framework (OSF) and can be accessed at https://osf.io/cgbw8/?view_only=d1e145f5f8724362bb93c7ec9f5d5992

Declarations

Conflict of interest We have no conflicts of interest, relevant financial, or non-financial interests to disclose.

References

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

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

*Benya, R. A. (1980). The effect of the placement of guiding questions for listening passages on the retention of factual material by third quarter college Spanish students (Order No. 8100117). Available from ProQuest Dissertations & Theses Global. (303056025). https://www.proquest.com/dissertations-theses/effect-placement-guiding-questions-listening/docview/303056025/se-2. Accessed July 2021.

*Boker, J. R. (1974). Immediate and delayed retention effects of interspersing questions in written instructional passages. *Journal of Educational Psychology, 66*(1), 96–98. https://doi.org/10.1037/h0036072

Bostan, E., & Ozcelik, E. (2024). The effect of integrative prequestions on learning from text: An eye-tracking study. *The Journal of Experimental Education*, 1–13.

*Bull, S. G., & Dizney, H. F. (1973). Epistemic curiosity-arousing prequestions: Their effect on long-term retention. *Journal Of Educational Psychology, 65*(1), 45–49. https://doi.org/10.1037/h0034817

Cai, Z., Gui, Y., Mao, P., Wang, Z., Hao, X., Fan, X., & Tai, R. H. (2023). The effect of feedback on academic achievement in technology-rich learning environments (TREs): A meta-analytic review. *Educational Research Review, 39*, Article 100521. https://doi.org/10.1016/j.edurev.2023.100521

Carpenter, S. K., King-Shepard, Q., & Nokes-Malach, T. (2023). The prequestion effect: Why it is useful to ask students questions before they learn. *In their own words: What scholars want you to know about why and how to apply the science of learning in your academic setting*, 74–82. Available from: https://www.researchgate.net/profile/Logan-Fiorella/publication/369588588_Learning_by_Teaching/links/64234a8ca1b72772e431adec/Learnin-by-Teaching.pdf#page=81

*Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied, 24*(1), 34–42. https://doi.org/10.1037/xap0000145

*Carpenter, S. K., & Toftness, A. R. (2017). The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition, 6*(1), 104–109. https://doi.org/10.1016/j.jarmac.2016.07.014

*Chimezie, J. A. (1970). The effect of predictability on the value of post-questions in learning from written discourse (Order No. 7113537). Available from ProQuest Dissertations & Theses Global. (302527671). https://www.proquest.com/dissertations-theses/effect-predictability-on-value-post-questions/docview/302527671/se-2. Accessed July 2021.

