

Are you sure? Examining the potential benefits of truth-checking as a learning activity

Quarterly Journal of Experimental Psychology
1–15
© Experimental Psychology Society 2023



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/17470218231206813
qjep.sagepub.com



Karen Arcos , Hannah Hausman and Benjamin C Storm 

Abstract

Learners may be uncertain about whether encountered information is true. Uncertainty may encourage people to critically assess information's accuracy, serving as a kind of desirable difficulty that benefits learning. Uncertainty may also have negative effects, however, leading people to mistrust true information or to later misremember false information as true. In three experiments, participants read history statements. In one condition, all statements were true, and the participants knew it. In the other two conditions, some statements were true, and others were false. Participants were either told the statements' accuracy or they guessed the statements' accuracy prior to feedback, a manipulation we refer to as truth-checking. All participants were then tested on recalling the true information and on recognising true versus false statements. We observed a significant benefit of truth-checking in one of the three experiments, suggesting that truth-checking may have some potential to enhance learning, perhaps by inducing people to encode to-be-learned information more deeply than they would otherwise. Even so, the benefit may come at a cost—truth-checking took significantly longer than study alone, and it led to a greater likelihood of thinking false information was true, suggesting costs of truth-checking may tend to outweigh benefits.

Keywords

Memory; learning; uncertainty; truth-checking

Received: 13 March 2023; revised: 18 August 2023; accepted: 15 September 2023

How one studies is often more important for learning than how much one studies. For example, reading a chapter of a textbook once and engaging in elaborative interrogation—that is, periodically pausing to ask oneself, “Why might that be?”—can improve comprehension compared with reading the same chapter twice (Smith et al., 2010). Furthermore, some strategies can produce more learning than others with the same amount of studying. For example, practice recalling information from a text can improve memory for the text more than spending the same amount of time rereading it (Roediger & Karpicke, 2006). Research has identified several study strategies like these that improve long-term learning. Some of the promising strategies that have been shown to improve learning across a range of materials include elaborative interrogation, self-explanation, spacing, and retrieval practice (for a review, see Dunlosky et al., 2013).

These study strategies can be considered as types of desirable difficulties because they are beneficial for learning but feel more challenging to students than their less-effective counterparts (E. L. Bjork & Bjork, 2011; R. A.

Bjork & Bjork, 2020). When participants practised vocabulary flashcards spaced across multiple days or massed in a single day, spacing led to better performance on the final test for 90% of participants, yet 72% of participants believed that they learned more from massing (Kornell, 2009). This example reflects the typical finding that although a sense of difficulty or challenge often accompanies durable learning, students tend to mistake effort as a negative sign and opt to use strategies that feel effortless but that are also less effective (e.g., Kirk-Johnson et al., 2019).

Study strategies that constitute desirable difficulties often require grappling with confusion or uncertainty

Division of Social Sciences, Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, USA

Corresponding author:

Karen Arcos, Division of Social Sciences, Department of Psychology, University of California, Santa Cruz, Social Sciences 2, Room 277, Santa Cruz, CA 95064, USA.

Email: karcos1@uci.edu

Passive Reading (All True Condition)

- ✓ Germans are not indigenous to Germany but migrated there from Scandinavian countries.
- ✓ Abraham Lincoln issued the Emancipation Proclamation because of its military necessity.
- ✓ Elizabeth Cady Stanton and Susan B. Anthony supported the 13th Amendment to the U.S. Constitution abolishing slavery and collected signatures for its passage.
- ✓ Ulysses S. Grant, as president, sent Federal troops into the South to fight the Ku Klux Klan.

Truth-Checking Condition

Five statements are shown below. Four statements are true; one statement is false. Please read the statements carefully, and select the four statements that you think are true.

- Julius Caesar was stabbed 23 times on the Ides of May (May 15th), 144 BCE.
- Germans are not indigenous to Germany but migrated there from Scandinavian countries.
- Abraham Lincoln issued the Emancipation Proclamation because of its military necessity.
- Elizabeth Cady Stanton and Susan B. Anthony supported the 13th Amendment to the U.S. Constitution abolishing slavery and collected signatures for its passage.
- Ulysses S. Grant, as president, sent Federal troops into the South to fight the Ku Klux Klan.

Below is the list of statements you just read. Notice which are true.

- x False: Julius Caesar was stabbed 23 times on the Ides of May (May 15th), 144 BCE.
- ✓ True: Germans are not indigenous to Germany but migrated there from Scandinavian countries.
- ✓ True: Abraham Lincoln issued the Emancipation Proclamation because of its military necessity.
- ✓ True: Elizabeth Cady Stanton and Susan B. Anthony supported the 13th Amendment to the U.S. Constitution abolishing slavery and collected signatures for its passage.
- ✓ True: Ulysses S. Grant, as president, sent Federal troops into the South to fight the Ku Klux Klan.

Passive Reading Including Misinformation (Accuracy Given Condition)

- x False: Julius Caesar was stabbed 23 times on the Ides of May (May 15th), 144 BCE.
- ✓ True: Germans are not indigenous to Germany but migrated there from Scandinavian countries.
- ✓ True: Abraham Lincoln issued the Emancipation Proclamation because of its military necessity.
- ✓ True: Elizabeth Cady Stanton and Susan B. Anthony supported the 13th Amendment to the U.S. Constitution abolishing slavery and collected signatures for its passage.
- ✓ True: Ulysses S. Grant, as president, sent Federal troops into the South to fight the Ku Klux Klan.

Figure 1. Sample study phase block in Experiment 1.

Note. In the all-true condition, participants passively read all true statements (top). In the accuracy-given condition, participants passively read four true and one false statement; each statement's accuracy was labelled (bottom). In the truth-checking condition, participants were shown the same five statements as in the accuracy-given condition, but had to guess which was the one false statement and which were the remaining true statements (middle; first box). After guessing, participants received corrective feedback about each statement's accuracy (middle; second box). Screen reader compatible version can be found in the Online Supplementary Material B.

while learning (Overoye & Storm, 2015). For example, explaining concepts in a reading can improve comprehension, even if participants lacking prior knowledge about the topic are confused. Another type of uncertainty or confusion that learners often experience is not knowing whether something they hear or read is true. Students may wonder if what their classmate claimed in a discussion was true or if their study partner used the right formula for the math problem.

We examined whether being uncertain about the accuracy of to-be-learned information could be harnessed to

promote learning. Specifically, we tested whether a new activity to learn information, which we refer to as truth-checking, could serve as another kind of desirable difficulty. In our study, participants who truth-checked read true and false statements about history. They then indicated which statements they believed were true and received feedback about the veracity of each statement. We predicted that truth-checking before receiving feedback (Figure 1, middle) would improve subsequent memory for true information compared with two less demanding baseline conditions: passively reading only true information

(Figure 1, top), and reading accurate and inaccurate statements labelled as such (Figure 1, bottom). The difference between truth-checking and reading accurate and inaccurate information is that participants who truth-checked received feedback on statement accuracy after guessing accuracy for each, whereas participants who read each statement and its accuracy were already aware of each statement's accuracy.

A growing body of research has examined how exposure to false statements can generally increase later memory for misinformation (for a review, see Brashier et al., 2020). In contrast, the goal of the current research was to examine how exposure to false information affects the learning of true information. We presented false information to induce uncertainty during learning, and we examined the potential benefits of this uncertainty for learning true information.

Potential benefits of truth-checking

Prior research has found that truth-checking can reduce acquiring misinformation (Brashier et al., 2020; Marsh & Fazio, 2006; Rapp et al., 2014; Umanath et al., 2012). The present research was designed to test whether truth-checking would be beneficial for learning new true information. Drawing conclusions from prior research is difficult because the past studies used true statements that participants generally already knew prior to their exposure in the experiment (Brashier et al., 2020; Rapp et al., 2014), or failed to include a condition that presented only accurate information (Marsh & Fazio, 2006).

