Journal of Experimental Psychology:
Learning, Memory, and Cognition
2012, Vol. 38, No. 6, 1792–1800

0278-7393/12/$12.00 DOI: 10.1037/a0028460

RESEARCH REPORT

Inoculation Against Forgetting: Advantages of Immediate Versus Delayed Initial Testing Due to Superior Verbatim Accessibility

Ainat Pansky
University of Haifa

In this study, potential benefits of early memory testing were examined in terms of "inoculating" eyewitness memory against forgetting. As predicted by fuzzy trace theory (e.g., Reyna & Titcomb, 1997), a larger testing advantage in the delayed recall of event details was expected after immediate testing than after delayed testing because of the decline in accessibility of verbatim traces over time. However, memory for only the gist of these details was expected to be relatively stable over time, resulting in a smaller (if any) effect of the timing of interpolated testing. After viewing a target event, participants were questioned about event items immediately, after a 24-hr delay, or after a 48-hr delay and were free to respond at either the gist or the verbatim level. Verbatim memory for event details was tested 72 hr after the event. As expected, immediate interpolated testing improved verbatim memory performance on the final test more than delayed testing did, yielding a larger testing effect. Furthermore, the effect of the timing of interpolated testing on the magnitude of the testing effect was mediated by verbatim accessibility at interpolated testing. In contrast, memory for only the gist of event details was unaffected by the timing of interpolated testing, both on the interpolated test and on the final test. The findings highlight the role of declining verbatim memory over time in accounting for the advantage of immediate over delayed interpolated testing in inoculating eyewitness memory against forgetting of detailed information.

Keywords: eyewitness memory, testing effect, repeated recall, memory over time, forgetting

Extensive research has documented conditions in which questioning eyewitnesses after exposure to an event can have deleterious effects on subsequent memory performance (see, e.g., Loftus & Palmer, 1974). However, it has also been suggested that under certain conditions, memory testing soon after an event may have beneficial effects of "inoculating" eyewitness memory against forgetting (e.g., Brainerd & Ornstein, 1991). This notion is consistent with a remarkable amount of research demonstrating that taking a memory test improves subsequent retrieval of the tested information, in what is known as the *testing effect* (for a recent review, see Roediger & Butler, 2011), supporting Bjork's (1975) claim that the act of recollection not only retrieves stored information but also modifies the memory representations of that information. Several studies that have examined the effects of testing in simulated eyewitness situations have confirmed that initial memory testing can inoculate eyewitness memory against forgetting (e.g., Chan & Langley, 2011; Gabbert, Hope, & Fisher, 2009; Pansky & Tenenboim, 2011; Poole & White, 1991).

In the present study, I examine how inoculation against forgetting is affected by the timing of initial memory testing. According to fuzzy trace theory (FTT; e.g., Brainerd & Reyna, 1993), each event item is encoded in memory at various levels of precision, from verbatim traces representing detailed episodic information to gist traces capturing its meaning. Over time, verbatim traces become inaccessible more rapidly than gist traces (e.g., Brainerd & Reyna, 1998). Therefore, FTT predicts that, compared with delayed interpolated testing, earlier interpolated testing is more likely to refurbish verbatim traces before they become inaccessible and is more likely to counteract their disintegration over time, yielding stronger inoculation against forgetting of these details (e.g., Reyna & Titcomb, 1997).

Two previous studies in which memory for prose passages was examined have found this expected trend of a larger benefit of earlier compared with later interpolated testing for correct recall on a delayed final test (Bergman & Roediger, 1999; Spitzer, 1939). However, these studies used repeated testing, such that earlier interpolated testing incorporated more interpolated tests than later interpolated testing, which may have accounted for at least some of the advantage observed for earlier testing. In the present study, the effect of the timing of a single interpolated test on subsequent memory performance is examined.

In the majority of studies that implemented interpolated testing, the participants were required to report the test items exactly as they were presented at study (i.e., verbatim responding). However, in real-life situations, eyewitnesses usually have control over the level of generality in which they choose to report event information. Several studies have shown that when rememberers are

This article was published Online First May 14, 2012.

I thank Einat Tenenboim and Adi Ben-Nun Inbar for their valuable assistance at various stages of this research.

Correspondence concerning this article should be addressed to Ainat Pansky, Department of Psychology, University of Haifa, Haifa 31905, Israel. E-mail: pansky@research.haifa.ac.il

allowed to control the *grain size* (i.e., precision or coarseness) of their responses, they respond at a level of generality at which they are likely to be correct (e.g., Goldsmith, Koriat, & Pansky, 2005; Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Weber & Brewer, 2008). Goldsmith et al. (2005) have also demonstrated that, over time, rememberers coarsen the grain size of their answers, resulting in a smaller proportion of fine-grain answers as retention interval increases. The choice of grain size was found to be determined primarily by one's confidence in the fine-grained answer. Previous findings indicate that such confidence is largely based on the amount and quality of the accessible information in memory and the ease with which it comes to mind (see, e.g., Kelley & Lindsay, 1993; Koriat, 1993). Thus, the answers that one freely chooses to provide at a precise grain size can be seen to reflect fine-grained information that comes to mind and is assumed to be correct.