*Chobot, M. C. (1984). The interactive effects of field dependence and adjunct questions on learning from prose by library/information science Students (trait treatment, mathemagenic effect, regression analysis, psychological differentiation, field orientation) (Order No. 8416040). Available from ProQuest Dissertations & Theses Global. (303288000). https://www.proquest.com/dissertations-theses/interactive-effects-field-dependence-adjunct/docview/303288000/se-2. Accessed July 2021.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587
Craik, F. I. M., Routh, D. A., & Broadbent, D. E. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 302*(1110), 341–359.
*Davidson, H. M. (1978). The use of pre-questions in the teaching of comprehension (Order No. 27919405). Available from ProQuest Dissertations & Theses Global. (2372044107). https://www.proquest.com/dissertations-theses/use-pre-questions-teaching-comprehension/docview/2372044107/se-2. Accessed July 2021.
*De Lima, NKd., & Jaeger, A. (2020). The effects of prequestions versus postquestions on memory retention in children. *Journal of Applied Research in Memory and Cognition, 9*(4), 555–563. https://doi.org/10.1016/j.jarmac.2020.08.005
*Dickerson, C. T. (1987). The effects of adjunct questions on learning from text inconsistent with prior knowledge (Order No. 8804837). Available from ProQuest Dissertations & Theses Global. (303587896). https://www.proquest.com/dissertations-theses/effects-adjunct-questions-on-learning-text/docview/303587896/se-2. Accessed July 2021.
Duell, O. K. (1974). Effect of type of objective, level of test question, and the judged importance of tested materials upon posttest performance. *Journal of Educational Psychology, 66*(2), 225.
Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Sage Publications Inc.
Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction, 49*, 13–20. https://doi.org/10.1016/j.learninstruc.2016.11.010
*Feil, P. H. (1977). A quasi-experimental analysis of the effects of adjunct questions on retention (Order No. 7727028). Available from ProQuest Dissertations & Theses Global. (302816658). https://www.proquest.com/dissertations-theses/quasi-experimental-analysis-effects-adjunct/docview/302816658/se-2. Accessed July 2021.
Felker, D. B., & Dapra, R. A. (1975). Effects of question type and question placement on problem-solving ability from prose material. *Journal Of Educational Psychology, 67*(3), 380–384.
Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. https://doi.org/10.48550/arXiv.1503.02220
Frase, L. T. (1968). Effect of question location, pacing, and mode upon retention of prose material. *Journal Of Educational Psychology, 59*(4), 244–249.
Frase, L. T., Patrick, E., & Schumer, H. (1970). Effect of question position and frequency upon learning from text under different levels of incentive. *Journal of Educational Psychology, 61*(1), 52–56. https://doi.org/10.1037/h0028823
Geller, J., Carpenter, S. K., Lamm, M. H., Rahman, S., Armstrong, P. I., & Coffman, C. R. (2017). Prequestions do not enhance the benefits of retrieval in a STEM classroom. *Cognitive Research: Principles and Implications, 2*(1), 1–13. https://doi.org/10.1186/s41235-017-0078-z
*Goldberg, E. S. (1980). The effects of three question properties on prose memory and forward text processing (Order No. 8103932). Available from ProQuest Dissertations & Theses Global. (303016641). https://www.proquest.com/dissertations-theses/effects-three-question-properties-on-prose-memory/docview/303016641/se-2. Accessed July 2021.
*Haimowitz, C. (1972). Effect of adjunct questions on learning from prose at the junior college level by reading or listening with rate controlled Available from APA PsycInfo. (615861413; 1973–05416–001). https://www.proquest.com/dissertations-theses/effect-adjunct-questions-on-learning-prose-at/docview/615861413/se-2. Accessed July 2021.
Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review Of Educational Research, 56*(2), 212–242. https://doi.org/10.3102/00346543056002212
*Hausman, H., & Rhodes, M. G. (2018). When pretesting fails to enhance learning concepts from reading texts. *Journal of Experimental Psychology: Applied, 24*(3), 331–346. https://doi.org/10.1037/xap0000160
Hidi, S., Renninger, K., & Krapp, A. (2004). Interest, A motivational variable that combines affective and cognitive functioning. In D. Y. Dai, & R. J. Sternberg (Eds.), *Motivation, emotion, and*

cognition: Integrative perspectives on intellectual functioning and development (pp. 89–115, 455 Pages). Lawrence Erlbaum Associates Publishers. https://doi.org/10.4324/9781410610515

*Hillman, S. B. (1972). The effect of question type and position on four types of learning among mentally handicapped children (Order No. 7302714). Available from ProQuest Dissertations & Theses Global. (302705419). https://www.proquest.com/dissertations-theses/effect-question-type-position-on-four-types/docview/302705419/se-. Accessed July 2021.

*Hollen, T. T., JR. (1970). Interaction of individual abilities with the presence and position of adjunct questions in learning from prose materials (Order No. 7111554). Available from ProQuest Dissertations & Theses Global. (302447369). https://www.proquest.com/dissertations-theses/interaction-individual-abilities-with-presence/docview/302447369/se-2. Accessed July 2021.

Hoyer, W. J., Cerella, J., & Onyper, S. V. (2003). Item learning in cognitive skill training: Effects of item difficulty. *Memory & Cognition, 31*, 1260–1270. https://doi.org/10.3758/BF03195809

*James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested? *Journal of Experimental Psychology: Applied, 25*(4), 576–587. https://doi.org/10.1037/xap0000231

*Janelli, M. (2019). Effects of pre-tests and feedback on performance outcomes in massive open online courses: What works and what doesn't? (Order No. 13895238). Available from ProQuest Dissertations & Theses Global. (2290797938). https://www.proquest.com/dissertations-theses/effects-pre-tests-feedback-on-performance/docview/2290797938/se-2. Accessed July 2021.

