Pretesting Boosts Item but Not Source Memory

Tina Seabrooke[1], Chris J. Mitchell[2], and Timothy J. Hollins[2]

[1] Department of Psychology, University of Southampton

[2] Department of Psychology, University of Plymouth

Word count: 6326

Author Note

Correspondence concerning this article should be addressed to Tina Seabrooke, Department of Psychology, University of Southampton, Hampshire, SO17 1PS, UK. E-mail: t.seabrooke@soton.ac.uk.

This manuscript is the version accepted for publication in *Memory*. It is not the copy of record and may not exactly replicate the authoritative document later published in *Memory*. The final article will be available upon publication in *Memory*.

Abstract

Two experiments examined the effect of pretesting on target recognition and source memory. In an initial encoding phase, participants attempted to learn the common English definitions of rare English words. For each rare word, the participants either guessed the definition of the rare English word before it was revealed (Pretest condition), or just studied the complete word pair without first guessing the definition (Read-only condition). To manipulate source information, the targets were either presented in different colours (Experiment 1) or lists (Experiment 2). In both experiments, the participants correctly recognised more targets from Pretest trials than Read-only trials, but showed no difference in source memory. Pretesting therefore appears to improve target recognition memory, but not memory for contextual information. The results are discussed in relation to semantic and episodic theories of the pretesting effect.

Keywords. Tests, Errors, Generation, Item recognition, Source memory.

Pretesting Boosts Item but Not Source Memory

Tests are a fundamental feature of many education systems. Although tests are widely to use to assess what students have learnt at the end of a module, there is now abundant evidence that tests can also improve memory in and of themselves (for a review, see Roediger & Butler, 2011). Relative to an equivalent period of studying, tests improve memory for information that students studied before the test (the *backwards testing effect*; Roediger & Karpicke, 2006; Rowland, 2014), and for unrelated information that they study *after* the test (the *forwards testing effect*; Szpunar et al., 2008; Yang et al., 2018). These backwards and forwards testing effects demonstrate that tests are powerful tools not only for assessment, but also for learning.

Recent research has examined whether even unsuccessful tests for individual items can improve learning when accompanied with corrective feedback. Kornell et al. (2009), for example, asked their participants to learn semantically related word pairs such as *pond-frog* and *whale-mammal*. On Pretest trials, the participants were shown the "cue" (e.g., *pond*) and had to guess the "target" (*frog*) before it was revealed. Since the cue and target were only weakly associated, and the targets were not presented before, the participants' guesses were usually incorrect (and hence the test was deemed "unsuccessful"). On Read-only trials, by contrast, the complete word pairs (e.g., *whale-mammal*) were simply presented for study for a duration matched to the Pretest trials. In a subsequent cued recall test, the participants recalled more targets (when prompted by the cue, e.g., *pond-?*) from Pretest trials than Read-only trials. Importantly, this pattern was seen even when the few Pretest targets that the participants correctly guessed at encoding were removed from the dataset. Thus, pretesting improved subsequent cued recall

(relative to study-alone), even when the participants generated incorrect responses on every Pretest trial. This basic effect – which we refer to as the *pretesting effect (PTE) –* has now been replicated many times (Carneiro et al., 2018; Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012; Knight et al., 2012; Kornell, 2014; Seabrooke, Mitchell, Wills, & Hollins, 2021; Vaughn & Rawson, 2012).

Metcalfe and Huelser (2020) recently conducted an elegant set of experiments to examine the mechanisms that mediate the PTE. In their experiments, participants were presented with word triplets, in which the second word in each triplet had multiple meanings (e.g., *palm*, which refers to both a type of tree and part of the hand). On congruent trials, the first and last word in each triplet tapped into the same interpretation of the second word (e.g., *wrist-palm-hand*). On incongruent trials, by contrast, the first and last word in each triplet tapped into *different* interpretations of the second word (e.g., *tree-palm-hand*). On Pretest trials, participants were presented with the first two words from the triplet and had to guess the third word before it was revealed. Thus, the participants were much more likely to correctly guess an associate of the target on congruent compared to incongruent Pretest trials. On Read-only trials, the complete triplet was presented for study for an equivalent duration. Participants then completed a cued recall test, where the first two words from each triplet were presented, and they had to recall the third word, as well as their original guess. In two experiments, participants showed better cued recall of triplets from the Pretest condition than the Read-only condition, regardless of whether the triplet was presented in the congruent or incongruent condition. Importantly, however, this pattern was only seen when the participants could also recall their original *guess* on Pretest trials. For trials where participants could not

recall their original guess, pretesting either produced no benefit (for congruent trials) or impaired recall (for incongruent trials). Thus, pretesting only improved subsequent cued recall of the targets when the participants could also recall their guess from the encoding phase. When the participants successfully recalled their guess, pretesting was beneficial for both congruent and incongruent trials.

Metcalfe and Huelser (2020) proposed an episodic recollection account of their data and the PTE more generally. They argued that the fact that the PTE depends on the participants being able to recall their original guesses suggests that the PTE relies on having good recollection of the original encoding trial. In essence, this reasoning equates recollection to the participants' ability to recall their guesses, which is a form of criterial-based recollection measure (Gallo et al., 2004). This finding is also consistent with previous failures to observe the PTE in participants who could not recall their original guesses (Knight et al., 2012; Vaughn & Rawson, 2012; Yan et al., 2014). Together, these results support the suggestion that a strong episodic memory trace is essential for the PTE. In the present work, we test this idea using a different criterial test of recollection - source memory - as described below.