In the current study, we predicted that truth-checking would enhance learning of true information compared with reading only true information for several reasons. Truth-checking involves initial uncertainty in information's accuracy and actively guessing, or predicting, what is true. Reading only true information, in contrast, involves passively encoding information with certainty that it is accurate. Despite the negative connotations associated with uncertainty and guessing, previous research suggests that they can benefit learning in numerous, related ways, which we may sometimes collectively refer to as deep processing (Craik & Lockhart, 1972).

First, truth-checking may increase attention to the to-be-learned true statements, thus improving learning. Truth-checking requires participants to consistently respond, making it more active than reading only true information. As a learning activity progresses, attention to the task tends to decline, and off-task mind wandering can increase. Periodically answering questions has been shown to help people sustain their attention throughout a task and improve learning as a result (Pan et al., 2020; Peterson & Wissman, 2020; Szpunar et al., 2013). Therefore, attempting to determine statements' truth may be a simple way to

enhance engagement during learning compared with passively receiving true information that is known to be true.

Truth-checking may also induce certain affective states, such as curiosity, surprise, and confusion, increasing attention and therefore learning. Previous research has shown that guessing a question's answer tends to increase curiosity (Brod & Breitwieser, 2019) or surprise (Brod et al., 2018) compared with only reading the correct answer. For example, Brod and colleagues (2018) presented pairs of countries (e.g., Denmark and Germany) and either immediately indicated which country had a larger population or required participants to guess before presenting the answer. When participants guessed first, the answer surprised them more, and higher levels of surprise were associated with better memory for the correct answer on a subsequent test. Similarly, Stare and colleagues (2018) found that participants learned the answers to trivia questions better when they were more curious. The benefits of curiosity even extended to incidental learning of unrelated information that was presented in the experiment around the same time as the trivia answers, highlighting the importance of curiosity as an affective state for new learning. Elevated surprise and curiosity are typically associated with better learning, in part because they increase attention to the correct answer once it is presented (Brod, 2021; Brod & Breitwieser, 2019; Brod et al., 2018; Butterfield & Metcalfe, 2006; Fazio & Marsh, 2009; Kang et al., 2009). Guessing may lead participants to become more curious about the to-be-learned material possibly by revealing gaps in their knowledge (Gruber & Ranganath, 2019).

Similarly, not knowing statements' accuracy while truth-checking may cause confusion. Although students may want learning to feel clear from the start, research suggests that people tend to learn more from activities during which they report more confusion, provided the confusion is eventually resolved (Craig et al., 2004; D'Mello et al., 2014; D'Mello & Graesser, 2011; Muller et al., 2008; for a review, see Overoye & Storm, 2015). For example, D'Mello and colleagues (2014) exposed participants to either consistent, true claims or conflicting true and false claims as they learned about research methods in an online tutoring environment. Participants self-reported higher levels of confusion when they encountered conflicting information than when they encountered consistent, true information. Results suggested the confusion resulting from initial exposure to conflicting information was associated with better learning outcomes. The authors posited that to improve learning, confusion may increase attention and deeper processing, such as more thoroughly explaining the content to oneself. By explaining content using elaboration, participants may create more cues to facilitate future recall, such as by tying the material to their everyday lives or connecting it to prior knowledge. Thus, truth-checking may serve as a desirable difficulty because

it causes uncertainty or confusion, influences students' affective states, improves attention, and encourages deep processing of true information.

Finally, we may expect truth-checking to impact memory directly, hence improving learning. Previous research has repeatedly shown that guessing a question's answer can enhance learning, even when guessing occurs before the information is taught. Kornell (2014), for example, showed that participants are better able to remember the answers to trivia questions when they are asked to initially guess the answers before receiving feedback than when given the same amount of time to study the questions and answers together. This benefit of guessing incorrectly prior to learning has been demonstrated across a range of materials and conditions (for reviews, see Mera et al., 2022; Metcalfe, 2017). Participants in Richland et al. (2009) read a text passage and were then tested on the material from that passage. Participants who attempted to answer questions about the material prior to reading the passage performed better on the final test than participants who were given additional time to study the passage, an effect that was observed even when the initial attempts were unsuccessful.

Retrieving information has been shown to make memory more labile and easily updated with new information (e.g., Nadel et al., 2012). Even if unsuccessful, guessing is thought to activate related information, which can become associated with the correct answer once it is presented. On a later test, the correct answer can be recalled both directly from the question and indirectly from the question via the related information that was activated during the pretest. In contrast, when learners immediately read the correct answer without first guessing, memory searching does not occur. Prior knowledge may not be activated or will be activated to a lesser degree, and fewer retrieval routes will be created. Thus, pretesting, or guessing, may create more elaborated memories that incorporate prior knowledge and thus more retrieval routes to correct information than reading alone (Cyr & Anderson, 2015, 2018; Huelser & Metcalfe, 2012; Overoye et al., 2021). Truth-checking may affect memory in a similar way, leading participants to integrate the information they are learning more effectively within their existing knowledge structures.

Given the theoretical considerations outlined above, we predicted that truth-checking would benefit memory for accurate information, because it involves activating prior knowledge to guess a statement's accuracy. If a participant has to guess the truthfulness of whether Germans originally migrated from Scandinavian countries, for example, they may activate prior knowledge, including a mental map of Europe. They may note that Germany borders Denmark and is also directly across the Baltic Sea from Sweden. The participant may think the statement seems plausible and therefore (accurately) judges it as true. On a later test, when asked where Germans originate from, the

participant would think back to the map, including Denmark and Sweden, and remember the answer was Scandinavia. In contrast, without needing to judge a statement's truthfulness, participants who read that Germans originated from Scandinavia are unlikely to search memory for prior knowledge or generate an explanation for the fact that could help them recall the test's answer.

In short, we reasoned that compared with reading true information, truth-checking would enhance learning of true information via different types of deep processing, including curiosity, surprise, attention, retrieving prior knowledge, explaining, and elaborating. If truth-checking enhances learning, it could be thought of as a new type of desirable difficulty, because participants would feel challenged to grapple with the uncertainty, guess, and inevitably make mistakes (Huelser & Metcalfe, 2012; Yang et al., 2017).

Potential negative effects of truth-checking

Despite truth-checking's predicted benefits on memory for true information, exposure to false information may simultaneously yield negative consequences. Based on the illusory truth effect, repeated exposure to false information can make it more believable (for a meta-analysis, see Dechêne et al., 2010), with subsequent processing feeling more familiar or easier (for a review, see Unkelbach et al., 2019). Being exposed to misinformation, even just once, can increase the likelihood of later judging that information as true (Hassan & Barber, 2021). For example, reading the false statement that tomatoes are originally from Ecuador increases the likelihood that participants will later judge the statement as true. The continued influence of false information on memory, beliefs, and reasoning persists, even if the information is labelled as incorrect (for a meta-analysis, see Chan et al., 2017).

Therefore, truth-checking could have the negative effect of increasing belief in false statements, even if learners are told which information is false. In fact, because the truth-checking activity increases deep processing, uncertainty may improve encoding of both the false and true statements. As a result, the false statements could become more available in memory, producing a greater feeling of familiarity or fluency on a later test and further increasing one's belief in the misinformation. Thus, the same cognitive processes that we predicted would cause truth-checking to benefit memory for the true information, such as deep processing, may also exacerbate the illusory truth effect for the false information (Unkelbach & Rom, 2017).