By the same token, in the present study, interim verbatim accessibility was defined in terms of the ease of access with which the target verbatim information comes to mind at interpolated testing, as reflected in the proportion of accurate verbatim information that is freely provided by the participants (see also Koriat, 1993). The expected decline in such free-level verbatim accessibility over time was predicted, in turn, to result in decreasing benefits of interpolated testing for subsequent verbatim recall, with increasing delay between the target event and the timing of interpolated testing.

The participants first viewed a slide sequence containing target details and were initially tested either immediately or after a delay of either 24 or 48 hr. The initial test contained questions about half of the target items, with the remaining items serving as control untested items. On this test, the participants were free to respond at either the gist or the verbatim level by recalling each item at either the basic level or the subordinate level, respectively (see, e.g., Brainerd & Reyna, 1998).[1] Finally, all of the participants were tested 72 hr after exposure to the event for their verbatim memory of the event details. This retention interval was held constant across the three interpolated-testing groups to equate baseline performance in the untested condition of the three groups against which the testing effects could be compared. Given the large variation in the retention intervals used in previous studies with similar eyewitness-type materials (although, typically, in the order of days), the retention intervals of the present study were chosen on the basis of preliminary testing with the specific experimental materials that were used.[2]

On the basis of the findings of Goldsmith et al. (2005), both the proportion of responses provided at the verbatim level (regardless of their accuracy) and the proportion of accurate verbatim responses provided at initial testing were expected to be higher the earlier such testing was conducted. Consistent with the findings of Spitzer (1939) and Bergman and Roediger (1999) and as derived from FTT (e.g., Reyna & Titcomb, 1997), a larger testing effect on verbatim recall was expected the earlier interpolated testing was performed. Furthermore, if the effect of the timing of interpolated testing on the magnitude of the testing advantage is due to the declining accessibility of verbatim traces over time, this effect should be mediated by the proportion of accurate verbatim responses provided on the interpolated test.

In contrast, because of the superior stability of gist over time (e.g., Brainerd & Reyna, 1998), memory for gist was expected to show a different pattern of results. First, the proportion of accurate responses at the gist (but not verbatim) level on the interpolated test was expected to be less affected (or even unaffected) by the timing of the test. Second, the magnitude of the testing effect in terms of accurate recall at only the gist level (on the final test) was expected to be less affected (or even unaffected) by the timing of interpolated testing.

Method

Participants

The participants were 96 Hebrew-speaking undergraduates at the University of Haifa, who took part in the experiment for a payment of 70 NIS (approximately US$18) or course credit.

Materials

The target event was a 6.5-min narrated slide show about a day in a female student's life, containing 33 slides. Sixteen concrete items (e.g., WOODEN CHAIR, POODLE), each presented visually on a separate slide, constituted the target items (see Appendix A, column 2). The interpolated test consisted of cued-recall questions about half of the target items, whereas the final test consisted of cued-recall questions about all of them (see Appendix A, column 3).

Procedure

In the first stage of the experiment, the participants viewed the slide show, which was followed by a 10-min nonverbal filler task. The *immediate-testing* group then performed the interpolated test, whereas the *delayed-testing* groups performed this test after a delay of either 24 or 48 hr. The interpolated test contained eight cued-recall questions, each about one of the eight target items assigned to the tested condition. The participants were requested to answer all of the questions, to avoid answers like "don't remember," and to provide the best answer that they could. Note, though, that at this stage they were free to choose the level of detail of their responding. The remaining eight target items constituted the untested items. The assignment of items to experimental condition was counterbalanced across participants. The order of the questions corresponded to the chronological order in which the items had appeared in the slide show.

After a retention interval of 72 hr (since viewing the slide show), all of the participants performed the final cued-recall test on all of the target items. For each item, the same question that was presented on the interpolated test was followed by a second question, designed to solicit a response at a more detailed level. Thus, after

[1] Thus, recalling a test item at the basic level (e.g., DOG) would constitute a gist response, whereas recalling it at the subordinate level (e.g., POODLE) would constitute a verbatim response.

[2] Preliminary testing using the format of the final test indicated that memory performance (i.e., the proportion of accurate verbatim responses) declined from .57 when tested 10 min after the event, to .49 after 24 hr, to .38 after 48 hr, and to .28 after 72 hr. On the basis of these data, interpolated testing after delays of 10 min, 24 hr, and 48 hr all had the potential of yielding a testing effect on the delayed final test.

answering an initial question about an event item (e.g., "What was in the oven . . . ?"), the participant was asked a second question about that item, with his or her response to the first question (e.g., PIZZA) inserted in the second question (e.g., "What kind of *pizza*?"). An example was provided in the instructions to give the participants an idea of the level of specificity that was expected in each phase. The questions were presented in chronological order and the participants were required to answer all of them. See Appendix B for the instructions that were presented to the participants at the various stages of the experiment.

