Given the complexity of human behavior, it can be difficult to identify factors that affect it so consistently that they can confidently be labeled as laws or principles. A notable exception is the testing effect in human memory: Taking a test on learned information, compared with simply restudying it, renders the information more likely to be remembered in the future. The testing effect has been demonstrated in numerous studies over the last century (e.g., Roediger & Butler, 2011; Roediger & Karpicke, 2006a), and it was recently featured in a practice guide for educators as a technique for enhancing student learning that is backed by strong evidence (Pashler et al., 2007).

Studies of the testing effect typically involve an encoding phase (e.g., a phase in which participants learn French-English word pairs, such as le chien—dog), followed by an attempt to retrieve the encoded information (le chien—?) or restudy it (le chien—dog). The final phase involves another test that is usually similar to the initial test and typically reveals better memory for information that was tested than for information that was restudied.

Most research on the testing effect has shown that taking a test enhances performance on a later test that is similar. We know much less about the potential of testing to promote the application—i.e., transfer—of learning. In this article, I review recent studies that have begun to address this issue, specifically with regard to the benefits of testing on transfer across temporal contexts, test formats, and knowledge domains. The small but growing number of studies on this topic have so far reported robust benefits of testing on transfer of learning. Future research is encouraged that explores the potential of tests to promote not just direct retention of information, but also the application of knowledge to new situations.

Transfer may be broadly defined as the application of learned information to novel contexts. What is meant by “novel context”? Barnett and Ceci (2002) provide a taxonomy of the types of contextual variation that have been explored in transfer studies. Using this taxonomy as an organizational guide, the small but growing body of research on test-enhanced transfer can be described with respect to (a) transfer across temporal context, (b) transfer across test format, and (c) transfer across knowledge domain.

Transfer Across Temporal Contexts

How well is information remembered after one week, one month, or one year? Compared with memory for information that was encountered just a few moments ago, memory after a delay is likely to be represented differently due to the effects of decay, interference, or consolidation. One way to measure transfer, therefore, is to conduct assessments of memory for information under a different temporal context than that in which the information was originally learned.

The benefits of testing appear to hold across a variety of temporal contexts. Carpenter, Pashler, Wixted, and Vul (2008) had participants learn verbal information (e.g., Swahili-English...
word pairs, such as *farasi–horse*) through either testing (*farasi–?*) or restudying (*farasi–horse*); they then tested participants’ memory for these items in the same way after various delays that ranged from 5 minutes to several weeks. Across all intervals, tested items were remembered better than restudied items were. Carpenter, Pashler, and McDermott (2009) also found that after a 9-month delay, middle-school-aged children had superior memory for U. S. history facts if they had been previously tested than if they had been restudied.

Additional studies have confirmed that testing benefits memory assessed after several days (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008), and that sometimes the testing effect is stronger when memory is assessed after a delay than when it is assessed soon after learning (e.g., Coppen, Verkoeijen, & Rikers, 2011; Kornell, Bjork, & Garcia, 2011; Roediger & Karpicke, 2006b; Toppino & Cohen, 2009). These studies provide evidence that the testing effect can transfer across novel temporal contexts when the final memory test is similar to the initial memory test.

**Transfer Across Test Formats**

What about when the final memory test is different from the initial test? This is often the case in everyday testing situations. For example, students studying for the GRE may use flash cards to remember word definitions (e.g., *ephemeral: lasting only a short time*); then, on the test, they may encounter a studied word in the form of an analogy problem (e.g., “ephemeral is to perennial as temporary is to permanent”). Does testing benefit memory even when information is tested later in a new way?

Some studies have addressed this question by administering an initial test of one type, followed by a final test of a different type. For example, Carpenter, Pashler, and Vul (2006) found that retention of word pairs (e.g., *train–plane*) was better after cued recall (e.g., *train → ?*) than after restudying, and this benefit held whether final recall was assessed in the same direction (*train → ?*) or in the opposite direction (*? → plane*). Kang, McDermott, and Roediger (2007) had participants learn information from journal articles by completing short-answer questions on some of the articles, and multiple-choice questions on other articles. Later, some of the content from each article was tested via multiple-choice questions, and other content was tested via short-answer questions. This way, some information was tested in the same format from the initial test to the final test (e.g., first with a short answer question and then with a short-answer question), and some information was tested in a different format (e.g., first with a short-answer question and then with a multiple-choice question). When corrective feedback was provided, short-answer tests enhanced later memory more than rereading the material did, and this benefit held whether the final test consisted of short-answer or multiple-choice questions.

Along similar lines, Carpenter and DeLosh (2006) found that retention of words from lists was better following an initial free-recall test than following either a cued-recall or a recognition test, and this advantage did not depend on whether the final test required free recall, cued recall, or recognition. An initial cued-recall test has also been shown to enhance retention of word pairs more than restudying does, even when the final test requires free recall (e.g., Carpenter, 2009) or cued recall using different cues than the ones used in the initial cued recall test (e.g., Carpenter, 2011).