An alternative, *attentional* account of the PTE posits that pretesting increases participants' curiosity and/or motivation to study, which then boosts attention to the correct feedback on Pretest trials. Support for this attentional account comes from Potts et al.'s (2019) Experiment 4, where participants had to guess the English definitions of rare English words (e.g., *roke-mist*) before they were revealed. On each trial, the participants also rated their curiosity to learn each answer either before or after their guess. Higher curiosity ratings were recorded when the participants rated their curiosity *after* having

made a guess. Similarly, Seabrooke, Mitchell, et al. (2019) found that participants rated their motivation to learn face-fact pairings more highly after having guessed the fact. Both findings suggest that increased attention (via increased curiosity/motivation) may play an important role in the PTE. While the episodic recollection and attentional accounts of the PTE need not be mutually exclusive, they do posit different mechanisms for the effect.

To this point, we have considered both the episodic and attentional accounts of the PTE in their broadest sense, and they are entirely compatible with one another. Pretesting may encourage participants to pay greater attention to targets, and thereby facilitate the formation of a richer memory trace that later enhances recall and source memory. Indeed, the *hypercorrection effect*, where participants show better memory for answers after making high-confidence errors than low-confidence errors (Butterfield & Metcalfe, 2001), is consistent with this view. Fazio and Marsh (2009), for example, found that participants remembered both the correct answers and the colour that they were presented in if the answer was presented after a high-confidence error (versus a low-confidence error). This finding was taken as evidence to suggest that high-confidence errors led participants to pay more attention, and therefore deeply encode, the correct answers, much as we have characterised the episodic and attentional accounts to date.

An alternative possibility is that, rather than evoking *more* processing, pretesting changes *how* target items are processed (relative to Read-only trials). Following a guess about a target's meaning, participants may focus particularly on the meaning of the target, potentially even at the expense of contextual information. This perspective is conceptually close to Mulligan's (2004) processing account of the observed dissociations

between item and source memory seen in experiments on the *generation effect.* In a series of studies, Mulligan (2004) had participants either read antonym pairs (e.g., *hot-cold*) or generate them (e.g., *hot-c___?*). In a series of experiments, the words were either presented in one of two colours (Experiments 1, 3 and 5), in one of two locations (Experiments 2 and 4), or against one of two background colours (Experiments 6 and 7). In each experiment, participants were tested on recognition memory for the items and the contextual source the words were presented in. In all experiments, source memory was reliably above chance and item memory was boosted by generation. The effects of generation on context memory (item colour, item location, background colour), however, varied across experiments (see also Mulligan, 2011). When item colour was manipulated, the Generation condition produced significantly *worse* memory than the Read condition. When spatial location and background colour was manipulated, however, generation had no significant impact on source memory. To explain these findings, Mulligan (2004) proposed that generation changes *what* is encoded about the stimulus. In particular, he argued that the Read condition promoted greater perceptual encoding of the target word (but not extrinsic features of the context, such as location or background colour) relative to the Generation condition, which instead promoted conceptual processing.

Mulligan's (2004) processing account could also be applied to the pretesting effect to predict that pretesting, relative to study-alone, would have differential effects on item and source memory. That is, participants may encode the conceptual aspects, but not the perceptual features, of Pretest targets more deeply than that of Read-only targets. Mulligan's (2004) processing account would therefore predict that pretesting would boost memory for the target, but not its source.

In two experiments, we examined the effect of pretesting on source memory by presenting the targets in one of two colours (Experiment 1) or lists (Experiment 2) at encoding. Following previous research, we predicted that participants would show better recognition of targets that they had incorrectly guessed at encoding than those that they had simply studied (Potts et al., 2019; Seabrooke, Hollins, et al., 2019; Seabrooke, Mitchell, et al., 2019; Seabrooke, Mitchell, Wills, & Hollins, 2021; Seabrooke, Mitchell, Wills, Inkster, et al., 2021). The question of main interest here was whether pretesting would similarly improve source memory. The episodic recollection account of the PTE (Metcalfe & Huelser, 2020) predicts that pretesting should only improve subsequent target memory for targets in which they have good episodic memory of the initial encoding event. The episodic recollection theory therefore predicts that pretesting will only improve subsequent target recognition when participants can also recollect the source. Moreover, improvements in target memory should be accompanied by improvements in source memory. That is, the episodic recollection theory predicts that pretesting will improve source memory, relative to study-alone. The attentional account of the PTE, in which pretesting is suggested to direct attention and improve processing of the target, also predicts that pretesting should improve source memory, relative to study-alone (Potts et al., 2019; Seabrooke, Mitchell, et al., 2019; Zawadzka & Hanczakowski, 2019). Such a finding would be consistent with results seen in the hypercorrection literature (Fazio & Marsh, 2009). Thus, both the episodic recollection and attentional accounts of the PTE predict that pretesting should improve both target and source memory, relative to study-alone.