The current study

Across three similar experiments, the current study tested the effectiveness of a truth-checking activity for learning

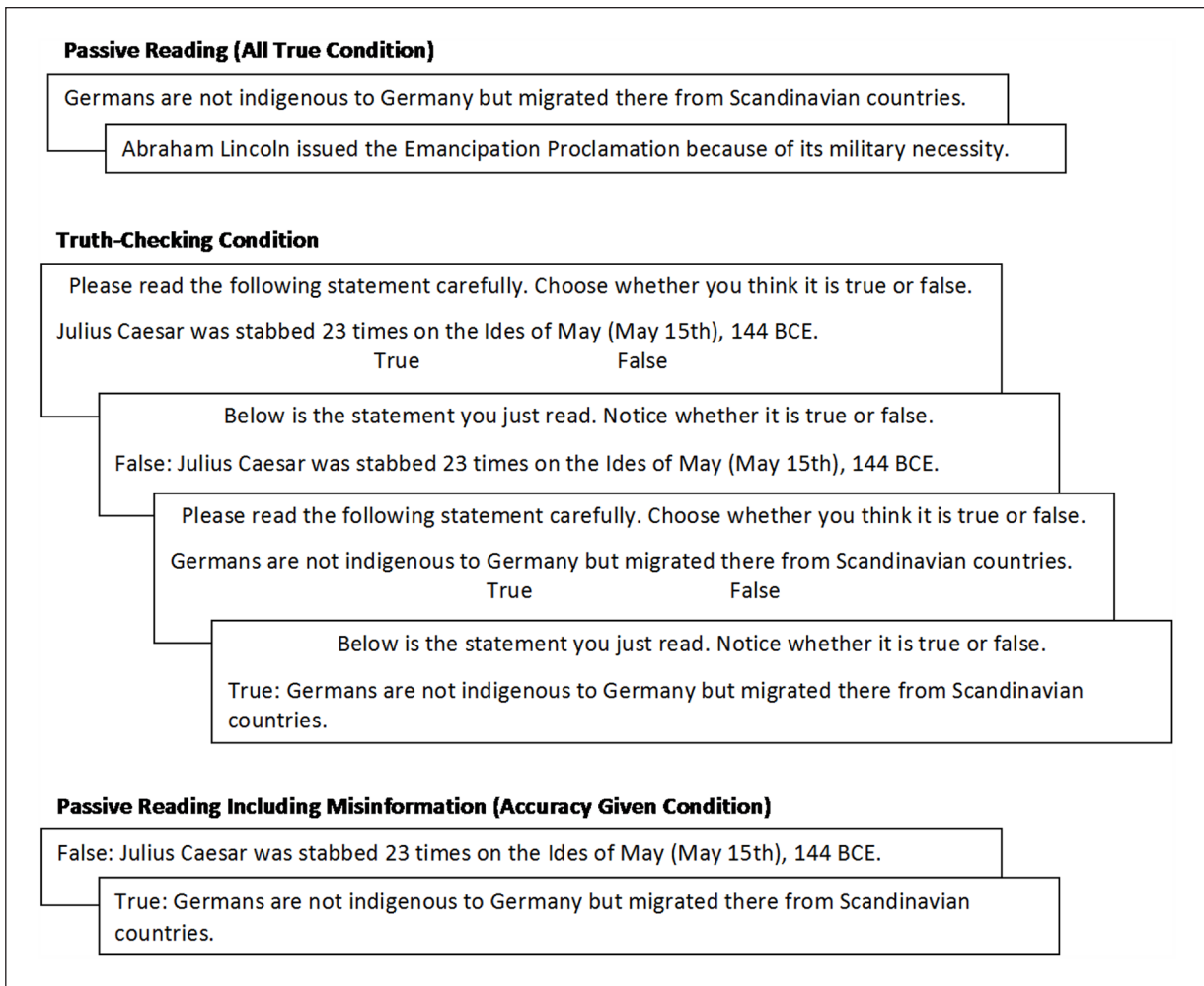


Figure 2. Two sample study phase trials in Experiments 2 and 3.
 Note. Screen reader compatible version in the Online Supplementary Material B.

accurate information. Participants learned historical facts in one of three conditions (see Figures 1 and 2 for the methods used in Experiment 1 and Experiments 2–3, respectively). Participants then completed a cued recall test in which the true facts were presented with a missing key word or phrase, which they were asked to remember. During learning, participants who engaged in the truth-checking activity (referred to as the truth-checking condition) were told they would read a mix of true and false historical statements. They guessed each statement’s accuracy, followed by receiving immediate feedback about whether each statement was true or false. The truth-checking activity was compared with a more typical learning activity: passively reading all true statements (referred to as the all-true condition). We hypothesised that the benefits of the truth-checking activity would not come from exposure to false information, *per se*. Instead, we reasoned that benefits of truth-checking for learning true information would come from the deep processing associated with initial uncertainty about statements’ accuracy and being

required to actively guess which statements were true. Therefore, we included a third condition (referred to as the accuracy-given condition) in which participants passively read the same true and false statements as participants in the truth-checking condition, except that statements were labelled as true or false when presented. The accuracy-given condition controlled for exposure to the false statements while not requiring participants to engage in any active truth-checking. Based on the findings and ideas presented in the introduction, we predicted that participants in the truth-checking condition would learn the true information best, performing better on the final cued recall test than participants in both the accuracy-given condition and the all-true condition. False statements were not tested to ensure that all participants were tested on the same statements, to which all participants were exposed. We predicted no difference in cued recall performance between the accuracy-given and all-true conditions because, in both conditions, participants passively read the true statements with the certainty that they were true.

Subsequently, we tested for potential side effects of being exposed to false information in the truth-checking activity. After completing the cued recall test on true statements, participants were administered a two-choice judgement task. Specifically, participants were presented with a true and false version of the false statements to which they were previously exposed. Participants indicated which version of the statements they believed to be true. Participants in the accuracy-given and truth-checking conditions read the false statements during learning. Participants in the all-true condition did not read these false statements previously. Each false statement was presented alongside the true version of the same statement. The statements' true versions were new to all participants. The judgement task therefore assessed participants' abilities to reject inaccurate information they were previously told was false. If exposure to false information produces an illusory truth effect and makes previously read statements feel familiar, participants in the all-true condition should outperform participants in the truth-checking and accuracy-given conditions, who may tend to misjudge the previously read false statements as true. By exposing participants in the truth-checking condition to four true statements and one false statement in Experiment 1, we hoped to maximise the benefits of truth-checking for learning the true statements while minimising the potential costs of truth-checking due to exposure to false information.

Experiment 1

Method

Participants. A total of 243 adults completed the first experiment. Participants were recruited from the participant pool in the Psychology Department at the University of California, Santa Cruz (UCSC). Participants could be from any academic major other than history. The majority ($n=237$) were compensated for participating with course credit; the remaining participants ($n=6$) completed the study voluntarily. Eleven participants were excluded from analysis for failing to answer a single cued recall test question (final $n=232$). Of the 11 excluded participants, 2 were from the all-true condition, 4 were from the accuracy-given condition, and 5 were from the truth-checking condition. More demographic information may be found in the Online Supplementary Material A.

Design. The experiment consisted of three between-subject conditions: all-true, accuracy-given, and truth-checking. An a priori power analysis using G*Power indicated that 240 participants (80 per condition) would be needed to have 80% power to detect a moderately sized effect of $d=0.45$ with $\alpha=.05$ for the independent samples t -test comparing final test performance between the truth-checking and accuracy-given conditions (Faul et al., 2007). This was our primary comparison of interest, and the sample

size was determined prior to beginning data collection. Our final sample size of 232 participants was also sufficient to detect a moderately sized difference in the final test performance among the three conditions (Cohen's $f=.21$, $d=0.42$) using the omnibus test of a one-way ANOVA with 80% power and $\alpha=.05$.