Jenkins, L., & Hoyer, W. J. (2000). Instance-based automaticity and aging: Acquisition, reacquisition, and long-term retention. *Psychology And Aging, 15*(3), 551–565. https://doi.org/10.1037/0882-7974.15.3.551

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

*Karjala, Y. (1984). Objectives, questions, and memory for instruction (social Work Education) (Order No. 8415471). Available from ProQuest Dissertations & Theses Global. (303343225). https://www.proquest.com/dissertations-theses/objectives-questions-memory-instruction-social/docview/303343225/se-2. Accessed July 2021.

*Kealy, W. A., Bakriwala, D. J., & Sheridan, P. B. (2003). When tactics collide: Counter effects between an adjunct map and prequestions. *Educational Technology Research And Development, 51*(2), 17–39. https://doi.org/10.1007/BF02504524

Kepes, S., Wang, W., & Cortina, J. M. (2023). Assessing publication bias: A 7-step user's guide with best-practice recommendations. *Journal of Business and Psychology, 38*(5), 957–982. https://doi.org/10.1007/s10869-022-09840-0

*Khoynejad, G. (1980). The effects of behavioral objectives, prequestions, and a combination of both on intentional and incidental learning from written text by secondary school students (Order No. 8100921). Available from ProQuest Dissertations & Theses Global. (303073359). https://www.proquest.com/dissertations-theses/effects-behavioral-objectives-prequestions/docview/303073359/se-2. Accessed July 2021.

Kim, E. S., & Willson, V. L. (2010). Evaluating pretest effects in pre–post studies. *Educational and Psychological Measurement, 70*(5), 744–759. https://doi.org/10.1177/0013164410366687

*Kirschner, P., & Van den Brink, H. (1979). The effect of adjunct questions on learning from a videotape lesson. (ERIC Document Reproduction Service No. ED 206264)

Kuhn, D., & Dean, D. (2004). A bridge between cognitive psychology and educational practice. *Theory Into Practice, 43*(4), 268–273. https://doi.org/10.1207/s15430421tip4304_4

*Lackman, T. W. (1970). Effects upon comprehension and reading rate under four prequestion conditions (Order No. 7106439). Available from ProQuest Dissertations & Theses Global. (302518877). https://www.proquest.com/dissertations-theses/effects-upon-comprehension-reading-rate-under/docview/302518877/se-2. Accessed July 2021.

Lai, E. R. (2011). Metacognition: A literature review. *Always Learning: Pearson Research Report, 24*, 1–40.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

*Lenz, B. K., Adams, G. L., Bulgren, J. A., Pouliot, N., & Laraux, M. (2007). Effects of curriculum maps and guiding questions on the test performance of adolescents with learning disabilities. *Learning Disability Quarterly, 30*(4), 235–244.

Lewis, M. R., & Mensink, M. C. (2012). Prereading questions and online text comprehension. *Discourse Processes, 49*(5), 367–390. https://doi.org/10.1080/0163853X.2012.662801

*Little, J. L. (2011). Optimizing multiple-choice tests as learning events (Order No. 3493389). Available from ProQuest Dissertations & Theses Global. (919052703). https://www.proquest.com/dissertations-theses/optimizing-multiple-choice-tests-as-learning/docview/919052703/se-2. Accessed July 2021.

Little, J., & Bjork, E. (2011). Pretesting with multiple-choice questions facilitates learning. In Proceedings of the annual meeting of the cognitive science society (Vol. 33, No. 33).

*Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44*(7), 1085–1101. https://doi.org/10.3758/s13421-016-0621-z

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75.

Lyday, N. L. (1983). A meta-analysis of the adjunct question literature. *Unpublished Dissertation.* The Pennsylvania State University. Available from ProQuest Dissertations & Theses Global. (303172082). https://www.proquest.com/dissertations-theses/meta-analysis-adjunct-question-literature/docview/303172082/se-2. Accessed July 2021.