Mulligan's (2004) processing account (as applied to the PTE), by contrast, predicts that pretesting will improve processing of the conceptual (meaning) aspects of the target, but not perceptual features such as target colour. When Mulligan (2004) manipulated the colour of the stimuli in his work on the generation effect, he found that generation improved target memory but impaired source memory, relative to study-alone. In the generation effect paradigm, however, participants must generate the targets themselves on Generate trials. The targets themselves were therefore not presented in colour. In PTE paradigms, by contrast, the participants generate an error on the Pretest trials, but then see corrective feedback in the same format as in the Study-only condition. Thus, they see the same perceptual information, and so there may well be no impairment for this information. From the perspective of the processing account, it is therefore plausible that pretesting will have no impact on source memory.

Experiment 1

Experiment 1 combined the typical pretesting paradigm used by Potts and colleagues (Potts et al., 2019; Potts & Shanks, 2014) and Seabrooke, Hollins, et al. (2019) with a source memory procedure. Participants first attempted to learn the common English definitions of rare English words (e.g., *roke-mist*). The participants either studied the complete word pair for the full trial duration (Read-only trials) or guessed the target (*roke-?*) before it was revealed (Pretest trials). To manipulate source information, half the targets (the definitions) in each encoding condition were presented in blue and half in pink, conceptually replicating Mulligan's (2004) colour manipulation. To ensure that only unsuccessful pretests were examined, any Pretest trials in which the participants generated the correct answer at encoding were not analysed at test. In a subsequent old-

new recognition test, the targets from the encoding phase, plus new words, were individually presented in black. The participants first had to determine whether the words had been presented before and, then, regardless of their initial response, which colour those words had been were presented in, guessing where necessary. We expected the participants to show better recognition of the Pretest targets than the Read-only targets. The critical question was whether differences between the encoding conditions would also be seen in source memory.

Method

Participants

Thirty-nine participants were recruited from Prolific (www.prolific.co) for £2 each. We excluded participants who stated that English was not their first language, those who failed a question check to determine whether their screen and keyboard was working correctly, and those who admitted to writing the word pairs down or looking the answers up during the experiment. Three participants were excluded on this basis, leaving 36 participants (23 females, 12 males, 1 unknown), who were aged between 18 and 49 years ($M = 29.47$ years, $SD = 9.31$ years). This sample size has good power (94%) to detect the average pretesting effect size (Cohen's $d_z = 0.61$) that was observed in our previous target recognition tests (Seabrooke, Hollins, et al., 2019, Experiments 1 and 3). All power analysis were conducted using the *pwr* package (Champely, 2017) in RStudio (R Core Team, 2021). The experiment was approved by the University of Plymouth Psychology Ethics Committee.

Materials

Forty-four rare English word pairs and their definitions were used as stimuli. The word pairs were selected from those used by Potts and Shanks (2014) and Seabrooke, Hollins, et al. (2019). For each participant, the word pairs were randomly allocated to the Pretest or Read-only conditions, or as novel words (foils) in the target recognition test. The experiment was programmed in JavaScript using the jsPsych library (de Leeuw, 2015), hosted on JATOS (Lange et al., 2015), and was presented in full screen on the participants' own web browsers. All text was presented in black and on a white screen, unless otherwise stated. The participants completed the experiment remotely using a range of desktop and laptop PCs (not tablets or phones).

Procedure

The experiment was advertised on Prolific (www.prolific.co), with the maximum experiment duration advertised as 16 minutes. After providing informed consent, the participants provided information about the browser they were using to view the experiment, their age, gender, and English language fluency. They also were also shown a grid of letters, with one letter presented in red and the rest in black. The participants were asked to select the keyboard button that corresponded to the red letter. This simple question check was included to screen potential computer 'bots' or non-attentive participants, and to ensure that stimuli and responses were recording properly for each participant. Before the encoding phase, the participants were asked to turn off any music and distractions, to complete the experiment in one sitting, and to keep the browser in full screen and not visit other webpages during the experiment.

Before the encoding phase, the participants were instructed that they should attempt to learn the common English definitions of rare English words, by either just

studying the word pairs, or by guessing the definitions before they were revealed. They were told that the definitions would be presented in either blue or pink, and they were asked to study both the words and the colours in anticipation of a later test. Participants were required to agree to complete the experiment without using additional memory aids (e.g., writing the words down). The participants then completed four practice trials (two Pretest and Read-only trials each, with one target from each condition presented in blue and pink each), before moving on to the main encoding phase.

The main encoding phase consisted of 10 Pretest and 10 Read-only trials. The targets were presented in blue on half of the trials in each encoding condition, and pink on the rest. All other text was presented in black. On Pretest trials, the participants were presented with a rare English word (e.g., *roke*) and were asked to guess the common English definition and type it into a textbox that was presented on-screen. A reminder to "Please guess the definition" was presented above the textbox. The textbox was presented for seven seconds, during which time the participants could use the Backspace key to change their answer. After the seven seconds, the complete word pair (e.g., *roke = mist*) was presented for a further five seconds. On Read-only trials, the complete word pair was simply presented for five seconds. We matched the two conditions on target duration because we were interested in memory for the *targets*[1]. Throughout the experiment, the

[1] Previous work on the pretesting effect has sometimes matched the Pretest and Read-only conditions on total trial duration rather than target duration. Most research has observed comparable findings when the encoding conditions were matched on total trial time and target presentation time (Grimaldi & Karpicke, 2012; Knight et al., 2012).

trials were randomly ordered for each participant and were separated by one-second intervals.