Materials. Fifty statements about history were used (and are available on Open Science Framework https://osf.io/d3qs4/?view_only=a284af15f78d431b8e0441173eef6fd). A university history instructor with extensive experience teaching undergraduates generated the statements. Specifically, we asked for a list of 50 statements (associated with 10 general categories) that would seem intuitively true but that would not be common knowledge among undergraduates. One statement from each of the categories was then modified to create a false version. The statements ranged from 5 to 32 words in length ($M=18.6$, $SD=6.5$), focusing on various information categories associated with world history: people, products, processes, theories, wars, monuments, museums, sources and methods, perspectives, and rumours. An example of a true statement is "Germans are not indigenous to Germany but migrated there from Scandinavian countries." To falsify the 10 statements, information such as dates, war, country, revolution, and names were modified. An example of a false statement is "Tomatoes are originally from Ecuador." The position of the false information in the statements varied from the beginning to the end of the statements. To design the cued recall test, keywords such as dates, war names, countries, regional terms, and people's first or last names were removed from each of the 40 true statements.

Procedure. The UCSC Human Subjects IRB approved all study procedures for this and subsequent experiments (IRB #: HS-FY2022-9), and all participants consented prior to participating. The online study was completed via Google Forms. Using a between-subjects design, participants were randomly placed in one of three conditions: all-true ($n=76$), accuracy-given ($n=71$), and truth-checking ($n=85$). The last digit of participants' telephone numbers determined to which condition they were assigned.

The study consisted of a learning phase, a cued recall test, and a two-choice judgement task. Participants in all three conditions were told they would read blocks of 40–50 statements that they would be tested on after reading the final block. They were not told about the final truthfulness judgements test. The study needed to be completed in one session, and participants had no time limit for completing it. Reading/truth-checking times were not collected.

During the study phase, participants in the all-true condition were instructed to carefully read the statement blocks (no mention was made regarding the statements' accuracy). Ten blocks were presented, corresponding to

each of the topics. For each block, the four true statements about the topic were presented on the screen at the same time, meaning participants read 40 statements in total. The 10 blocks were presented one at a time. The blocks and the statements within each block were presented in a fixed pseudorandom order. Participants took as much time as they needed to read each block before moving on to the next. In the accuracy-given condition, participants read 50 statements, 5 per each of the 10 topics, 4 true and 1 false. The ratio of true to false information was selected to maximise exposure to true information and minimise exposure to false information, which was important given our interest in learning of true information. As in the all-true condition, statements were presented on the screen, one at a time, in the same fixed pseudorandom order. Each statement was also labelled as true or false.

As in the accuracy-given condition, participants in the truth-checking condition encountered 10 blocks of 5 statements per block. They were told that within each block, four statements were true, and one statement was false. Critically, participants were not initially told which statements were true or false. To increase likelihood that participants attended to all five statements, participants were asked to choose which four statements they thought were true, and not just to select which statement they thought was false. Participants then received immediate feedback in which all five statements were presented again, labelled as true or false as in the accuracy-given condition. Participants were not provided with the true version of the false statement during feedback.

The cued recall test immediately followed the study phase. Test questions were made from the 40 true statements with one to three critical words missing from each (e.g., “Germans are not indigenous to Germany but migrated there from ___ countries.”). Participants were not tested on false statements. Statements were shown simultaneously in random order. Participants were instructed to fill in the blank with statements’ missing word(s).

Finally, participants completed the two-choice judgement task. Ten pairs of statements were presented, one per topic category. Pairs of statements comprised true and false versions of the same fact. The false version in each statement pair was the false statement that participants in the accuracy-given and truth-checking conditions read during the study phase. Participants in the all-true condition had not read these false statements during the study phase. The true version in each statement pair was new to all participants. Thus, to answer correctly, participants in the accuracy-given and truth-checking conditions could reject the statement they were told was false in the study phase. Participants in the all-true condition would need to guess or rely on prior knowledge.¹

Data analysis. Two raters manually coded each response. Our interrater reliability on cued recall coding was

near-perfect (Cohen’s kappa = .99; Landis & Koch, 1977). If raters disagreed, a third rater was consulted to resolve any unclear responses. Responses were correct if spelling was off by at most one letter. If the correct response was “Scandinavian” and the participant responded “Scandinavian,” the answer was coded as correct. On the contrary, if the correct response was “Teotihuacan” and the participant responded “Tenotchtilan,” the answer was coded as incorrect. This relatively strict spelling criteria was used to allow some typos while ensuring reliable coding across questions and raters. Responses were considered correct if wording from the original statement was present. For example, if the answer was “Antikythera” and the participant responded “Antikythera mechanism,” the response was coded as correct because the word “Antikythera” was used. Had only “mechanism” been written, the response would have been coded as incorrect. For responses over one word long, all the words needed to be present for the response to be coded as correct. For example, if the correct answer was Emancipation Proclamation and the participant responded “Emancipation,” the response was coded as incorrect. Responses that consisted of human first and last names were considered correct if at least the last name was correct. For example, if the correct answer was “Elizabeth Cady Stanton” and the participant responded “Stanton,” the response was correct. Responses with the correct first name and an incorrect last name were coded as incorrect. For responses consisting of first names and Roman numerals, responses were considered correct if at least the correct first name was included regardless of the Roman numeral. Responses consisting of years were only correct if the exact year was used, for example, 1450.

Results and discussion

Truth-checking performance. During the study phase, participants in the truth-checking condition were instructed to select which four of the five presented statements were true for each of the 10 topics. Their accuracy suggests that participants were largely guessing, with performance levels being only slightly above chance ($M=71\%$, $SD=8\%$, 95% confidence interval [CI] = [69%, 72%]). Importantly, if a participant was incorrect about which statement was false on a given trial, they would necessarily respond incorrectly to two of the five statements (i.e., they would be incorrect on the true statement marked false and on the false statement marked true). Thus, assuming participants in the truth-checking condition guessed correctly at chance levels, 20% of the time, they would earn five out of five on correct trials and three out of five on incorrect trials, resulting in an overall chance performance level of 68%. Our observation that participants in the truth-checking condition performed only slightly above chance suggests that they experienced considerable uncertainty during the study phase.

Table 1. Proportion correct on the cued recall test (assessing the learning of true information) and the two-choice judgement task (assessing memory for the false information) in Experiments 1–3.

Experiment	Condition	Cued recall <i>M</i> (<i>SE</i>)	Two-choice judgement task <i>M</i> (<i>SE</i>)
Ex. 1	All-true	.48 (.03)	.66 (.02)
	Accuracy-given	.47 (.03)	.72 (.02)
	Truth-checking	.45 (.02)	.74 (.02)
Ex. 2	All-true	.42 (.02)	.67 (.02)
	Accuracy-given	.37 (.02)	.68 (.02)
	Truth-checking	.50 (.02)	.65 (.02)
Ex. 3	Accuracy-given	.36 (.02)	.68 (.02)
	Truth-checking	.40 (.02)	.59 (.02)

Final tests. Table 1 shows proportion correct on the cued recall test. A one-way ANOVA of cued recall with condition (all-true, accuracy-given, truth-checking) as the between-subjects factor revealed no main effect of condition on proportion correct on the cued recall test, $F(2, 229) = 0.28$, $p = .755$, $\eta_p^2 = .002$. This result suggests we failed to find evidence that the uncertainty manipulation acted as a desirable difficulty to improve encoding true information.

Table 1 also shows proportion correct on the two-choice judgement task for identifying true versions of false statements. A one-way ANOVA with condition (all-true, accuracy-given, truth-checking) as the between-subjects factor found a significant main effect of study condition on proportion correct on the two-choice judgement task, $F(2, 229) = 4.40$, $p = .013$, $\eta_p^2 = .04$. Independent samples *t*-tests revealed that participants in the accuracy-given condition were more accurate than participants in the all-true condition, $t(145) = 2.25$, $p = .026$, $d = 0.37$. Similarly, participants in the truth-checking condition were more accurate than participants in the all-true condition, $t(159) = 3.03$, $p = .003$, $d = 0.48$. No differences in judgement accuracy were found between the truth-checking and accuracy-given conditions, $t(154) = 0.52$, $p = .607$, $d = 0.08$.