*Manchester, N. E. (1984). The effects of conceptual level and position of adjunct questions on recall measures of reading comprehension in beginning college french (Order No. 8418971). Available from ProQuest Dissertations & Theses Global. (303321991). https://www.proquest.com/dissertations-theses/effects-conceptual-level-position-adjunct/docview/303321991/se-2. Accessed July 2021.

*McEeady, B. J. (1982). The effects of adjunct questions on prose comprehension in high vocabulary/low comprehension high school students (Order No. 8305937). Available from ProQuest Dissertations & Theses Global. (303267773). https://www.proquest.com/dissertations-theses/effects-adjunct-questions-on-prose-comprehension/docview/303267773/se-2. Accessed July 2021.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103*, 399–414. https://doi.org/10.1037/a0021782

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *Journal Of Educational Psychology, 88*(3), 508–519. https://doi.org/10.1037/0022-0663.88.3.508

Medin, D. L., & Ross, B. H. (2014). The specific character of abstract thought: Categorization, problem solving, and induction. In *Advances in the psychology of human intelligence* (pp. 189–223). Psychology Press.

*Memory, D. M. (1981). An extended comparison of adjunct aids with low-average and good readers in the sixth grade. *The Journal of Educational Research, 74*(6), 405–410. https://doi.org/10.1080/00220671.1981.10885340

*Memory, D. M. (1983). Main idea prequestions as adjunct aids with good and low-average middle grade readers. *Journal of Reading Behavior, 15*(2), 37–48.

Miyagi, P. S. (1995). The effect of level and location of adjunct questions (Order No. 9603918). Available from ProQuest Dissertations & Theses Global. (304235846). https://www.proquest.com/dissertations-theses/effect-level-location-adjunct-questions/docview/304235846/se-2. Accessed July 2021.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 3–27. https://doi.org/10.1037/0096-1523.17.1.3

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hofmann, T., Mulrow, C., Shamseer, L., Tetzlaf, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*. https://doi.org/10.1136/bmj.n71

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and pretesting effects: A review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review, 35*(4), 40. https://doi.org/10.1007/s10648-023-09814-5

*Pan, S. C., & Sana, F. (2021). Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice. *Journal of Experimental Psychology: Applied, 27*(2), 237–257. https://doi.org/10.1037/xap0000345