Results

For each response that was provided on the interpolated test, two independent judges determined whether it was reported at the verbatim level (e.g., POODLE), reported at the gist level (e.g., DOG), or constituted an invalid response (e.g., "don't remember"), and whether it was accurate (a) at the verbatim level and (b) at the gist level. They also determined for each response provided on the final cued-recall test whether it was accurate (a) at the (requested) verbatim level and (b) at the gist level. The classifications made by these two judges were identical in 99% of the cases. A third judge determined the scoring of the controversial 1% of the responses. Analyses of variance (ANOVAs) and planned comparisons (*t* tests) were used for statistical testing, except where noted.

Interpolated Responses

The proportion of invalid responses provided on the interpolated test at neither the verbatim nor the gist level did not vary with the timing of interpolated testing (.02 for all three groups), $F(2, 93) = 0.17$, $p = .84$, $\eta_p^2 = .004$. However, the timing of interpolated testing had the predicted effect on the proportion of responses provided at the verbatim level, $F(2, 93) = 6.24$, $p = .003$, $\eta_p^2 = .12$. As shown in Figure 1A, the proportion of verbatim-level responses declined from .43 when interpolated testing was conducted immediately to .28 when it was delayed by 24 hr, $t(62) = 2.58$, $p = .01$, $d = 0.64$, with an additional nonsignificant decline to .25 when interpolated testing was delayed by 48 hr, $t(62) = 0.62$, $p = .54$, $d = 0.15$.

Not surprisingly, as shown in Figure 1B, the same pattern emerged for the proportion of responses that were accurate at the verbatim level, $F(2, 93) = 11.19$, $p < .001$, $\eta_p^2 = .19$, dropping from .37 when interpolated testing was conducted immediately to .20 when it was delayed by 24 hr, $t(62) = 3.27$, $p = .002$, $d = 0.83$, with an additional nonsignificant reduction (to .16) when interpolated testing was delayed by 48 hr, $t(62) = 1.05$, $p = .297$, $d = 0.27$. In contrast, the proportion of responses that were accurate only at the gist level (.48) was not affected by the timing of the test, $F(2, 93) = 0.03$, $p = .973$, $\eta_p^2 = .001$.[3]

To summarize the findings obtained for the interpolated test, both the total proportion of responses provided at the verbatim level (regardless of their accuracy) and the proportion of accurate verbatim responses decreased when the test was delayed by 24 hr and then asymptoted between a delay of 24 and 48 hr. In contrast, the proportion of responses that were accurate only at the gist level remained stable over time.

Testing Effect

A mixed-model ANOVA was conducted on the proportion of accurate responses on the final test, with interpolated testing condition (tested, untested) and level of assessment (verbatim, gist) as within-subject factors and timing of interpolated testing (immediate, delayed 24 hr, delayed 48 hr) as a between-subjects factor. A significant testing effect, $F(1, 93) = 17.93$, $p < .001$, $\eta_p^2 = .16$, was qualified by a significant interaction between testing condition and level of assessment, $F(1, 93) = 21.07$, $p < .001$, $\eta_p^2 = .19$, and by a significant interaction between testing condition, timing of interpolated testing, and level of assessment, $F(2, 93) = 4.37$, $p = .015$, $\eta_p^2 = .09$. When accuracy was assessed at the verbatim level, a significant testing effect was found, with a higher proportion of accurate verbatim recall for previously tested items (.42) than for previously untested items (.28), $F(1, 93) = 35.58$, $p < .001$, $\eta_p^2 = .28$, as well as a significant interaction between interpolated testing condition and timing of interpolated testing, $F(2, 93) = 9.45$, $p < .001$, $\eta_p^2 = .17$. As shown in Figure 2A, the testing effect was substantially larger when interpolated testing was immediate (.30) than when it was delayed by 24 hr (.10), $F(1, 62) = 12.56$, $p = .001$, $\eta_p^2 = .17$, which was not significantly larger than when interpolated testing was delayed by 48 hr (.05), $F(1, 62) = 0.72$, $p = .4$, $\eta_p^2 = .01$.[4]

However, when accuracy was assessed at the gist level, a negative testing effect was found, with a lower proportion of accurate gist (only) recall for previously tested items (.30) than for previously untested items (.36), $F(1, 93) = 4.50$, $p = .037$, $\eta_p^2 = .05$, which seems to derive from cases in which only gist information was recalled accurately in the untested condition yet both verbatim and gist information were recalled in the tested condition. In such cases, interpolated testing yielded an increase in accurate verbatim recall in tandem with a reduction in the absolute proportion of items recalled accurately only at the gist level. More important, the interaction between interpolated testing condition and timing of interpolated testing was not significant, $F(2, 93) =$

[3] The proportion of responses that were inaccurate at both levels increased with time, $F(2, 93) = 12.97$, $p < .001$, $\eta_p^2 = .22$, from .15 when interpolated testing was conducted immediately to .32 when it was delayed by 24 hr, $t(62) = 4.27$, $p < .001$, $d = 1.08$, with an additional nonsignificant increase (to .35) when interpolated testing was delayed by 48 hr, $t(62) = 0.664$, $p = .509$, $d = 0.17$. When these entirely inaccurate responses were removed from the analysis, to examine the relative proportion of responses that were accurate at the verbatim level as opposed to those that were accurate only at the gist level, a significant effect of the timing of the test was found, $F(2, 93) = 3.83$, $p = .025$, $\eta_p^2 = .08$. The relative proportion of responses that were accurate at the verbatim level declined from .43 to .25 with increasing delay (from 10 min to 48 hr), confirming that verbatim memory was more affected than gist memory by the passage of time.