Although we failed to find evidence of truth-checking improving the learning of correct information, we also failed to find evidence of it producing or exacerbating an illusory truth effect, causing participants to misremember false statements as true. Indeed, participants exposed to the false information during learning identified false information as false better than participants who were not exposed to the false information during learning. Two factors may help explain this finding. First, the illusory truth effect has typically been observed when participants are unaware of the accuracy of the information to which they are exposed (e.g., Brashier et al., 2020). In the current study, they received feedback regarding statements' accuracy. Second, participants would have likely benefitted from studying

the false information during learning, thereby allowing them to remember which statements were false to support their performance on the subsequent two-choice judgement task.

Experiment 2

We found no effect of learning condition on cued recall performance in Experiment 1. We hypothesised that presenting the four true statements and the one false statement simultaneously on the same screen may have reduced the benefit of truth-checking. For example, if a participant evaluated the first of the five statements as false, they could immediately select the remaining four statements as true without engaging in deep processing to evaluate each statement's accuracy. Said differently, the procedure may have succeeded in minimising the potential negative effects of exposure to false information, but it may have also diminished the potential benefits of truth-checking for learning true information. In Experiment 2, participants were shown statements one at a time, with participants in the truth-checking condition told that each statement might be false. We predicted that having participants judge each statement's accuracy individually would encourage participants to spend more time reading and engaging in processing each statement more deeply, leading to a benefit of truth-checking and thus better cued recall performance in the truth-checking condition than in the accuracy-given and all-true conditions.

Method

Participants. A total of 285 participants completed Experiment 2. To be eligible, participants could be from any academic major other than history. Most participants ($n = 258$) were compensated for participating with course credit and were recruited through the psychology department's subject pool; other participants from the university community completed the study voluntarily ($n = 10$). An additional 17 participants completed the study but did not indicate how they were recruited or whether they received course credit for their participation. Forty participants were removed from analyses for not attempting to answer a single cued recall test question (final $n = 245$). Demographic information may be found in the Online Supplementary Material A.

Materials, design, and procedure. The study was completed online via Qualtrics software (Qualtrics, Provo, UT, USA). In this between-subjects design, participants were randomly placed in one of three conditions: all-true ($n = 82$), accuracy-given ($n = 82$), and truth-checking ($n = 81$). An a priori power analysis using G*Power identical to experiment 1 indicated that 240 participants (80 per condition) would be needed to have 80% power to detect a

moderately sized effect of $d=0.45$ with $\alpha=.05$ for the independent samples t -test comparing final test performance between the truth-checking and accuracy-given conditions (Faul et al., 2007). This was our primary comparison of interest, and the sample size was determined prior to beginning data collection. Our final sample size of 245 participants was also sufficient to detect a moderately sized difference in final test performance among the three conditions (Cohen's $f=.20$, $d=0.40$) using the omnibus test of a one-way ANOVA with 80% power and $\alpha=.05$.

The materials and procedure were similar to those employed in Experiment 1, with a few notable exceptions. Most importantly, during the learning phase, participants read each statement individually (Figure 2), rather than the four to five statements per topic all at once. Participants in the all-true condition read 40 true statements, one statement per page, with statements presented in a fixed pseudorandom order and not grouped by topic. In the accuracy-given condition, participants read 50 statements, 5 per each of the 10 topics, 4 true, and 1 false. As in the all-true condition, the statements were presented on the screen, one at a time, in the same fixed pseudorandom order. Each statement was also labelled as true or false. In the truth-checking condition, the same 50 statements were presented on screen, one at a time, in the same order. Statements were not initially labelled as true or false. Participants were asked to judge the accuracy of each statement before receiving immediate feedback. Feedback was presented one at a time and displayed the same statement again, along with the label of true or false. Participants were not told the number of true and false statements they would read. In all three conditions, the statements were presented for unlimited time, and reading/truth-checking times were collected.

Immediately following the learning phase, all participants completed the same cued recall test and two-choice judgement task as in Experiment 1. Two independent raters manually coded each participant's cued recall responses as described in Experiment 1. Interrater reliability was near-perfect ($\kappa=.99$).

Results and discussion

Truth-checking performance. During the study phase, participants in the truth-checking condition judged whether each presented statement was true or false but were not told how many true and false statements they would read. Across all 50 items, participants correctly judged statements at a rate similar to that which was observed in Experiment 1 ($M=70%$, $SD=7%$, 95% CI = [68%, 71%]). Chance is somewhat difficult to calculate in this experiment because it is unclear when, or to what extent, participants became aware of the true/false ratio through feedback. However, signal detection theory analyses (Stanislaw & Todorov, 1999) suggested that participants

Table 2. Median total seconds per true statements in the study phase and learning efficiency scores in Experiments 2 and 3.

Experiment	Condition	Median time M (SE)	Learning efficiency M (SE)
Ex. 2	All-true	6.11 (0.63)	.11 (.02)
	Accuracy-given	4.66 (0.32)	.12 (.01)
	Truth-checking	8.52 (0.59)	.07 (.003)
Ex. 3	Accuracy-given	5.10 (0.32)	.10 (.01)
	Truth-checking	7.80 (0.32)	.05 (.003)

entered the experiment with little prior knowledge, with near-zero ability to discriminate true from false facts originally (d' : $M=0.13$, $SD=0.53$, 95% CI = [0.01, 0.24]). Our observation that participants were correct over 50% of the time appeared to be driven by the fact that participants showed a significant bias to judge a statement as true (c : $M=-89%$, $SD=37%$, 95% CI = [-97%, -81%]), likely due to becoming aware that substantially more statements were true than were false. In any case, the level of performance suggests that as in Experiment 1, participants in the truth-checking condition experienced uncertainty during the study phase and that final test performance primarily reflected new learning.

We then calculated each participant's median time for the 40 true statements during the study phase (Table 2). For participants in the truth-checking condition, time per statement was the sum of the time spent judging a statement's accuracy and then reading the feedback containing each statement and its true or false label. For participants in the all-true and accuracy-given conditions, time per statement was the total amount of time spent reading the statements during learning. A one-way ANOVA with condition (all-true, accuracy-given, truth-checking) as the between-subjects factor found study condition significantly affected how long participants studied true statements, $F(2, 242)=13.35$, $p<.001$, $\eta_p^2=.1$. Follow-up independent samples t -tests revealed that the time taken for participants in the truth-checking condition to truth-check each statement and then read its feedback was significantly longer than for participants to read the statements in the all-true condition, $t(161)=2.78$, $p=.006$, $d=0.44$ and the accuracy-given condition, $t(161)=5.76$, $p<.001$, $d=0.90$. Participants in the all-true condition also took significantly longer than those in the accuracy-given condition, $t(162)=2.05$, $p=.042$, $d=0.32$.

Final tests. Table 1 shows proportion correct on the cued recall test in each condition. Unlike Experiment 1, a one-way ANOVA with condition (all-true, accuracy-given, truth-checking) as the between-subjects factor revealed a significant main effect on proportion correct on the cued recall test, $F(2, 242)=9.07$, $p<.001$, $\eta_p^2=.07$. Follow-up independent samples t -tests revealed that participants in the truth-checking condition recalled significantly more

than those in the all-true condition, $t(161)=2.63$, $p=.009$, $d=0.41$, and the accuracy-given condition, $t(161)=4.34$, $p<.001$, $d=0.68$. A significant difference was not found between participants in the all-true and accuracy-given conditions, $t(162)=1.54$, $p=.125$, $d=0.24$. Consistent with our predictions, we observed a benefit of truth-checking, with the effect emerging in Experiment 2 (and not in Experiment 1) perhaps because encouraging participants to attend to each statement individually led to deeper processing of all statements.