The instructions for the test phase were presented immediately after the encoding phase. The participants read the instructions and completed four practice trials (using the targets from the practice encoding trials). The main test then consisted of 40 trials, which included the 20 targets from the encoding phase, plus 20 novel foils. On each test trial, the participants were first presented with a target or foil and were asked to indicate whether the word was presented at encoding (choosing between 'yes' and 'no' options and guessing if necessary). Regardless of their response, they were then asked what colour the word was presented in (choosing between 'blue' and 'pink'). The participants were explicitly instructed to choose a colour even if they did not remember seeing the word before, so that the number of source memory trials was not dependent on target recognition performance. All text was presented in black, except for the response options for the source memory test ("blue" and "pink"), which were presented in their respective colours. Responding was not time-limited during the test, and a response was required for each question to progress.

After the test, the participants were asked to indicate whether they used additional memory aids during the experiment (e.g., writing the word pairs down). They were paid regardless of their answer (and were instructed so) to encourage honesty, but the data from participants who admitted to using additional aids were not analysed.

Results

Three participants guessed the correct target for one cue each at encoding. Those targets were removed from further analysis for those participants. The foils were novel

words that were not presented at encoding, so were not attached to either encoding

condition. Any differences in discrimination (d') and response bias (c) between the

Pretest and Read-only conditions must, therefore, reflect differences in the hit rates. We

therefore took the percentage of correct responses to foils (correct rejections) and targets

(hits) as our primary measures of interest, rather than d' or c (see Mulligan, 2004, for the

same approach).

In this experiment and the next, we examined the relationship between the PTE

and source and item memory in two ways, so that our findings are comparable with other

research that used different approaches. We first examined the impact of our

experimental manipulations directly on item and source memory. Then, in line with

Metcalfe and Huelser (2020), we looked at the magnitude of the PTE both when the

participant did and did not report correct source information for each item.

During the test phase, the participants were very good at identifying the foils as

novel ($M = 90.69\%$, $SD = 11.72\%$). Figure 1a shows the percentage of correct responses

(hits) for Pretest and Read-only trials in the target recognition and source memory tests.

In the target recognition test, participants correctly recognised more Pretest targets than

Read-only targets, $t(35) = 7.51$, $p < .001$, $d_z = 1.25$, $BF_{10} > 100$. All Bayes Factors were

calculated using Version 0.9.12.4.2 of the *BayesFactor* package (Morey et al., 2015) in

RStudio (R Core Team, 2021).

In the source memory test, overall accuracy (excluding the foils; $M = 55.42\%$, SD

$= 13.37\%$) was significantly above chance in a one-sample t-test, $t(35) = 2.43$, $p = .02$,

but the Bayes Factor was inconclusive, $BF_{10} = 2.33$. Source memory for the Pretest and

Read-only conditions did not differ, $t(35) = 0.47$, $p = .64$, $d_z = 0.08$, $BF_{10} = 0.20$.

We also examined the size of the PTE for target recognition across trials in which the participants did and did not accurately recall the source. An Encoding condition (Pretest, Read-only) × Source accuracy (Correct, Incorrect) repeated-measures ANOVA revealed a significant main effect of encoding condition, $F(1, 35) = 43.46$, $p < .001$, generalised eta squared (η_g^2) = .20, $BF_{10} > 100$. The participants correctly recognised more targets from Pretest trials ($M = 90.33\%$, $SD = 15.81\%$) than Read-only trials ($M = 68.03\%$, $SD = 28.12\%$). No significant main effect of source accuracy, $F(1, 35) = 0.22$, $p = .64$, $\eta_g^2 = .0008$, $BF_{10} = 0.19$, or encoding condition × source accuracy interaction, $F(1, 35) = 0.19$, $p = .67$, $\eta_g^2 = .0007$, $BF_{10} = 0.25$, was observed.

Discussion

Experiment 1 revealed a strong PTE in target recognition. When participants incorrectly guessed the definitions of rare English words at encoding, they were more likely to correctly recognise the corrective feedback in a later test than if they had simply studied the definitions without generating a guess. This pattern is consistent with previous results (Potts et al., 2019; Potts & Shanks, 2014; Seabrooke, Hollins, et al., 2019; Seabrooke, Mitchell, et al., 2019; Seabrooke, Mitchell, Wills, & Hollins, 2021; Seabrooke, Mitchell, Wills, Inkster, et al., 2021). The pattern also reiterates that, in contrast to earlier thoughts, pretesting can also improve memory for targets that were presented in semantically unfamiliar and unrelated word pairs, at least in target recognition tests.

The second analysis, based upon whether participants remembered the colour that the targets were presented in at encoding, revealed that the target recognition PTE is not dependent on good episodic memory of the source. That is, there was no evidence that

the PTE was modulated by whether the participants recalled accompanying source information. Thus, there was a clear dissociation between item and source memory. Pretesting boosted item memory, but not source memory, and there was no evidence that the PTE was dependent on successful recollection of the original event (defined by accurate source recall). Although participants showed above chance source memory, no difference was observed between the Pretest and Read-only conditions (and the Bayes Factor supported the null).