[4] Performance on the tested and untested items was also compared separately for each group, to examine whether a significant testing effect was obtained. Whereas immediate interpolated testing and interpolated testing conducted 24 hr after the event yielded significant testing effects, $t(31) = 7.75$, $p < .001$, $d = 1.37$, and $t(31) = 2.50$, $p = .018$, $d = 0.44$, respectively, interpolated testing conducted 48 hr after the event did not, $t(31) = 0.94$, $p = .357$, $d = 0.17$.

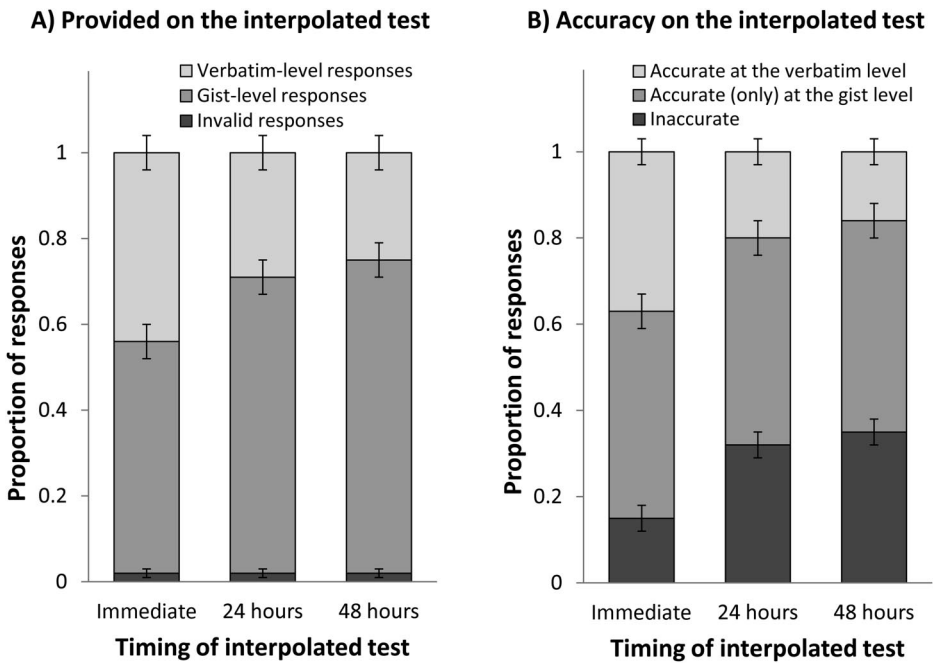

Figure 1. Mean proportion of responses provided on the interpolated test at the verbatim level, gist level, or neither (A) and the accuracy of these responses at the verbatim level, gist level, or neither (B), as a function of the timing of interpolated testing. Error bars indicate ±1 standard error of the mean.

0.42, $p = .661$, $\eta_p^2 = .01$, suggesting that the timing of interpolated testing did not influence the magnitude of the testing effect for gist (only) memory (see Figure 2B). This conclusion is also supported by the findings that 93% of items that were recalled accurately only at the gist level on the final test were also recalled accurately only at the gist level on the interpolated test and that neither this percentage nor the absolute proportion of interpolated responses that were accurate at the gist level were affected by the timing of interpolated testing, $F(2, 86) = 0.30$, $p = .739$, $\eta_p^2 = .007$, and $F(2, 93) = 0.03$, $p = .973$, $\eta_p^2 = .001$, respectively.