Table 1 also shows proportion correct on the two-choice judgement task in each condition. Unlike Experiment 1, a one-way ANOVA with condition (all-true, accuracy-given, truth-checking) as the between-subjects factor found no effect of condition on proportion correct on the two-choice judgement test, $F(2, 242)=0.64$, $p=.529$, $\eta_p^2=.005$. Thus, the benefits of truth-checking on remembering true information were not associated with an illusory truth effect for the false information.

Learning efficiency. Truth-checking improved learning of true information but also required more study time. One way to contextualise the benefits of truth-checking is to numerically assess whether the learning benefits outweigh the time costs. Therefore, we calculated a learning efficiency score for each participant by dividing their proportion correct on the cued recall test by their median time per true statements, with higher scores indicating the participant learned more per second of studying (Table 2). To capture the totality of the time costs across the learning activity, we measured study time as participants' total amount of time spent per trial. For participants in the truth-checking condition, this included the time spent reading the statements, making a decision, and attending to feedback. A one-way ANOVA with condition (all-true, accuracy-given, truth-checking) as the between-subjects factor found that learning efficiency differed significantly as a function of condition, $F(2, 242)=5.74$, $p=.004$, $\eta_p^2=.05$. Follow-up independent samples t -tests revealed that participants in the truth-checking condition learned significantly less per second of studying than participants in the all-true condition, $t(161)=2.97$, $p=.003$, $d=0.47$, and the accuracy-given condition, $t(161)=3.71$, $p<.001$, $d=0.58$. A significant difference in learning efficiency was not found between participants in the all-true and accuracy-given conditions, $t(162)=0.35$, $p=.724$, $d=0.06$.

In sum, in Experiment 2, truth-checking improved memory for true information and did not lead to an enhanced belief in false information. The only drawback to truth-checking was time because evaluating each statement and reading the feedback for each statement took significantly longer than just reading the statements. Although a medium-sized benefit of truth-checking emerged, the efficiency analyses revealed that uncertainty produced the least learning per second of studying.

Experiment 3

Given that the pattern of results changed from Experiment 1 to Experiment 2 after making the procedural change to the study phase, we sought to replicate Experiment 2. Experiment 3 was nearly identical to Experiment 2 except that the cued recall test and two-choice judgement task took place after a brief delay, rather than immediately after the study phase. We also removed the all-true condition to prioritise data collection in the two conditions most important for concluding about the effects of truth-checking. If the benefits of truth-checking can be attributed to being uncertain and actively guessing whether each statement was true, and not to exposure to false information per se, then the accuracy-given condition should provide a better comparison condition than the all-true condition.

Methods

A total of 228 participants completed the experiment for course credit, none of whom had completed either Experiment 1 or 2. Eligibility and recruitment were the same as in the previous experiment. A total of 63 participants were excluded due to either a software malfunction or not attempting to answer any of the cued recall test questions (final $n=165$). Demographic information about participants may be found in the Online Supplementary Material A. In this between-subjects design, participants were randomly assigned to the accuracy-given condition ($n=83$) or the truth-checking condition ($n=82$). Similar to Experiment 2, an a priori power analysis using G*Power indicated that 160 participants (80 per condition) would be needed to have 80% power to detect a moderately sized effect of $d=0.45$ with $\alpha=.05$ for the independent samples t -test comparing final test performance between the truth-checking and accuracy-given conditions (Faul et al., 2007). Sample size was determined prior to beginning data collection. The procedure was identical to Experiment 2, except that participants watched a 12-min video about flooding as a distractor between the study phase and the cued recall test.

Results and discussion

Study phase. Across all 50 items, participants in the truth-checking condition correctly judged statements as true or false at a rate similar to that of the prior experiments ($M=68\%$, $SD=7\%$, 95% CI = [67%, 70%]). A signal detection theory analysis suggested that participants entered the experiment without prior knowledge, unable to discriminate true from false facts originally (d' : $M=0.05$, $SD=0.62$, 95% CI = [-0.09, 0.18]). As in Experiment 2, participants were correct over 50% of the time possibly because they realised statements were more likely to be true than false, and thus developed a bias to judge statements as true (c : $M=-0.85$, $SD=0.35$, 95% CI = [-0.93,

−0.77]). Based on these results, we can assume that participants in the truth-checking condition experienced uncertainty during the study phase and that performance on the final tests primarily reflected new learning.

Table 2 shows participants' average median time per statement in each learning condition. We replicated the finding that participants in the truth-checking condition took significantly longer per statement than those in the accuracy-given condition, $t(163)=5.97, p<.001, d=0.93$.

Final tests. Interrater reliability for scoring cued recall test responses was near perfect ($\kappa=.99$). Although the methods of Experiment 3 closely replicated those of Experiment 2, the pattern of results differed. Although in the same numerical direction, an independent samples t -test failed to reveal a significant difference in proportion correct on the cued recall test between participants in the truth-checking and accuracy-given conditions, $t(163)=1.49, p=.137, d=0.23$ (Table 1). However, an independent samples t -test revealed that proportion correct on the two-choice judgement task was significantly higher in the accuracy-given condition than the truth-checking condition, $t(163)=2.66, p=.009, d=0.41$ (Table 1).

Learning efficiency. We again calculated learning efficiency scores for each participant by dividing proportion correct on the cued recall test by their median time per true statement in the study phase (Table 2). As in Experiment 2, an independent samples t -test revealed that participants in the truth-checking condition learned significantly less per second of studying than participants in the accuracy-given condition, $t(163)=4.23, p<.001, d=0.66$.

In sum, the method used in Experiment 3 closely replicated the method used in Experiment 2, with the main change being the 12-min delay between the study phase and the final test. If or why this small difference could account for the discrepant results is unclear. Whereas Experiment 2 found truth-checking significantly benefited the recall of true information without affecting the judgement of false information, Experiment 3 failed to find evidence that truth-checking affects recalling true information. However, truth-checking significantly increased participants' likelihood of identifying previously read false statements as true. Moreover, when evaluated in terms of learning efficiency, both experiments' results suggested that truth-checking produced the least amount of learning per second.

General discussion

The goal of the current experiments was to better understand how applying uncertainty as a desirable difficulty could impact learning new information. Participants in the truth-checking condition were exposed to both true and false information about history. Using their existing

knowledge, they were asked to try to identify which statements were true and which statements were false. This activity, which we refer to as truth-checking, could be expected to lead learners to encode statements more deeply and/or integrate the to-be-learned information more effectively with existing knowledge, thus enhancing memory for the true information. In Experiment 1, participants were exposed to blocks of four true statements and one false statement, and they were asked to determine which four statements were true and which statement was false. In Experiments 2 and 3, participants were exposed to each statement individually and were asked to determine whether each statement was true or false. After judging, participants in Experiments 1–3 were informed whether each statement was indeed true or false. Learning was then assessed by a final cued recall test for the true information, which was administered either immediately following the study phase (Experiments 1 and 2) or after a 12-min delay (Experiment 3).

The results observed across the experiments were mixed. In Experiment 1, we failed to observe any evidence of truth-checking enhancing learning of the true information. Specifically, participants in the truth-checking condition performed no differently than participants in the accuracy-given and all-true conditions on the cued recall test. A significant effect was observed in Experiment 2, however, with participants in the truth-checking condition outperforming participants in the other two conditions. We hypothesised that the benefit of truth-checking was observed in Experiment 2 because of the procedural change, which we anticipated would increase attention and deep processing of each statement. Unlike in Experiment 1, in which participants only needed to select the one false statement out of the five presented, participants in Experiment 2 judged the truthfulness of each statement individually. This significant benefit of truth-checking, however, failed to replicate in Experiment 3. No significant difference was observed between the truth-checking condition and the accuracy-given condition for cued recall test performance, but the results were in the same numerical direction as in Experiment 2. The only difference between Experiments 2 and 3 was the 12-min delay added between the study phase and the cued recall task. Between-experiment comparisons should be interpreted with caution. One possible explanation is that the benefit of truth-checking in this context is relatively small and unreliable. Although it may have emerged as significant in Experiment 2, it did not emerge as significant in Experiment 3.