The observation that pretesting improves item recognition but not source memory speaks against both the episodic recollection (Metcalfe & Huelser, 2020) and the general attentional (Potts et al., 2019; Seabrooke, Mitchell, et al., 2019; Zawadzka & Hanczakowski, 2019) accounts of the PTE. These theories both predict that pretesting should have improved target and source memory. Instead, the data are more consistent with the idea that pretesting improves processing (and therefore subsequent memory) of the *conceptual*, but not *perceptual*, aspects of the target (Mulligan, 2004).

It should be noted, however, that while the level of source memory was above chance, it was not high. It is entirely possible that successful (versus unsuccessful) guessing washed out any potential benefit of pretesting on source memory. To test this idea, we changed the source dimension under manipulation to list-membership, with the materials presented in two separated lists. In other research from our laboratory, we have found that, relative to source manipulations of colour, list membership shows greater evidence of clustering and higher source discrimination (Randle, 2021). Thus, in Experiment 2 we anticipated that source memory would be well above chance, which would give greater opportunity to detect a PTE on source memory.

Experiment 2

Experiment 2 employed a list-discrimination task to further examine the effect of pretesting on source memory. The task was the same as Experiment 1, except that targets were presented in black, the participants completed an interim maths task half-way through the encoding phase, and the source memory test involved recalling whether the targets were presented before or after the interim maths task As with Experiment 1, we anticipated a PTE on target recognition, but the crucial question of interest was whether pretesting would also improve source memory.

Method

The method was the same as in Experiment 1, except in the following respects.

Participants

Forty Psychology undergraduates from the University of Southampton (32 females, eight males), who were aged between 18 and 25 years ($M = 19.00$ years, $SD = 1.41$ years), completed the experiment for course credit. As in Experiment 1, the participants completed the experiment remotely, and the same exclusion criteria applied (although no participants were excluded). The *pwr* package indicated that, to obtain 80% power to replicate the target recognition PTE ($d_z = 1.25$) from Experiment 1 with an alpha of .05, we would require eight participants. However, to increase the generalisability of the results and the potential to detect a (perhaps substantially smaller) PTE in source memory, we recruited a much larger sample of 40 participants. The experiment was approved by the University of Southampton Psychology Ethics Committee.

Procedure

The encoding phase was the same as Experiment 1, except that all targets were presented in black. Half-way through the encoding phase, the participants completed an interim maths task, where they were presented with 10 simple maths equations (e.g., *(6 x 3) + 2 = 17)*. The participants had to indicate whether the answer provided was correct or not by means of a button-press (z = incorrect, m = correct). Response time was unlimited, and the participants completed the second half of the encoding phase upon completion. As in Experiment 1, the participants were instructed to attend to both the words and whether they were presented before or after the interim maths task. The test phase followed the same format as Experiment 1, except that the source memory question concerned whether the item was presented before or after the interim maths task, rather than the colour of the target.

Results

During the encoding phase, none of the participants correctly guessed any of the targets on Pretest trials. During the final test, the participants were once again very good at identifying the foils as novel (M = 91.75%, SD = 12.43%). Figure 1b shows the percentage of correct responses for Pretest and Read-only trials in the target recognition and source memory tests. In the target recognition test, the participants correctly recognised more Pretest targets as old than Read-only targets, $t(39)$ = 7.09, $p < .001$, d_z = 1.12, $BF_{10} > 100$. In the source memory test, overall accuracy (excluding the foils) was well above chance (M = 66.50%, SD = 9.14%), $t(39)$ = 11.41, $p < .001$, $BF_{10} > 100$, but the Pretest and Read-only conditions did not differ, $t(39)$ = 0.36, $p = .72$, d_z = 0.06, BF_{10} = 0.18.

As in Experiment 1, we also examined the size of the PTE for target recognition across trials in which the participants did and did not accurately recall the source. An Encoding condition (Pretest, Read-only) × Source accuracy (Correct, Incorrect) repeated-measures ANOVA revealed a significant main effect of encoding condition, $F(1, 39) = 35.70$, $p < .001$, $\eta_g^2 = .16$, $BF_{10} > 100$. The participants correctly recognised more targets from Pretest trials ($M = 86.32\%$, $SD = 20.66\%$) than Read-only trials ($M = 65.67\%$, $SD = 27.42\%$). There was also a significant main effect of source accuracy, $F(1, 39) = 11.08$, $p = .002$, $\eta_g^2 = .06$, $BF_{10} = 16.64$. The participants recognised significantly more targets on trials where they recalled the source correctly ($M = 81.98\%$, $SD = 19.87\%$) than trials when they did not ($M = 70.00\%$, $SD = 30.44\%$). However, no significant interaction was observed, $F(1, 39) = 0.68$, $p = .41$, $\eta_g^2 = .003$, $BF_{10} = 0.32$.

Discussion

The change of context manipulation, from colour to list membership, was successful. Overall, memory for source was higher in Experiment 2 (66%), than in Experiment 1 (55%). Despite this, however, the results largely replicated those seen in Experiment 1, with one exception that we discuss below.