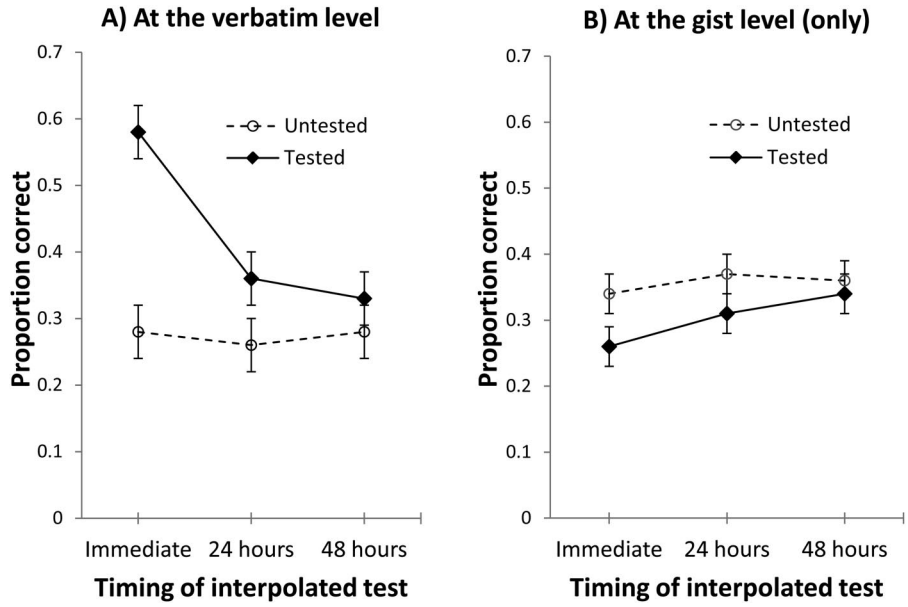

Figure 2. Mean proportion correct with accuracy assessed at the verbatim level (A) and (only) at the gist level (B) on the final test for tested versus untested items as a function of the timing of interpolated testing. Error bars indicate ±1 standard error of the mean.

Given the trade-off between verbatim memory and gist memory as a result of interpolated testing, an additional analysis was conducted to test whether the timing of interpolated testing had a differential effect on verbatim memory than on gist memory. The effects of interpolated testing and its timing were examined on the relative proportion of accurate verbatim responses from the subset of responses that were accurate at either (gist only or verbatim) level. Interpolated testing was found to increase the relative proportion of accurate verbatim responses (by .15), $F(1, 93) = 21.02$, $p < .001$, $\eta_p^2 = .18$, suggesting that interpolated testing had a stronger effect on verbatim memory than on gist memory. Furthermore, a significant interaction was found between interpolated testing condition and timing of interpolated testing, $F(2, 93) = 3.30$, $p = .041$, $\eta_p^2 = .07$, confirming that the timing of interpolated testing had a more pronounced effect on verbatim than on gist memory. As shown in Figure 3, the benefit of testing in terms of the relative proportion of accurate verbatim responses was substantially larger when interpolated testing was immediate (.26) than when it was delayed by 24 hr (.12), $F(1, 62) = 3.68$, $p < .001$, $\eta_p^2 = .30$, which was not significantly larger than when interpolated testing was delayed by 48 hr (.06), $F(1, 62) = 0.45$, $p = .503$, $\eta_p^2 = .01$.

Obviously, information recalled accurately at the verbatim level is also accurate at the gist level, so overall gist memory (i.e., accurate recall of gist regardless of whether verbatim information was also recalled accurately) was expected to benefit from interpolated testing, particularly when conducted immediately, because of the enhancement of verbatim memory. Indeed, overall gist memory was superior after interpolated testing (.73) versus after no interpolated testing (.63), demonstrating a significant testing effect, $F(1, 93) = 18.57$, $p < .001$, $\eta_p^2 = .17$. The interaction between interpolated testing condition and timing of interpolated testing was significant, $F(2, 93) = 7.81$, $p = .001$, $\eta_p^2 = .14$, with a larger testing effect when interpolated testing was immediate (.23) than when it was conducted after 24 hr (.05), $F(1, 62) = 11.62$, $p = .001$, $\eta_p^2 = .16$, which was not significantly larger than when interpolated testing was conducted after 48 hr (.02), $F(1, 93) = 0.16$, $p = .693$, $\eta_p^2 = .003$.

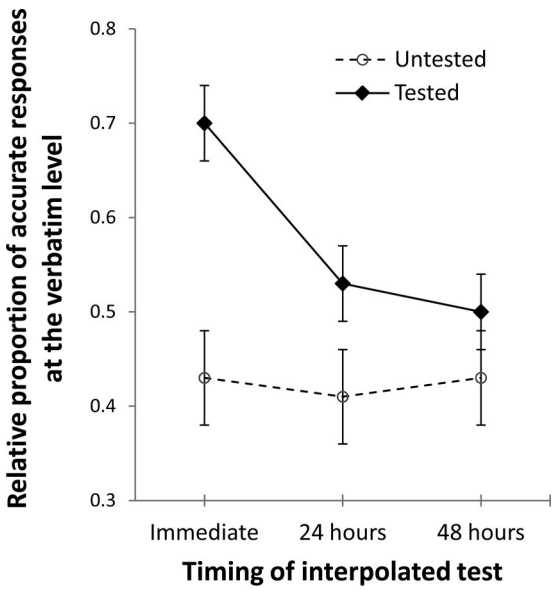

Figure 3. Mean relative proportion of responses that were accurate at the verbatim level from the subset of responses that were accurate at either (verbatim or gist only) level on the final test for tested versus untested items as a function of the timing of interpolated testing. Error bars indicate ±1 standard error of the mean.