*Pan, S. C., Schmitt, A. G., Bjork, E. L., & Sana, F. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition, 9*(4), 542–554. https://doi.org/10.1016/j.jarmac.2020.07.004
*Park, I. (1993). The effects of orienting questions and prior knowledge on learning in hypertext (Order No. 9410166). Available from ProQuest Dissertations & Theses Global. (304035489). https://www.proquest.com/dissertations-theses/effects-orienting-questions-prior-knowledge-on/docview/304035489/se-2. Accessed July 2021.
*Patrick, E. M. (1976). The effects of variations in the uncertainty of prequestions and information feedback on the short and long term retention of prose materials Available from APA PsycInfo. (616180279; 1977–31013–001). https://www.proquest.com/dissertations-theses/effects-variations-uncertainty-prequestions/docview/616180279/se-2. Accessed July 2021.
*Peeck, J. (1970). Effect of prequestions on delayed retention of prose material. *Journal of Educational Psychology, 61*, 241–246. https://doi.org/10.1037/h0029104
Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International Encyclopedia of Education, 2*, 6452–6457.
*Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology, 15*(1), 27–35. https://doi.org/10.1016/0361-476X(90)90003-J
*Rauscher, W. C. (1978). An Investigation into the facilitative effects of two kinds of adjunct questions on the comprehension of students at two grade levels during the reading of textual materials (Order No. 7902934). Available from ProQuest Dissertations & Theses Global. (302882642). https://www.proquest.com/dissertations-theses/investigation-into-facilitative-effects-two-kinds/docview/302882642/se-2. Accessed July 2021.
Richey, J. E., & Nokes-Malach, T. J. (2015). Comparing four instructional techniques for promoting robust knowledge. *Educational Psychology Review, 27*(1), 181–218. https://doi.org/10.1007/s10648-014-9268-0
*Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*(3), 243–257. https://doi.org/10.1037/a0016496
*Richmond, M. G. (1973). The Relationship of the uniqueness of prose passages to the effect of question placement and question relevance on the acquisition and retention of information (Order No. 7402698). Available from ProQuest Dissertations & Theses Global. (302653002). https://www.proquest.com/dissertations-theses/relationship-uniqueness-prose-passages-effect/docview/302653002/se-2. Accessed July 2021.
*Rickards, J. P. (1972). The effects of frequency, position, conceptual level and type of orienting direction on recall of prose material (Order No. 7314040). Available from ProQuest Dissertations & Theses Global. (302684213). https://www.proquest.com/dissertations-theses/effects-frequency-position-conceptual-level-type/docview/302684213/se-2. Accessed July 2021.
Rickards, J. P. (1976). A new look at the position effect. *Journal of Reading Behavior, 8*(1), 37–45.
*Rickards, J. P., Anderson, M. C., & McCormick, C. B. (1976). Processing effects of common-word and number questions inserted in reading materials. *The Journal of Educational Research, 69*(7), 274–277. https://doi.org/10.1080/00220671.1976.10884897
Rogers, W. A., Hertzog, C., & Fisk, A. D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 359–394. https://doi.org/10.1037/0278-7393.26.2.359
Rothkopf, E. Z., & Billington, M. J. (1979). Goal-guided learning from text: Inferring a descriptive processing model from inspection times and eye movements. *Journal Of Educational Psychology, 71*(3), 310–327. https://doi.org/10.1037/0022-0663.71.3.310
*Rowls, M. D. (1975). The facilitative and interactive effects of adjunct questions on retention of eighth graders across three prose passages Available from APA PsycInfo. (616167773; 1977–28323–001). https://www.proquest.com/dissertations-theses/facilitative-interactive-effects-adjunct/docview/616167773/se-2. Accessed July 2021.
*Ruff, A. C. (1975). The effect of embedded factual questions in different positions on the recall of relevant and incidental information in three subject matter areas (Order No. 7516989). Available from ProQuest Dissertations & Theses Global. (302741095). https://www.proquest.com/dissertations-theses/effect-embedded-factual-questions-different/docview/302741095/se-2. Accessed July 2021.

Sagaria, S. D., & Di Vesta, F. J. (1978). Learner expectations induced by adjunct questions and the retrieval of intentional and incidental information. *Journal of Educational Psychology, 70*(3), 280–288. https://doi.org/10.1037/0022-0663.70.3.280

Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review, 30*(5), 1908–1916.

*Sana, F., Forrin, N. D., Sharma, M., Dubljevic, T., Ho, P., Jalil, E., & Kim, J. A. (2020). Optimizing the efficacy of learning objectives through pretests. *CBE—Life Sciences Education*. https://doi.org/10.1187/cbe.19-11-0257

*Sana, F., Yan, V. X., Clark, C. M., Bjork, E. L., & Bjork, R. A. (2021). Improving conceptual learning via pretests. *Journal of Experimental Psychology: Applied, 27*(2), 228–236. https://doi.org/10.1037/xap0000322

*Shanahan, T. (1986). Predictions and the limiting effects of pre-questions. In J. A. Niles & R. V. Lalik (Eds.), Solving problems in literacy: Learners, teachers, and researchers (pp. 92–98). Rochester, NY: The National Reading Conference. https://www.proquest.com/scholarly-journals/predictions-limiting-effects-prequestions/docview/617399378/se-2. Accessed July 2021.

*Sinnott, L. T., & Alderman, D. (1977). The effects of pre- and postquestions on learning from textual material in a cai format (Order No. 0512642). Available from ProQuest Dissertations & Theses Global. (302860487). https://www.proquest.com/dissertations-theses/effects-pre-postquestions-on-learning-textual/docview/302860487/se-2. Accessed July 2021.

Simkins, S., & Allen, S. (2000). Pretesting students to improve teaching and learning. *International Advances in Economic Research, 6*(1), 100–112.