For the analysis by experimental condition, Experiment 2 was entirely consistent with Experiment 1; pretesting boosted item recognition but had no impact on source accuracy. The pattern differed slightly for the second analysis, where we examined whether the PTE was affected by whether the participants correctly recalled the source. In both experiments, a main effect of encoding condition was observed, with participants demonstrating better recognition memory for targets from Pretest trials than Read-only trials. In Experiment 2 only, however, there was also a main effect of source accuracy,

such that the participants demonstrated better target recognition memory when their source decision was also correct. Thus, better target memory was accompanied by higher source memory. Crucially, however, both experiments revealed no significant interaction between source recollection status and the size of the PTE. Experiment 2 therefore confirms that pretesting boosts target recognition memory, but not memory for contextual information.

General Discussion

Two experiments examined the effect of pretesting on recognition of targets and their associated sources. In both experiments, pretesting produced clear improvements in target recognition memory, relative to Read-only trials. This pattern is consistent with previous observations using unfamiliar word definitions (Potts et al., 2019; Potts & Shanks, 2014; Seabrooke, Hollins, et al., 2019; Seabrooke, Mitchell, Wills, Inkster, et al., 2021), semantically unrelated word pairs (Seabrooke, Mitchell, Wills, & Hollins, 2021), and unrelated face-fact pairs (Seabrooke, Mitchell, et al., 2019). In tests of source memory, however, no effect of pretesting was observed, regardless of whether source was assessed for target colour (Experiment 1) or temporal context (Experiment 2). This failure to detect benefits of pretesting on source memory was seen in the context of good overall source memory (particularly in Experiment 2), suggesting that source memory was not simply at floor. Thus, while pretesting *does* improve recognition of targets, it does not appear to improve accessibility of contextual information that accompanies those targets. Our results demonstrate a dissociation between item and source memory.

Our choice of contextual information was based on Tulving's (1985) original conception of episodic memory as the mental reconstruction of a past event, experienced

as autonoetic consciousness. There is a long history of testing episodic memory either through self-report measures (e.g., the remember-know procedure; Gardiner et al., 1994), the process-dissociation procedure (Jacoby, 1991), or, as used here, through measures of context memory that demonstrate that the participant can reconstruct the past (e.g., Chan & McDermott, 2007). Although recollection can be supported by different aspects of source on different trials, there is clear evidence that these approaches measure largely the same concept (Chan & McDermott, 2007; Meiser & Bröder, 2002; Perfect et al., 1996). Thus, we argue that our source memory tasks reflect a broad view of episodic memory. Our source memory measures are not better or worse measures of episodic memory than those used by Metcalfe and Huelser (2020), but they are different, and they produce a different pattern of results. They therefore provide information for testing the theoretical accounts of the PTE.

The failure to observe generalised benefits of pretesting to source memory contradicts our broad interpretation of both the episodic recollection and attentional accounts of the PTE. However, a narrower version of the attentional theory, along the lines suggested by Mulligan (2004) for the generation effect, could potentially explain the current data pattern. Under this account, pretesting specifically increases attention to the *meaning* of feedback presented after a pretest, rather than to all the features of the episode. This processing account can explain both our and Metcalfe and Huelser's (2020) results. For the current data, greater attention to the meaning of the feedback boosts subsequent recognition of the target, but gives no information about the colour or temporal context of the target. In Metcalfe and Huelser's (2020) paradigm, by contrast, recollection of the participant's guess is directly related to the meaning of the target

presented as feedback, because the feedback provides information about the interpretation of the ambiguous homonym in the cue.

While we suggest that the current results may be interpreted within an application of Mulligan's processing account of the generation effect to the pretesting effect, we also acknowledge that there are some outstanding issues with this interpretation. Notably, Mulligan (2004, 2011) consistently observed negative generation effects for intrinsic source attributes such as target colour, where participants demonstrated better source memory for Read targets than Generate targets. This finding was explained by suggesting that generation increases conceptual processing of the target's meaning, while reading encourages processing of perceptual features such as the target's colour. No difference between the Generate and Read conditions was seen in memory for extrinsic source attributes, such as background colour or spatial location. Given these findings, it might be argued that, according to the Mulligan's processing account, we should have observed a negative pretesting effect for source memory, at least in Experiment 1. There are, however, potentially important methodological differences between the paradigms that are used in studies of the generation and pretesting effects. Perhaps most notably, in the generation effect, the target is generated by the participants (and therefore not presented in colour) on Generate trials. In our Experiment 1, by contrast, participants only generated errors on Pretest trials. The targets were always presented in colour on both Pretest and Read-only trials. It seems reasonable to anticipate that source memory would be particularly disadvantaged in Mulligan's Generate condition, because the full target was not presented in colour on those trials. When source memory for the cues (which were presented in colour on both Generate and Read trials) was examined, no difference

was observed between the two conditions for source memory (Mulligan, 2004,

Experiment 7).