Whether and under what conditions truth-checking might lead to strong and consistent learning benefits remains to be seen. True–false practice questions are similar to our truth-checking activity and have been shown to enhance memory for tested information (Brabec et al., 2021; Uner et al., 2022). For example, Brabec and colleagues (2021) had participants read a brief text, either

read true statements or answer true–false questions about what they read, and then take a final cued recall test. Participants who answered true–false questions recalled more true facts on the final test than participants who only read true statements. The primary difference between truth-checking and true–false testing is that truth-checking involves evaluating statements without prior explicit instruction about the content. Therefore, relevant prior knowledge may be necessary to benefit from truth-checking. Perhaps truth-checking was not a desirable difficulty because it was too difficult. Indeed, participants in the truth-checking condition were essentially unable to discriminate between true and false statements during the study phase. Nevertheless, prior research has shown that attempting to answer questions prior to learning the material—or pretesting—can enhance memory for the pretested content, although having at least some related prior knowledge increases the benefits of pretesting (Overoye et al., 2021). Future research should examine whether the amount of related prior knowledge moderates the benefits of truth-checking, perhaps making it more effective for some stages of the learning process than others.

Another possibility is that the truth-checking manipulation did not increase attention and deep processing as we predicted. The truth-checking manipulation may have been more effective if participants had been given both a true and false version of each statement during learning, thus forcing them to think critically about why a given statement was true or false. For example, had participants initially read two statements that tomatoes are originally from Peru (true) and Ecuador (false), they may have recalled relevant details such as Peru becoming independent before Ecuador to aid them in recalling the true statement that tomatoes originated in Peru first. Even if participants did not have this prior knowledge, seeing both versions of the same statement could have prompted participants to generate some explanation for what makes Peru the correct answer, improving recall of Peru on the final test (Bisra et al., 2018). Participants may have been less likely to engage in this type of elaboration because they were only presented the false versions of some statements and true versions of other statements. In related studies, refutation texts present participants with both true and false versions of the same statement, and seeing both statement versions has been shown to aid participants in overcoming false beliefs (Schroeder & Kucera, 2022). Furthermore, showing a true and false version of each statement may direct participants' attention to the key detail that would be tested (e.g., the country name). In the present studies, participants did not know which details from the true statement they would need to recall later, so participants' attention and elaboration efforts may have possibly been directed towards the wrong information in the sentence. Future research should investigate how different ways of operationalising truth-checking affect the

quality of encoding processes including attention, study time, and elaboration.

Taken together, the current results provide inconsistent evidence of uncertainty benefitting learning. However, even when participants in the truth-checking condition outperformed participants in the accuracy-given condition (Experiment 2), this benefit occurred alongside an increase in the amount of time spent studying. Therefore, when we calculated learning efficiency scores for the conditions, participants in the truth-checking condition showed the least benefit in learning per second of study time of any of the three conditions. This observation is important given that participants in the all-true and accuracy-given conditions could have used that extra time to further study the to-be-learned statements. Additional study time does not always translate into more learning (e.g., Nelson & Leonesio, 1988), and thus we hesitate to conclude that learning in the truth-checking condition would have necessarily been less efficient than in the other two conditions if we had controlled study time. The point is that learners have the option to use extra study time to engage in more powerful learning strategies. Effective learning strategies such as spaced retrieval practice can take additional time, but this time cost is often worth it given the robust benefits to long-term memory (Dunlosky et al., 2013). In contrast, at least as demonstrated in the current study, truth-checking yielded small and inconsistent benefits. Although truth-checking did not harm recall, the efficiency analyses suggest that any learning that could be gained may not be worth the time cost, especially if it takes time away from using more effective strategies such as retrieval practice.

A secondary goal of the current study was to examine a potential negative side effect of engaging in truth-checking, specifically being exposed to false information (Brashier et al., 2020). After completing the cued recall test, participants were administered a two-choice judgement task in which each false statement was presented along with the true version of the statements, and participants were asked to identify which statement was true. Participants in the truth-checking and accuracy-given conditions would be able to complete this task based on what they learned in the initial study phase. Specifically, they could rule out the statement versions that they were told were false (either immediately or after guessing). Participants in the all-true condition were never exposed to any of the false statements and would need to answer based on their pre-existing knowledge. Of note, participants in the truth-checking condition performed significantly better on this task than the other two conditions only in Experiment 1 but not in Experiments 2 or 3. Indeed, performance was significantly lower in the truth-checking condition than in the accuracy-given condition in Experiment 3. This pattern suggests that the truth-checking manipulation was not very effective at helping participants reject false information when it was later re-exposed.

Moreover, the results of Experiment 3 suggest that the dangers of being presented with false information may become more pronounced following even a very short delay, and an even longer delay could have possibly exacerbated this effect (e.g., Grady et al., 2021). Having the chance to study the true versions of the false statements likely would have reduced the probability of participants endorsing the false statements as true on the subsequent two-choice judgement test (Mullet & Marsh, 2016). In the current study, participants were simply told that the false statements were false, but not why they were false, or how they could be revised to be true.

Of course, when operationalised differently, truth-checking could become a far more effective learning manipulation applied in educational settings without causing memory for false information, but future research will be needed to identify if and under what conditions that is the case. Given the increased use of the internet for obtaining information, examining how truth-checking with the aid of the internet might affect learning of true information and subsequently endorsing false information is important. Internet consumers might apply truth-checking to assess the accuracy of information they are reading online, such as current events, commercials and advertisements, and social media posts. Truth-checking might also be applied to judge the accuracy of information such as political statements, sales pitches, or responses provided by artificial intelligence chatbots like Chat-GPT, which have been shown to occasionally produce hallucinatory or false information. Truth-checking could lead people to think more deeply about the information they seek to verify, thereby mitigating the negative effects of internet use on memory for to-be-learned information.

Several limitations of the current study deserve emphasis. First, because participants were tested remotely due to Covid restrictions, data collection was conducted online in a relatively uncontrolled environment. Thus, we had little control over how participants read the to-be-learned material, whether they took breaks between parts of the experiment, or whether they dropped out of the experiment completely. Although an in-person version of the study might have provided additional control (e.g., study time per statement, ensuring that participants attempted to answer all of the test questions), a substantial amount of learning—at the undergraduate level especially—does take place in uncontrolled, online environments, thus giving the current study a certain degree of ecological validity. Second, we examined the potential benefits of truth-checking in a relatively focused set of conditions and with a specific type of to-be-learned information. Possibly, other types of to-be-learned information might be affected differently. The history statements used in the present study were relatively difficult, for example, with participants unable to identify statements' accuracy much better than at chance. Perhaps truth-checking would be more

beneficial in contexts where learners can bring more of their prior knowledge to bear when judging the truthfulness of information to which they are exposed. Applying prior knowledge would thus support creating new retrieval pathways that have the potential to spur memory. Although the current study may have failed to observe a strong and reliable benefit of truth-checking, one might still exist under other yet to be identified learning conditions.

Acknowledgements

The authors gratefully acknowledge the study participants. They also thank the following individuals for collecting and analysing data, as well as for preparing survey designs, stimuli, and references: Lilia X. Diaz, Kathryn E. Echandia-Monroe, Sreejita Ghose, Nicholas T. Hilberg, Oscar O. Huete, Abigail Shamelashvili, Hanyue Tang, Kirstyn Tara, Calia P. Thompson, and especially Dr. Moyses Marcos for his help in generating the statements about history.

