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 — a process which is, presumably, reactivated at the time of a later test — 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 **??????**).

After 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 each subsequent study episode in the paradigm used to demonstrate TPL contains new materials (giving a now standard blocked design; **?**), one still observes that testing the memory for those new materials after each learning episode yields a greater number of correct responses and a decrease of proactive interference (PI) on a test administered to all subjects after the final learning episode (e.g. **???**). Following the reasoning of **?**, we use the term "test-potentiated *new* learning" (TPNL) to denote this effect. With studies mainly using the multilist learning paradigm to delineate the scope of TPNL, a particularly important question 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 (prose passages: **??**; video lectures: **??**).

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.Summarising the results of their metaregression, **?** 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.

Nonepisodic recall and feedback

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 (**??**), or short-term memory (**?**), although there have been unsuccessful attempts at replication (e.g. **?**). **?** demonstrated this using lists of words, while **?** replicated and extended these findings using prose passages.

Although corrective feedback is known to augment the testing effect (**?**), there is a paucity of research into the effect of feedback on TPNL. Feedback is particularly important for recognition tests 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 given in bulk showing superior effects compared to immediate, piecemeal feedback (**????**). 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 according to resource 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. Even though it has been shown that, in the standard TPNL 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. We used multiple-choice questions designed to assess memory both in terms

of correct answers and susceptibility to intrusions. Secondly, there is a relative dearth of investigations using nonepisodic retrieval and recognition, and furthermore a lack of studies introducing feedback in a blocked study design (**?**). 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.

Based on the preceding discussion, we predicted that participants in the retrieval groups would display TPNL, whereas a control rereading group would not. 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 in the rereading condition. We assumed that presenting feedback would have a positive effect on memory performance, but only for the participants engaging in episodic recall. We also predicted receiving feedback would significantly increase interference. 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.