Mediation Analyses

The final set of analyses was designed to test the hypothesis that the effect of the timing of interpolated testing on the magnitude of the testing effect (in terms of verbatim recall) was mediated by the proportion of interpolated responses that were accurate at the verbatim level, as a measure of verbatim accessibility. Given that the timing of interpolated testing is a discrete independent variable, Baron and Kenney's (1986) traditional approach for establishing mediation was not applicable, so two alternative approaches were adopted to examine the mediating role of verbatim accessibility at interpolated testing. The first approach is conducting an analysis of covariance (ANCOVA), in which the potential mediator is treated as a covariate and the effect of the experimental manipulation is retested while controlling for the mediator. Reduction of this effect once the covariate is controlled can be taken as support for the mediating role of the covariate. Indeed, when verbatim accessibility at interpolated testing was added as a covariate, the effect of the timing of interpolated testing on the magnitude of the testing effect was reduced and rendered nonsignificant, $F(2, 92) = 2.84$, $p = .06$, $\eta_p^2 = .06$, whereas verbatim accessibility at interpolated testing significantly covaried with the magnitude of the testing effect, $F(1, 92) = 17.82$, $p < .001$, $\eta_p^2 = .16$.

As an additional test of the mediating role of verbatim accessibility at interpolated testing, a bootstrapping analysis was conducted, as recommended by Preacher and Hayes (2004). Bootstrapping involves the repeated extraction of samples from the data set and the estimation of the indirect effect of the independent variable on the dependent variable (via the mediator) in each resampled data set, allowing the construction of a 95% confidence interval for the effect size of the indirect effect. If the confidence interval for the indirect effect does not include zero, this indicates a significant indirect effect, supporting a mediation hypothesis. Because it is a discrete variable, the timing of interpolated testing was coded using two dummy variables, with the first indicating whether interpolated testing was delayed by 24 hr and the second indicating whether it was delayed by 48 hr. Two bootstrapping analyses (each involving the extraction of 5,000 samples) were conducted with the magnitude of the testing effect as the dependent variable; verbatim accessibility as the mediator; and each dummy variable serving once as the independent variable and once as a covariate, with the other dummy variable playing the opposite role. In both analyses, the 95% confidence interval (−.18 to −.03 and −.19 to −.05, respectively) did not include zero, indicating statistically significant indirect effects.

To summarize, the ANCOVA and bootstrapping procedures provide converging evidence that the effect of the timing of interpolated testing on the magnitude of the testing effect was mediated by verbatim accessibility at interpolated testing, as in-

dexed by the proportion of interpolated responses that were accurate at the verbatim level.[5]

Discussion

My main goals in the present study were (a) to test the hypothesis that immediate interpolated testing is more beneficial than delayed interpolated testing for the subsequent recall of verbatim information and (b) to examine whether this effect is mediated by the accessibility of verbatim information at the time of interpolated testing.

Consistent with previous studies (e.g., Chan & Langley, 2011; Gabbert et al., 2009), interpolated testing was found to inoculate eyewitness memory against forgetting, as is evident in a higher proportion of accurate verbatim responses for tested than for untested items on the final memory test. The magnitude of the testing effect was affected by the timing of interpolated testing, such that interpolated testing was more effective when it was immediate than when it was delayed (see also Bergman & Roediger, 1999; Spitzer, 1939).

Several additional findings obtained in the present study suggest that this disadvantage of delayed testing is due to the decline in verbatim memory over time. First, consistent with previous studies (e.g., Goldsmith et al., 2005), fewer verbatim responses were freely reported at delayed than at immediate interpolated testing. Second, the proportion of accurate (freely reported) verbatim responses was also smaller at delayed interpolated testing. Finally, and perhaps most conclusively, the effect of the timing of interpolated testing on the magnitude of the testing effect was mediated by the verbatim accessibility of event details at the time of interpolated testing.

These results support the hypothesis derived from FTT: that weaker inoculation against forgetting of detailed information is achieved when initial testing is delayed rather than immediate, due to the decline in accessibility of verbatim traces over time (e.g., Reyna & Titcomb, 1997). The differential pattern of results obtained for gist memory supports the FTT distinction between verbatim and gist memory: In contrast to verbatim memory, memory for gist (only) was stable over time and was less influenced by interpolated testing and by its timing.

The present findings are also consistent with Glover's (1989) hypothesis that more complete retrieval on intervening tests yields larger benefits for later retrieval. The larger testing effect that was obtained after immediate testing than after delayed testing (both in terms of verbatim memory and in terms of either verbatim or gist memory) can be attributed to more complete retrieval (i.e., more verbatim responses) at immediate interpolated testing.

The findings can also be discussed in terms of the trade-off between retrieval difficulty and retrieval success (see also Finley, Benjamin, Hays, Bjork, & Kornell, 2011; Pyc & Rawson, 2009; Storm, Bjork, & Storm, 2010). On the one hand, it has been suggested that interpolated retrieval is more likely to enhance subsequent retrieval the more difficult or cognitively effortful it is (e.g., Bjork, 1975; Gardiner, Craik, & Bleasdale, 1973). On the other hand, unless one is provided with subsequent study opportunities or corrective feedback (e.g., Kornell, Hays, & Bjork, 2009), interpolated retrieval that is too difficult and hence unsuccessful might be of limited advantage, and the likelihood of successful retrieval is likely to decrease if the delay between encoding and testing is too long. In the present study, the success of interpolated verbatim retrieval was found to decline substantially between immediate and delayed testing conducted after 24 hr. Therefore, it was not surprising that the detrimental effect of declining retrieval success after a delay overshadowed any potential benefit of increased retrieval difficulty in this condition, resulting in a smaller testing effect after delayed interpolated testing.[6]