Declaration of conflicting interests

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

Funding

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

ORCID iDs

Karen Arcos  <https://orcid.org/0000-0003-0651-4160>

Benjamin C Storm  <https://orcid.org/0000-0001-7645-9991>

Supplemental material

The supplementary material is available at qjep.sagepub.com.

Availability of data and materials

Statements, data, and analyses for all experiments are available in the Open Science Framework (OSF) project: https://osf.io/d3qs4/?view_only=a284af15f78d431b8e0441173eecf6fd. None of the experiments were preregistered.

Note

1. We opted to have participants in the all-true condition study 40 statements instead of 50 statements to control for the total number of true statements, and to create a somewhat higher bar for comparison with the truth-checking condition. Given the item-specific cued-recall nature of the final test, however, we would not expect such a small difference in the number of statements to significantly impact final recall performance.

References

- Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational*

- Psychology Review*, 30, 703–725. <https://doi.org/10.1007/s10648-018-9434-x>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.
- Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, 9(4), 475–479. <https://doi.org/10.1016/j.jar-mac.2020.09.003>
- Brabec, J. A., Pan, S. C., Bjork, E. L., & Bjork, R. A. (2021). True-false testing on trial: Guilty as charged or falsely accused? *Educational Psychology Review*, 33(2), 667–692. <https://doi.org/10.1007/s10648-020-09546-w>
- Brashier, N. M., Eliseev, E. D., & Marsh, E. J. (2020). An initial accuracy focus prevents illusory truth. *Cognition*, 194, 104054. <https://doi.org/10.1016/j.cognition.2019.104054>
- Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin & Review*, 28, 1839–1847. <https://doi.org/10.3758/s13423-021-01904-1>
- Brod, G., & Breitwieser, J. (2019). Lighting the wick in the candle of learning: Generating a prediction stimulates curiosity. *NPJ Science of Learning*, 4(1), 1–7. <https://doi.org/10.1038/s41539-019-0056-y>
- Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, 55, 22–31. <https://doi.org/10.1016/j.learninstruc.2018.01.013>
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, 1(1), 69–84. <https://doi.org/10.1007/s11409-006-6894-z>
- Chan, M.-p. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241–250. <https://doi.org/10.1080/1358165042000283101>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. [https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Cyr, A.-A., & Anderson, N. D. (2015). Mistakes as stepping stones: Effects of errors on episodic memory among younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 841–850. <https://doi.org/10.1037/xlm0000073>
- Cyr, A.-A., & Anderson, N. D. (2018). Learning from your mistakes: Does it matter if you're out in left foot, I mean field? *Memory*, 26(9), 1281–1290. <https://doi.org/10.1080/09658211.2018.1464189>
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14(2), 238–257. <https://doi.org/10.1177/1088868309352251>
- D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, 25(7), 1299–1308. <https://doi.org/10.1080/02699931.2011.613668>
- D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170. <https://doi.org/10.1016/j.learninstruc.2012.05.003>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review*, 16(1), 88–92. <https://doi.org/10.3758/PBR.16.1.88>
- Grady, R. H., Ditto, P. H., & Loftus, E. F. (2021). Nevertheless, partisanship persisted: Fake news warnings help briefly, but bias returns with time. *Cognitive Research: Principles and Implications*, 6(1), 1–16. <https://doi.org/10.1186/s41235-021-00315-z>
- Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in Cognitive Sciences*, 23(12), 1014–1025. <https://doi.org/10.1016/j.tics.2019.10.003>
- Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive Research: Principles and Implications*, 6(1), 1–12. <https://doi.org/10.1186/s41235-021-00301-5>
- Huelsen, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition*, 40(4), 514–527. <https://doi.org/10.3758/s13421-011-0167-z>
- Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & 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>
- Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, 115, 101237. <https://doi.org/10.1016/j.cogpsych.2019.101237>
- Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(9), 1297–1317. <https://doi.org/10.1002/acp.1537>
- Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1), 106–114. <https://doi.org/10.1037/a0033699>

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, 34(5), 1140–1149. <https://doi.org/10.3758/BF03193260>
- Mera, Y., Rodríguez, G., & Marin-García, E. (2022). Unraveling the benefits of experiencing errors during learning: Definition, modulating factors, and explanatory theories. *Psychonomic Bulletin & Review*, 29, 753–765. <https://doi.org/10.3758/s13423-021-02022-8>
- Metcalf, J. (2017). Learning from errors. *Grantee Submission*, 68, 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>
- Muller, D. A., Bewes, J., Sharma, M. D., & Reimann, P. (2008). Saying the wrong thing: Improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning*, 24(2), 144–155. <https://doi.org/10.1111/j.1365-2729.2007.00248.x>
- Mullet, H. G., & Marsh, E. J. (2016). Correcting false memories: Errors must be noticed and replaced. *Memory & Cognition*, 44(3), 403–412. <https://doi.org/10.3758/s13421-015-0571-x>
- Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews*, 36(7), 1640–1645. <https://doi.org/10.1016/j.neubiorev.2012.03.001>
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the “labor-in-vain effect.” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4), 676–686. <https://doi.org/10.1037/0278-7393.14.4.676>
- Overoye, A. L., James, K. K., & Storm, B. C. (2021). A little can go a long way: Giving learners some context can enhance the benefits of pretesting. *Memory*, 29(9), 1206–1215. <https://doi.org/10.1080/09658211.2021.1974048>
- Overoye, A. L., & Storm, B. C. (2015). Harnessing the power of uncertainty to enhance learning. *Translational Issues in Psychological Science*, 1(2), 140–148. <https://doi.org/10.1037/tps0000022>
- Pan, S. C., Sana, F., Schmitt, A. G., & Bjork, E. L. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition*, 9(4), 542–554. <https://doi.org/10.1016/j.jarmac.2020.07.004>
- Peterson, D. J., & Wissman, K. (2020). Using tests to reduce mind-wandering during learning review. *Memory*, 28(4), 582–587. <https://doi.org/10.1080/09658211.2020.1748657>
- Rapp, D. N., Hinze, S. R., Kohlhepp, K., & Ryskin, R. A. (2014). Reducing reliance on inaccurate information. *Memory & Cognition*, 42(1), 11–26. <https://doi.org/10.3758/s13421-013-0339-0>
- 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>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Schroeder, N. L., & Kucera, A. C. (2022). Refutation text facilitates learning: A meta-analysis of between-subjects experiments. *Educational Psychology Review*, 34(2), 957–987. <https://doi.org/10.1007/s10648-021-09656-z>
- Smith, B. L., Holliday, W. G., & Austin, H. W. (2010). Students’ comprehension of science textbooks using a question-based reading strategy. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 47(4), 363–379. <https://doi.org/10.1002/tea.20378>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>
- Stare, C. J., Gruber, M. J., Nadel, L., Ranganath, C., & Gómez, R. L. (2018). Curiosity-driven memory enhancement persists over time but does not benefit from post-learning sleep. *Cognitive Neuroscience*, 9(3–4), 100–115. <https://doi.org/10.1080/17588928.2018.1513399>
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16), 6313–6317. <https://doi.org/10.1073/pnas.1221764110>
- Umanath, S., Butler, A. C., & Marsh, E. J. (2012). Positive and negative effects of monitoring popular films for historical inaccuracies. *Applied Cognitive Psychology*, 26(4), 556–567. <https://doi.org/10.1002/acp.2827>
- Uner, O., Tekin, E., & Roediger, H. L., III. (2022). True–false tests enhance retention relative to rereading. *Journal of Experimental Psychology: Applied*, 28, 114–129. <https://doi.org/10.1037/xap0000363>
- Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, 28(3), 247–253. <https://doi.org/10.1177/0963721419827854>
- Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, 160, 110–126. <https://doi.org/10.1016/j.cognition.2016.12.016>
- Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1073–1092. <https://doi.org/10.1037/xlm0000363>