The term "testing effect" refers to the finding that, when it comes to long-term retention of a piece of information, retrieving it from memory trumps restudying it (**???????**). Besides directly enhancing retention through repetition of *successful* retrieval, testing effects can be brought about indirectly (**??**; but for a different view, see **?**). For example, *unsuccessful* retrieval attempts can, through subsequent repeated encoding, generate test-potentiated (re)learning (TPL), whereby taking more tests increases the proportion of *newly* retrieved items in a test immediately following a restudy episode (**??????**).

More importantly, following an initial impetus provided by **?**, who built upon earlier findings (**??**), a decade of research has shown that retrieving previously studied information can even facilitate the acquisition of *new* information (**???**). If, within a multi-list learning paradigm, each subsequent study episode contains new items, testing the memory for those items after each learning episode still yields a greater number of correct responses and a decrease in proactive interference (PI) on a test pertaining to the final learning episode (e.g. **???**). Following the reasoning of **?**, we use the term "test-potentiated *new* learning" (TPNL) to denote this effect. Having been investigated mainly using simple materials, a particularly important question for real-world applications is whether TPNL generalises to materials more complex than single words and word pairs. The research conducted in the preceding decade mostly points to a positive answer.

? were the first to ask whether the scope of TPNL extends to prose passages, and they did find a consistent effect using a free recall measure throughout their five experiments. As for its possible cause, evidence for an improvement via integration of related information or via PI reduction — an explanation which figures prominently in the multi-list paradigm literature (**??**) — was rather unconvincing. The effect was observed regardless of the interrelatedness of passages, and, possibly owing to the usage of text passages and free recall, overall intrusion rates were quite low. The authors subscribe to a view that retrieval attempts may promote the use of more effective encoding strategies.

In a similar study using text passages, **?** hypothesised that retrieval attempts enhance context fluctuations, which may lead to TPNL. Context changes have mostly been studied in terms of their role in *directed forgetting*, where it has been proposed that they could reduce

interference (e.g. **?**). Since people store contextual alongside the target information during learning (**???**), the interplay between the two may affect memory performance. Regarding TPNL, it has been proposed that testing causes an internal context change relative to the one during study (**???**). Hence, items from successive study episodes might be updated with unique contextual information related to retrieval (**??**), while those encountered last remain associated only with the study context. For example, while knowing that each study episode contains new items, being aware that a recalled item was actually retrieved on a prior test could prevent subjects from providing erroneous responses, which may underlie the PI reduction observed in TPNL (**??**). In other words, additional contextual cues increase the disparity between pre- and post-retrieval items, providing an advantage to the latter by constraining the memory search set to items that are associated exclusively with the study context (**??**). Apart from this retrieval perspective on the effect of internal context changes, an encoding explanation proposes that they also induce a reset of encoding processes, making subsequent encoding more effective However, if this scenario were true, then perhaps *any* interpolated task causing an internal context change might also produce TPNL.

Nonepisodic recall

One of the more curious findings in the field is that TPNL may indeed ensue not only after retrieving the previously studied material (episodic retrieval), but also after the retrieval of information unrelated to the studied material from semantic (**??**), or even short-term memory (**?**), although there have been unsuccessful attempts at replication (e.g. **?**).

? let their participants learn five lists of 20 words while engaging in varied interlist activities. They either restudied the lists, recalled the words from the list, generated as many words as they could from one of four semantic categories (e.g. professions), engaged in a 2-back short-term memory task, or counted backwards from a random three-digit number. They found that only the three forms of retrieval induced TPNL. In their first experiment, **?** adapted the procedure from **?**. Participants learned five lists of 10 words, between which they either retrieved information from semantic memory (e.g. enumerated professions) or counted backwards by 3s from a three-digit number. Only semantic retrieval enhanced performance

in final list recall. They replicated and extended these findings in their second experiment by using complex learning materials. Participants read four text passages in a self-paced manner, while engaging in one of the same two types of intervening tasks between reading.

The argument these two groups of authors invoke to explain their results is that nonepisodic retrieval tasks sufficiently alter participant's internal context, and because the last study session is not affected by an additional context shift, a beneficial segregation of the final study context from the previous ones is produced. An additional prediction stemming from the context change hypothesis is that introducing a delay between the last study episode and the final test should annul the benefits conferred by the contextual segregation because the delay offers ample opportunity for a context change even without the retrieval (**?**). **?** confirmed this in their final experiment, but **?** provide evidence to the contrary, and side with the strategy change explanation suggested by **?**. **?**, however, did not include any other supposed context-changing interpolated tasks besides episodic retrieval, which prevented assessing other predictions of the hypothesis.