It is worth noting that Mulligan (2004, 2011) did take steps to avoid a "nuisance"

account that attributes the negative generation effect for colour features to the lack of

colour on Generate trials. For example, a negative generation effect for target colour was

seen even when the letters of the Generate targets were replaced by blocks of colour

(Mulligan, 2004, Experiment 5). Perhaps the participants here focused their attention on

the targets that they had written for study, at the expense of the colours presented on

screen. This would explain why Mulligan (2004) observed a negative generation effect

on memory for colour attributes, while we observed a comparable null effect of

pretesting. This explanation is, however, less amenable to Mulligan's later observation of

a negative generation effect for target colours even when the participants were required to

write the targets in different colours (Mulligan, 2011, Experiment 5). One possible

explanation for this negative generation effect is that the targets were presented in colour

both on-screen and in their own writing on Read trials, but only in their writing on

Generate trials. If this is responsible for the negative generation effect that Mulligan

(2011) observed, then we might again expect the null source memory effect that we saw

in our Experiment 1, where target colour was matched between the Pretest and Read-only

trials. It is also possible that overall source accuracy was simply not good enough in our

Experiment 1, and that a stronger intrinsic source manipulation would reveal a negative

pretesting effect. In sum, we acknowledge that our results do not align exactly with the

results seen in the generation effect paradigm. We nevertheless suggest that Mulligan's

(2004) processing account can go some way to explaining our findings in the pretesting

effect, but further research would clearly help to elucidate some of the issues that we have noted here.

The idea that pretesting improves processing of the conceptual meaning of the target might seem at odds with the clear evidence that pretesting does *not* improve associative cue-target memory for semantically unrelated word pairs. Seabrooke, Hollins, et al. (2019), for example, showed that pretesting improves subsequent recognition of targets from unfamiliar, novel word pairs (e.g., Euskara-English translations), but not subsequent cued recall or associative recognition. Similar results were also seen with familiar but semantically unrelated word pairs such as *pond-spanner* (Seabrooke, Mitchell, Wills, & Hollins, 2021). However, focusing on the meaning of a target (perhaps in relation to the guess) is not the same as focusing on the association between the target and the original cue, as tested in associative recognition. It is entirely possible that a guess encourages participants to pay more attention to the meaning of the corrective feedback, rather than to the relationship between the feedback and the cue that elicited the original error.

In summary, the current results add to a growing pattern that shows that pretesting does not have a general effect, but rather is restricted to particular tests. For cued recall, pretesting boosts performance on related, but not unrelated, word pairs (Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012; Knight et al., 2012). For target recognition, by contrast, the PTE is robust for both related and unrelated word pairs (Seabrooke, Mitchell, Wills, & Hollins, 2021). Seabrooke, Hollins, et al. (2019) demonstrated the same dissociation between cued recall and target recognition for unrelated materials, but also demonstrated that pretesting does not improve associative recognition. We can now

add a further dissociation; pretesting boosts target recognition, but not source memory, unless the source judgement is based on the meaning of the materials under test (Metcalfe & Huelser, 2020). Collectively, these findings refute theories that predict simple quantitative effects of pretesting, which predict the PTE across a range of memory measures. Instead, the results are more compatible with the view that the PTE is highly specific, and that pretests improve conceptual processing of the feedback that follows an incorrect guess. This enhanced conceptual processing may then boost performance on subsequent criterion tests that tap conceptual processing of the target, but not those that examine other aspects of the study event.

Funding details

Disclosure Statement

The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement

The experimental programs and materials, trial-level data, and analysis scripts for both experiments are publicly archived at https://osf.io/kxgj5/.

References

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for

ANOVA. *Behavior Research Methods*, *44*(1), 158–175.

https://doi.org/10.3758/s13428-011-0123-7

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are

hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and

Cognition*, *27*(6), 1491–1494. https://doi.org/10.1037/0278-7393.27.6.1491

Carneiro, P., Lapa, A., & Finn, B. (2018). The effect of unsuccessful retrieval on

children's subsequent learning. *Journal of Experimental Child Psychology*, *166*,

400–420. https://doi.org/10.1016/j.jecp.2017.09.010

Champely, S. (2017). pwr: Basic functions for power analysis. *R Package Version 1.2-1*.

https://cran.r-project.org/package=pwr

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A

dual process account. *Journal of Experimental Psychology: Learning Memory and

Cognition*, *33*(2), 431–437. https://doi.org/10.1037/0278-7393.33.2.431

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments

in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

https://doi.org/10.3758/s13428-014-0458-y

Fazio, L., & 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

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with

criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of

Memory and Language*, *51*(3), 473–493. https://doi.org/10.1016/j.jml.2004.06.002

Gardiner, J. M., Gawlik, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, *1*(1), 107–110. https://doi.org/10.3758/BF03200764

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition*, *40*(4), 505–513. https://doi.org/10.3758/s13421-011-0174-0

Huelser, 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

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541. https://doi.org/10.1016/0749-596X(91)90025-F

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language*, *66*(4), 731–746. https://doi.org/10.1016/j.jml.2011.12.008

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

Kornell, N., Hays, M., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, *35*(4), 989–998. https://doi.org/10.1037/a0015729

Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies"

(JATOS): An easy solution for setup and management of web servers supporting

online studies. *PLoS ONE*, *10*(6), e0130834.

https://doi.org/10.1371/journal.pone.0130834 Le

Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information.