One final issue warrants a discussion. In the present study, the retention interval between the event and final testing was held constant to equate the baseline untested condition in the three experimental groups against which the effect of the timing of initial testing could be examined. Therefore, delaying the initial test simultaneously shortened the interval between the two tests, introducing a potential alternative interpretation of the present findings in terms of the intertest interval. However, previous findings render such an interpretation unlikely to account for the present results. Using a variety of materials and procedures, several studies have found that extending the retention interval between initial and final testing beyond 24 hr (e.g., Carpenter, Pashler, Wixted, & Vul, 2008; Chan, 2010; Slamecka & Katsaiti, 1988) or beyond 48 hr (e.g., Roediger & Karpicke, 2006) did not affect the magnitude of the testing effect, with the tested items exhibiting comparable forgetting rates over time as the control (untested or restudied) items. Therefore, the larger testing effect obtained after immediate than after delayed testing was not likely to be due to the longer interval between the interpolated test and the final test in the former condition (72 hr) than in the latter condition (48 and 24 hr, respectively, in the delayed-24-hr and delayed-48-hr conditions).

Implications and Conclusions

At the theoretical level, the present findings bring to the fore the need to examine the testing effect in the context of the changes that take place in memory over time. Such examination is important not only for explaining the effects of the timing of interpolated testing but also for the understanding the mechanisms underlying the effects of other events that may take place after initial encoding, such as exposure to misleading postevent information (e.g., Pansky, Tenenboim, & Bar, 2011). At the practical level, the implication of the present study is that the timing of initial eyewitness questioning should be considered seriously. The findings demonstrate that delaying such questioning can substantially reduce its advantage for subsequent retrieval, indicating that time is of the essence with regard to the scheduling of initial questioning of

[5] This conclusion is also supported by the finding that 96% of the items that were accurately recalled at the verbatim level on the interpolated test were also accurately recalled at the verbatim level on the final test and that this percentage was unaffected by the timing of interpolated testing, $F(2, 78) = 1.13$, $p = .327$, $\eta_p^2 = .03$.

[6] It is possible that an optimal initial delay between 0 and 24 hr exists, after which verbatim retrieval would be as successful as at immediate testing but more effortful, consequently yielding a larger testing effect than the one obtained after immediate testing. However, Karpicke and Roediger (2007) have suggested that, in contrast to memory for paired-associate material that is characteristic of research on expanding retrieval, memory for more complex materials may not benefit from delaying the initial test. They proposed that, in such cases, an immediate interpolated test may be sufficiently difficult to yield testing benefits and need not be delayed to make it more difficult.

eyewitnesses. On the positive side, when such questioning takes place almost immediately after the event, the findings confirm that it can serve as an effective technique by which to inoculate eyewitnesses against forgetting event details over time.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182. doi:10.1037/0022-3514.51.6.1173

Bergman, E. T., & Roediger, H. L. (1999). Can Bartlett's repeated reproduction experiments be replicated? *Memory & Cognition, 27,* 937–947. doi:10.3758/BF03201224

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. C. Solso (Ed.), *Information processing and cognition* (pp. 123–144). New York, NY: Wiley.

Brainerd, C. J., & Ornstein, P. A. (1991). Children's memory for witnessed events: The developmental backdrop. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 10–20). Washington, DC: American Psychological Association. doi:10.1037/10097-002

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100,* 42–67. doi:10.1037/0033-295X.100.1.42

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71,* 81–129. doi:10.1006/jecp.1998.2464

Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36,* 438–448. doi:10.3758/MC.36.2.438

Chan, J. C. K. (2010). Long-term effects of testing on the recall of nontested materials. *Memory, 18,* 49–57. doi:10.1080/09658210903405737

Chan, J. C. K., & Langley, M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 248–255. doi:10.1037/a0021204

Finley, J. R., Benjamin, A. S., Hays, M. J., Bjork, R. A., & Kornell, N. (2011). Benefits of accumulating versus diminishing cues in recall. *Journal of Memory and Language, 64,* 289–298. doi:10.1016/j.jml.2011.01.006

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior, 33,* 298–307. doi:10.1007/s10979-008-9146-8

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1,* 213–216. doi:10.3758/BF03198098

Glover, J. A. (1989). The testing phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81,* 392–399. doi:10.1037/0022-0663.81.3.392

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52,* 505–525. doi:10.1016/j.jml.2005.01.010

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131,* 73–95. doi:10.1037/0096-3445.131.1.73

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704–719. doi:10.1037/0278-7393.33.4.704

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32,* 1–24. doi:10.1006/jmla.1993.1001

Koriat, A. (1993). How do we know that we know: The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639. doi:10.1037/0033-295X.100.4.609

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. doi:10.1037/a0015729

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589. doi:10.1016/S0022-5371(74)80011-3

Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory & Cognition, 39,* 155–170. doi:10.3758/s13421-010-0005-8