Although **?** did find a reduction in PI in their first experiment, which they had predicted based on the purported isolating effect of context fluctuations, when their methodology precluded such effects in their second experiment, the results, resonating with those of **?**, still showed TPNL. They conceded that PI reduction may not be the sole basis for the observed enhanced performance, and mention optimisation of encoding strategy as a viable cause. Finally, it is worth noting that, apart from prose passages, positive evidence for TPNL with complex learning materials was also found using lengthy video materials **???**, but in order to clarify the nature of TPNL within realistic settings, further inquiry is warranted.

Feedback

Even though it is not necessary for the testing effect to emerge (**?**), corrective feedback is known to augment it (**?**). When it is provided, corrective feedback increases learning because it promotes error correction (**?**) and reassurance (**?**). Feedback is particularly important for recognition tests such as multiple-choice tests since the usual benefit testing provides might turn into a disadvantage in case the test-taker selects a lure (**??**). Further, evidence points to

the timing of feedback being a relevant variable when gauging its influence on learning, with delayed feedback given in bulk showing superior effects compared to immediate, piecemeal feedback (**????**).

Perhaps unsurprisingly, there is a paucity of research into the potential role of feedback in TPNL because the process by which it could affect new learning is less apparent. One possibility is that feedback may act indirectly by guiding future efforts and motivating a change in encoding strategy (**?**). Indeed, even failed or unconfident retrieval can be conceptualised as feedback per se because it can provide insight into where additional cognitive resources should be allocated to improve performance (see e.g., **?**). Moreover, providing feedback may engender additional processing (**?**), further enhancing context fluctuations, and perhaps increasing the likelihood of intrusions during new learning (**?**). Thus, the variable of corrective feedback may be a fruitful avenue for research.

Present study

Our study had two main goals. Firstly, we sought to replicate the TPNL effect in an ecologically valid setting, by using complex learning materials and standard multiple-choice items. Secondly, there is a relative dearth of investigations using nonepisodic retrieval and recognition, and furthermore a lack of studies introducing corrective feedback (**?**). We therefore formed two memory tests, one of which tapped into episodic (assessing memory of the studied materials) while the other tapped into semantic memory (assessing general knowledge). Participants either were or were not given feedback upon completing an interpolated activity episode.

Even though it has been shown that, in the standard multi-list procedure, substantially larger effect sizes follow after using free recall rather than recognition-level retrieval (**?**), choosing to examine the impact of feedback on TPNL imposed constraints upon our choice of testing format; immediate provision of feedback would have been intractable had we chosen to use free recall. Recognition is thought to be based more on familiarity than on controlled memory search (**?**), which could be the reason behind the discrepant effect sizes. We, therefore, made an effort to heed the advice provided by **?** — constraining the time

for encoding, increasing the similarity between the correct answer and the distractors, and providing subjects with ample time to respond to questions — in order to encourage the use of controlled recollection. While relevant studies did find PI to be superfluous when complex learning materials are used, in order to explore this possibility more closely, we used multiple-choice questions designed to assess memory both in terms of correct answers and susceptibility to intrusions.

Based on the preceding discussion, if TPNL is mediated by context fluctuations, then we should find that both types of retrieval enhance new learning, whereas rereading, which is deemed unable to bring about context changes (**?**), does not. If PI is in fact nonessential, then we should observe the expected memory boost without necessarily seeing an accompanying decrease in PI[1]. We assumed that presenting feedback would have a positive effect on memory performance, but only for the participants engaging in episodic recall because only they could alter their encoding strategies accordingly. Finally, we expected to find an interaction effect of activity type and feedback presentation on the number of intrusions, but did not set a specific prediction regarding its pattern.

[1]Our original hypothesis suggested a differential effect of activity type on PI rates. Specifically, we expected that participants engaging in episodic retrieval would display the lowest susceptibility to PI, followed by participants in the semantic retrieval condition, and finally by those tasked with rereading. Upon further examination of the literature, our expectations changed, but this had no bearing on the analyses, which remained the same nevertheless.