english

The term "testing effect" refers to the finding that, when it comes to long-term retention of a piece of information, retrieving that information from memory trumps restudying it (**???????**). It is generally recognised that testing can have two types of effects — *direct* and *indirect* (**??**). Direct effects refer specifically to the increased retention that ensues from repeatedly *successfully* retrieving the target information — a process which is, presumably, reactivated at the time of a later test. A typical demonstration of the testing effect entails an initial learning phase, followed by a period during which participants either restudy the same material, engage in a memory test involving the studied material, or are not exposed to the original material at all. Finally, after a retention interval, an additional memory test reveals that the group subjected to a memory test during the intervening period has a distinct advantage over the other two groups.

On the other hand, indirect effects are brought about by some other process or processes besides the act of taking the test (**?**; but for a different view, see **?**). For example, *unsuccessful* retrieval attempts (which are not followed by feedback) can, through subsequent repeated encoding, also generate a testing effect, namely *test-potentiated (re)learning* (**?????**). With the aim of disentagling test-potentiated (re)learning from the testing effect, **?** let participants learn a list of 25 word pairs. One group completed nine cycles comprising a single test and a restudy session, while another completed three cycles comprising five tests and a restudy session. In order to isolate the effect of unsuccessful from successful retrieval attempts, they focused their analysis only on words not recalled on a test preceding a restudy episode. Results showed that, compared to taking fewer tests, taking more tests produces a greater increase in the proportion of *newly* retrieved items (i.e. words that were not retrieved on a pretest) in a test immediately following a restudy episode — a veritable potentiation of learning.

Test-potentiated new learning

Juxtaposed to the well established finding that attempting to recall studied material compared to restudying it, facilitates the long-term retention of *that* material, stand the results of a

decade of research showing that retrieving previously studied information can even facilitate the acquisition of *new* information (**??**). If each additional study episode in the paradigm used to demonstrate test-potentiated learning contains *new* materials (giving a now standard blocked design; **?**), one still observes that testing the memory of those new materials after each learning episode, compared to restudying the same materials, yields a greater number of correct responses and a decrease of proactive interference on a test administered to all subjects after the final learning episode (e.g. **???**). Following the reasoning of **?**, in this paper we will use the term "test-potentiated *new* learning" (TPNL) to denote this effect.

In one of the earliest studies showing the effect of TPNL, **?** observed that, when recalling studied lists of words, participants systematically produce more prior-list intrusions when probed for a given list, if their memory of a prior list had not been tested before they proceeded to study the given list. These findings were corroborated by **?**, who used an AB-AC interference paradigm, and found that omitting a test following the study of the AB list had a deleterious effect on learning AC items. Building on these results, **?** conducted a study using a blocked design wherein they told their subjects to study five lists of items in anticipation of a final cumulative test. All subjects were tested immediately after studying the final list, but they engaged in different intermittent activities between studying the first four lists. One group was tested on each list after studying it, another group restudied each list, and a third group completed a mathematical distractor task. Participants whose memory was tested after each list produced more correct responses and fewer prior-list intrusions on the immediate test administered after studying the last list. The authors explained the found benefit of testing in terms of a segregation mechanism that prevents overburdening of retrieval cues, which, in the absence of testing, causes a build-up of proactive interference. The following decade has seen a renewal of interest in TPNL (**???**), with studies mainly using the multilist learning paradigm to delineate the scope of the effect with respect to various moderating variables: the relationship between materials to be studied in successive study episodes, varieties of study designs (blocked vs. interleaved), and populations, to name a few.

A particularly important question for our study and for real-world applications is whether these results generalise to materials more complex than word lists, and research conducted

in the preceding decade mostly points to a positive answer (**????**). **?** were the first to demonstrate TPNL with complex textual materials. In a series of five experiments, they presented their participants with prose passages, and showed that interpolating study episodes with attempts at free recall of the studied materials increased performance relative to a no-study condition and a condition where the interpolated activity involved solving mathematical equations. Interestingly, their results provided compelling evidence that release from proactive interference may not be the most appropriate explanation for TPNL, at least when it comes to complex learning materials. Extending these results to a more contemporary eduactional setting, **?** showed that TPNL occurs when the learning materials are online video lectures, and inclined towards a source monitoring account of the observed effect (**?**). In the following section, we provide a short overview of heretofore proposed accounts of TPNL.

Theoretical overview

Recently, **?** provided a meta-analytic analysis and comprehensive overview of the literature, identifying four *nonconflicting* theoretical frameworks which were put forth throughout the years as viable explanations for TPNL. *Resource theories* generally posit that testing increases cognitive resources, but they propose different mechanisms by which this is achieved: (i) proactive interference reduction (e.g. **????**), (ii) restoration of encoding/attentional resources (e.g. **?**), or (iii) alteration of mind wandering patterns (e.g. **???**). Whereas resource theories focus on the amount of deployable cognitive resources, *metacognitive theories* emphasise the optimisation of encoding strategies induced by retrieval attempts (e.g. **??**). For example, in a recent investigation, **?** found that, compared to untested groups, the group whose memory for the first three word lists was subjected to interpolated testing displayed superior semantic organisation across lists. These findigs reflect a similar pattern obtained for the testing effect, where a greater number of tests is associated with improved organisation of output displayed upon testing (**??**).