*Skiba, D. T. (1975). Interactive effects of adjunct questions, field orientation, and intelligence on the learning of audio taped material by adult learners (Order No. 7611001). Available from ProQuest Dissertations & Theses Global. (302801433). https://www.proquest.com/dissertations-theses/interactive-effects-adjunct-questions-field/docview/302801433/se-2. Accessed July 2021.

*Smith, D. G. (1976). Position of inserted questions and ability in learning from prose (Order No. 7702012). Available from ProQuest Dissertations & Theses Global. (302823758). https://www.proquest.com/dissertations-theses/position-inserted-questions-ability-learning/docview/302823758/se-2. Accessed July 2021.

*St. Hilaire, K. J., & Carpenter, S. K. (2020). Prequestions enhance learning, but only when they are remembered. *Journal of Experimental Psychology: Applied, 26*(4), 705–716. https://doi.org/10.1037/xap0000296

*St. Hilaire, K. J., Carpenter, S. K., & Jennings, J. M. (2019). Using prequestions to enhance learning from reading passages: The roles of question type and structure building ability. *Memory, 27*(9), 1204–1213. https://doi.org/10.1080/09658211.2019.1641209

*St. Hilaire, K. J. (2017). The pretesting effect: How question-type and structure building ability impact learning (Order No. 10618238). Available from ProQuest Dissertations & Theses Global. (1954045888). https://www.proquest.com/dissertations-theses/pretesting-effect-how-question-type-structure/docview/1954045888/se-2. Accessed July 2021.

St. Hilaire, K. J., Chan, J. C., & Ahn, D. (2024). Guessing as a learning intervention: A meta-analytic review of the prequestion effect. *Psychonomic Bulletin & Review*, *31*(2), 411–441. https://doi.org/10.3758/s13423-023-02353-8

Szafran, R. F. (1986). What do introductory sociology students know and when do they know it? The results of pretesting students. *Teaching Sociology*, 217–223.

*Taylor, J. C. (1996). Attentional and non-attentional effects of prequestioning (Order No. NN18508). Available from ProQuest Dissertations & Theses Global. (193716025). https://www.proquest.com/dissertations-theses/attentional-non-effects-prequestioning/docview/193716025/se-2. Accessed July 2021.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multi-trial learning. *Psychonomic Bulletin & Review, 6*(4), 662–667. https://doi.org/10.3758/BF03212976

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal Of Educational Psychology, 95*(1), 66–73. https://doi.org/10.1037/0022-0663.95.1.66

*Toftness, A. R., Carpenter, S. K., Lauber, S., & Mickes, L. (2018). The limited effects of prequestions on learning from authentic lecture videos. *Journal of Applied Research in Memory and Cognition, 7*(3), 370–378. https://doi.org/10.1016/j.jarmac.2018.06.003

Touron, D. R., Hoyer, W. J., & Cerella, J. (2001). Cognitive skill acquisition and transfer in younger and older adults. *Psychology and Aging, 16*(4), 555–563. https://doi.org/10.1037/0882-7974.16.4.555

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age–cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122*(3), 231–249. https://doi.org/10.1037/0033-2909.122.3.231

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. https://doi.org/10.18637/jss.v036.i03

*Wiseman, E. D. (1982). A Study to determine the effects of both adjunct prequestions and adjunct postquestions on achievement (Order No. 8213737). Available from ProQuest Dissertations & Theses Global. (303055946). https://www.proquest.com/dissertations-theses/study-determine-effects-both-adjunct-prequestions/docview/303055946/se-2. Accessed July 2021.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10*, Article 487662. https://doi.org/10.3389/fpsyg.2019.03087

Yu, Y., Zhao, W., Li, A., Shanks, D. R., Hu, X., Luo, L., & Yang, C. (2025). Is covert retrieval an effective learning strategy? Is it as effective as overt retrieval? Answers from a meta-analytic review. *Educational Psychology Review, 37*(2), 52.

Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. *Journal Of Educational Psychology, 107*(4), 954–970. https://doi.org/10.1037/edu0000022