Journal of Experimental Psychology: Learning Memory and Cognition, *28*(1), 116–

137. https://doi.org/10.1037/0278-7393.28.1.116

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic

recollection rather than semantic mediation. *Neuropsychologia*, *138*.

https://doi.org/10.1016/j.neuropsychologia.2019.107296

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *Package 'BayesFactor.'*

https://richarddmorey.github.io/BayesFactor/

Mulligan, N. W. (2004). Generation and memory for contextual detail. *Journal of*

Experimental Psychology: Learning Memory and Cognition, *30*(4), 838–855.

https://doi.org/10.1037/0278-7393.30.4.838

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not

extrinsic context. *Quarterly Journal of Experimental Psychology*, *64*(8), 1543–1562.

https://doi.org/10.1080/17470218.2011.562980

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context

discriminate recollection from familiarity in recognition memory? *Quarterly Journal*

of Experimental Psychology Section A: Human Experimental Psychology, *49*(3),

797–813. https://doi.org/10.1080/713755644

Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during

 learning: What is the locus of the effect? *Journal of Experimental Psychology:*

 Learning Memory and Cognition, *45*(6), 1023–1041.

 https://doi.org/10.1037/xlm0000637

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning.

 Journal of Experimental Psychology: General, *143*(2), 644–667.

 https://doi.org/10.1017/CBO9781107415324.004

R Core Team. (2021). R: A language and environment for statistical computing. In *R*

 Foundation for Statistical Computing. https://www.r-project.org/

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-

 term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.

 https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L. I., & 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

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic

 review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463.

 https://doi.org/10.1037/a0037559

Seabrooke, T., Hollins, T. J., Kent, C., Wills, A. J., & Mitchell, C. J. (2019). Learning

 from failure: Errorful generation improves memory for items, not associations.

 Journal of Memory and Language, *104*, 70–82.

 https://doi.org/10.1016/j.jml.2018.10.001

Seabrooke, T., Mitchell, C. J., Wills, A. J., & Hollins, T. J. (2021). Pretesting boosts

recognition, but not cued recall, of targets from unrelated word pairs. *Psychonomic Bulletin & Review*, *28*, 268–273. https://doi.org/10.3758/s13423-020-01810-y

Seabrooke, T., Mitchell, C. J., Wills, A. J., Inkster, A. B., & Hollins, T. J. (2021). The benefits of impossible tests: Assessing the role of error-correction in the pretesting effect. *Memory & Cognition*. https://doi.org/10.3758/s13421-021-01218-6

Seabrooke, T., Mitchell, C. J., Wills, A. J., Waters, J. L., & Hollins, T. J. (2019). Selective effects of errorful generation on recognition memory: The role of motivation and surprise. *Memory*, *27*(9), 1250–1262. https://doi.org/10.1080/09658211.2019.1647247

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. I. (2008). Testing during study insulates against the buildup of proactive inteference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392–1399. https://doi.org/10.1037/a0014896

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review*, *19*(5), 899–905. https://doi.org/10.3758/s13423-012-0276-0

Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition*, *42*(8), 1373–1383. https://doi.org/10.3758/s13421-014-0454-6

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ Science of Learning*, *3*(8).

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

Zawadzka, K., & Hanczakowski, M. (2019). Two routes to memory benefits of guesing.

Journal of Experimental Psychology: Learning Memory and Cognition, *45*(10),
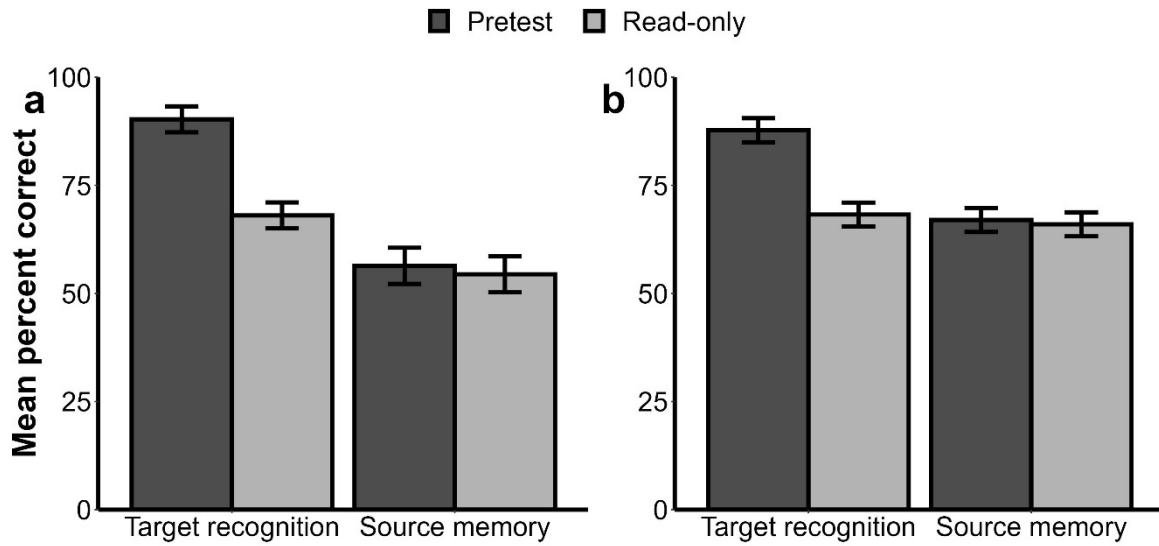
1748–1760. https://doi.org/10.1037/xlm0000676

Figure 1

Figure 1. Mean percentage of correct responses in the target recognition and source

memory tests of **(a)** Experiment 1 and **(b)** Experiment 2. Error bars represent difference-

adjusted, within-subject, 95% confidence intervals (Baguley, 2012).