The key idea underlying the third framework — *context theories* — is that, apart from storing the studied information themselves, people store the related contextual information as well (e.g. **?**). Afterwards, the accessibility of this contextual information can affect the

likelihood of successful retrieval of target information. Furthermore, the claim is that, unlike restudying, attempting retrieval causes an internal context change relative to the study context (**??**), and recalled items may be updated with contextual information from the retrieval attempt, while newly encountered information is still associated only with the study context. Therefore, recalling new-learning items is limited to only those items associated exclusively with the study context, providing them with the advantage observed upon testing. While this circumscription of separate learning episodes is at the core of both resource and context accounts, its effect on learning is supposedly different. According to the former, isolating a learning episode through attempts at recall increases resources for subsequent learning by preventing *encoding-based* proactive interference. On the other hand, the latter place the emphasis on later *retrieval* processes, whereby isolating an earlier learning episode reduces the memory search set for retrieval.

Finally, *integration theories* advance the notion that interpolated testing facilitates the integration of the new-learning material either with its retrieval cues or with the original-learning material. On one account, testing increases the likelihood of spontaneous covert retrieval of original-learning items during the study of new items, fostering their integration, thereby increasing conceptual organisation (e.g. **?**) and the effectiveness of retrieval cues (**?**). For example, **?** found that interpolated testing increased the clustering of related information that is acquired across different segments within a video-recorded lecture.

Nonepisodic recall

One of the more curious findings in the field is that TPNL can arise not only after retrieving the previously studied material (episodic retrieval), but also after retrieval of information unrelated to the studied material from semantic memory (**??**), or from 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 all three forms of retrieval induced TPNL. In their first experiment, **?** adapted the procedure from **?**, using only the semantic generation and distractor (counting backwards) tasks, and found that interleaved semantic retrieval enhanced performance for final list recall. They replicated and extended these findings in their second experiment by using complex learning materials: lists of words were replaced by texts related to animals, while learning was evaluated with short-answer and multiple-choice questions.

The argument these two groups of authors invoke to explain their results is that nonepisodic retrieval tasks sufficiently alter participant's internal context. 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, which reduces the memory search set. **?** conducted a metaregression, comparing the explanatory power of the four frameworks described above. Summarising their results, they highlighted resource and integration theories as accounts which have thus far garnered more empirical support, giving a slight upper hand to integration theories, while stating that context theories are least supported by extant research. Therefore, we opted to align our study design with the goal of comparing resource and integration frameworks.

Feedback

Although corrective feedback is known to augment the testing effect (**?**), there is a paucity of research into the effect of feedback on TPNL, especially when considering studies that have implemented the blocked design. Feedback is particularly important for recognition test such as multiple-choice tests since the usual benefit testing confers might turn into a disadvantage in case the test-taker selects a lure (**??**). Moreover, evidence points to the timing of feedback being a relevant variable when gauging its influence on learning, with delayed feedback showing superior effects compared to immediate feedback (**????**). For example, participants in a study by **?** read prose passages and then either took or did not take an initial multiple-choice test. If they took the test, corrective feedback was either not given, given immediately after each answer was provided, or given in bulk after the entire test. A final test administered one week after the initial test revealed that, relative to studying,

(1) taking an initial test alone tripled the success rate on the final test (22% performance increase), (2) giving immediate feedback on the initial test increased performance by 32%, but (3) that delayed feedback increased performance by 43%.

The variable of corrective feedback may be a fruitful avenue for research because resource and integration theories provide conflicting predictions regarding its effects on TPNL (**?**). Providing corrective feedback should increase the likelihood of intrusions during new learning, which are deemed beneficial from the standpoint of integration theories, but detrimental from the point of view of resource theories. Thus, feedback should reduce TPNL according to resource theories, but increase it according to integration theories.

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, guided by the analysis of gaps in the field provided by **?**, who identified a relative dearth of studies using nonepisodic retrieval and recognition (e.g. multiple-choice items) for the interpolated activity, and furthermore a lack of studies introducing feedback in a blocked study design, we attempted to expand the existing body of literature by employing a novel combination of variables, in order to examine their effects and interactions in the context of TPNL.

In particular, we assumed that retrieval could be the active component in interpolated activities that have been shown to give rise to TPNL. To test this, apart from using rereading as a control comparison task, we formed two memory tests, one of which tapped into episodic (assessing memory of the studied materials) while the other tapped into semantic (i.e. nonepisodic) memory (assessing general knowledge). Following the reasoning of **?**, in order to pit integration and resources accounts of TPNL against each other, participants either were or were not given feedback upon completing an interpolated activity episode. Bearing in mind the necessity of systematically examining the impact of proactive interference on participants' performance, we used multiple-choice questions designed to assess memory both in terms of correct answers and susceptibility to intrusions.

Based on the preceding discussion, we predicted that, compared to the rereading group, participants in the retrieval groups would have a significantly higher average total score on the final test. Furthermore, we expected to find no difference between the two groups having different types of retrieval. With regards to proactive interference, we expected that participants engaging in episodic retrieval would display the lowest susceptibility to proactive interference, followed by participants in the semantic retrieval condition, and finally by the participants in the rereading condition. Looking at the two retrieval groups, we expected to find a significant main effect of feedback on the average number of correctly answered questions, as well as an interaction effect between feedback presentation and type of interpolated activity. Specifically, we assumed that presenting feedback would have a positive effect on the average number of correctly answered questions, but only for the participants in the content-related test condition. We also predicted that participants receiving feedback would have a significantly higher average number of intrusions than participants receiving no feedback. 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.