Pansky, A., Tenenboim, E., & Bar, S. K. (2011). The misinformation effect revisited: Interactions between spontaneous memory processes and misleading suggestions. *Journal of Memory and Language, 64,* 270–287. doi:10.1016/j.jml.2010.12.003

Poole, D. A., & White, L. T. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27,* 975–986. doi:10.1037/0012-1649.27.6.975

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36,* 717–731. doi:10.3758/BF03206553

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447. doi:10.1016/j.jml.2009.01.004

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. In Payne, D. G., & F. G. Conrad (Eds.), *Intersections in basic and applied memory research* (pp. 157–174). Mahwah, NJ: Erlbaum.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. doi:10.1016/j.tics.2010.09.003

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Slamecka, N. J., & Katsaiti, L. T. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 716–727. doi:10.1037/0278-7393.14.4.716

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641–656. doi:10.1037/h0063404

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38,* 244–253. doi:10.3758/MC.38.2.244

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied, 14,* 50–60. doi:10.1037/1076-898X.14.1.50

Appendix A

The Target Items and Respective Test Questions Used in This Study (Translated From Hebrew)

Item no.	Target item	Test question
1	*Ma'ariv* newspaper	What was the reading material that was placed on Inbal's bed when she called her friend Yael to invite her for a get together?
2	*Meridol* toothpaste	When Inbal washed her face before she went out to meet Yael, what was in the tube that was placed on the bathroom shelf?
3	Red grapefruit	Which fruit was placed on the kitchen counter, while Inbal was preparing herself a drink in the morning and the phone rang?
4	Black pen	Which writing utensil did Inbal use to write her father the message delivered to him over the phone?
5	Orbit gum	Which sweet did Inbal take out of her pocket, when she was standing outside the door of her house waiting for her brother Danny to give her a lift to her friend Yael?
6	(the letter) *S*	Which symbol was printed on the cup Inbal was holding in her right hand, when she was sitting across from Yael and the two reminisced about the time they served in the army?
7	Poodle	Which animal did Itai hold in his arms when he met Inbal on her way back home from her visit at Yael's?
8	*Shufersal* supermarket	In which store did Inbal's mother realize she forgot her purse?
9	Subaru car	What was the vehicle out of which Inbal's mother took grocery bags, before she entered the house?
10	Danone yogurt	What kind of dairy product did Inbal's father eat while watching TV?
11	Crisphead lettuce	Which vegetable was placed in a salad bowl on the kitchen table, when Inbal's mother came home and the two discussed what Inbal should wear when she went out with Itai that evening?
12	Mushroom pizza	What was in the oven when Inbal was trying to identify the source of the delicious scent and accidentally burned her finger?
13	*Neka 7* shampoo	Which personal care product did Inbal reach for when she was in the shower, singing to herself?
14	Carlsberg beer	Which alcoholic beverage, other than wine, did Inbal and Itai order when they sat across from each other at the pub?
15	Silver ring	When Inbal opened the drawer in order to find a book that she wanted to read, which piece of jewelry was placed inside the drawer in the open box that was near the book?
16	Wooden chair	When Inbal came back home after the pub and started reading her book, what was she sitting on?

Note. All words that appear in italics refer to well-known Israeli products, brands, or celebrities.

(*Appendices continue*)

Appendix B

Instructions to Participants

Instructions for the Witnessed/Slide Show Event

You will now be presented with a slide show. Please concentrate and attend to it carefully. We will later ask you several questions about the slide show.

Instructions for the Interpolated Test

In this section, we are interested in examining your memory regarding some items that had appeared in the slide show that you viewed at the beginning of the experiment. For this purpose, we will ask you a few questions. Each question refers to a certain item that you are requested to recall.

When answering the questions, please do your best and try to avoid answers such as "I don't remember" or "I don't know." If you don't remember a certain item, give the most suitable answer that you can. Consider each response carefully, because as soon as you press the OK button, you will not be able to change your response.

Instructions for the Final Memory Test

In this section, we aim to examine once again your memory regarding some items that had appeared in the slide show that you viewed in the first experimental session.

You will be asked two questions about each item. In the first question, you will be asked **what** was the item that had appeared in a certain scene. The question refers to the underlined word. For example: "What did Inbal drink at the neighborhood pub?" The correct answer is "wine."

In the second question, we will ask you to provide a more detailed answer regarding the item you provided in response to the first question: We will ask you **what kind** of item was it. For example, if your answer to the first question was "wine," you will now be asked, "What kind of wine?" The correct answer is "red wine."

Consider your responses carefully. As soon as you press the OK button, you will not be able to change them.

A reminder: You are requested to answer the following questions by relying on the **slide show** that you viewed in the first experimental session.

Please do your best and try to avoid answers such as "I don't remember" or "I don't know." If you don't remember a certain item, provide the most suitable answer that you can.

This is the final section of the entire experiment. Please give it your full attention and best efforts. Your responses to the following questions are very important to the goals of this research.

If you wish, you can ask the experimenter to review a printed version of these instructions at any stage.

Received May 4, 2011
Revision received February 16, 2012
Accepted March 19, 2012 ■