Other studies conducted in applied educational settings have confirmed that testing can promote transfer of learning across novel test formats. McDaniel, Anderson, Derbish, and Morrisey (2007) assessed memory for material that was learned in an online course either through weekly quizzes or through additional reading. Weekly quizzes, but not additional reading, produced benefits over nonquizzed information on the unit exam. These benefits were observed even though the items on the unit exams (e.g., “All ______ axons, whether sympathetic or parasympathetic, release acetylcholine as a neurotransmitter*”) required different responses than the items on the original quizzes (e.g., “All preganglionic axons, whether sympathetic or parasympathetic, release ________ as a neurotransmitter”).

Rohrer, Taylor, and Sholar (2010) had elementary-school-aged children learn the locations of cities on a map by either matching the city name to its location (testing) or simply viewing the correct locations of the cities (restudying). On a final test, the children showed superior memory for the locations of cities they had learned through testing than for those they had learned through restudying, and this advantage held whether the final test was similar to the initial test (i.e., requiring the matching of a city name to its location) or different (i.e., requiring recall of a city that lay along a route between two other cities).

Finally, there is one known study that has addressed the test-enhanced transfer of spatial knowledge. Carpenter and Kelly (2012) had participants learn the locations of several objects within a virtual environment. After a brief encoding phase, participants were asked to imagine standing at the location of one object facing a second object, and to point in the direction of a third object. This would be akin to imagining standing in Chicago facing Detroit, and pointing in the direction of Kansas City. Participants had to point to the location of the third object either from memory (i.e., testing) or by following a marker that indicated the direction of the object (i.e., restudying). On a final test, participants were required to estimate some of the objects’ locations from vantage points they had not previously encountered. By analogy, if one originally estimates the direction of Kansas City by standing in Chicago facing Detroit, the final test would be akin to imagining standing in Kansas City facing Chicago, and estimating the direction of Detroit. Even from these novel vantage points, participants who had learned the objects’ locations through testing were more accurate on the final test than were participants who had learned them through restudying.
Transfer Across Knowledge Domains

Some situations call for the application of learned information from one knowledge base to another. Like all types of transfer, this application can vary along a continuum from “near” (e.g., the application of a rule or concept from one physical-science problem to another; Chen & Klahr, 1999) to “far” (e.g., the application of a rule or concept that was acquired in the context of a military problem to a medical problem that requires a similar underlying solution; Gick & Holyoak, 1980). There is evidence that testing can enhance the application of learned information both within and across knowledge domains. For example, Chan, McDermott, and Roediger (2006; see also Chan, 2009, 2010) tested participants on a question about a passage (e.g., question: “Where do toucans sleep at night?”; answer: “In tree holes”), and found that this testing facilitated later memory for related content that was never tested (e.g., question: “What other bird species is the toucan related to?”; answer: “Woodpeckers”).

Testing has also been shown to promote transfer of rules to novel, never-before-seen material within the same knowledge domain. For example, Kang, McDaniel, and Pashler (2011) had participants learn a mathematical function by either estimating the value of \( y \) given \( x \) (i.e., testing) or simply seeing the corresponding \( x \)-\( y \) values together (i.e., study). On a final test requiring participants to estimate the same \( y \) values from the \( x \) values, participants performed better if they had learned the \( x \)-\( y \) relationships through testing than if they had learned them through restudying. Furthermore, when presented with new \( x \) values outside the range that was previously learned, participants who had learned the function through testing estimated the novel \( y \) values more accurately than did participants who had learned the function through restudying. Similarly, in a study of natural-concept learning, Jacoby, Wahlheim, and Coane (2010) found that learning to classify birds into particular familial categories (e.g., orioles, finches, etc.) benefited more from testing (i.e., trying to classify birds into their appropriate families and then receiving feedback) than from studying (i.e., merely seeing the birds with their family labels). Learning this information through testing benefited not only later retention of these birds’ families but also the later classification of never-before-seen birds into their correct familial categories.

The beneficial effect of tests on knowledge-based inferences was nicely demonstrated in a recent study by Butler (2010). After reading a text passage, participants either restudied it or completed an initial test on it (e.g., question: “Approximately how many bat species are there in the world?”; answer: “More than 1,000”). Learning this information through testing enhanced performance on a final test that required participants to make inferences on the basis of the learned information (e.g., question: “There are about 5,500 species of mammals in the world. Approximately what percent of all mammal species are species of bat?”; answer: “If there are about 5,500 species of mammals and more than 1,000 species of bat, then bats account for approximately 20% of all mammal species”). Butler also found that initial testing on a given concept, as compared with restudying, led to better transfer across knowledge domains. The final test contained inference questions that differed quite dramatically from the initial questions in their surface details but shared similar underlying concepts. For example, answering a question about the differences between the wing structure of bats and the wing structure of birds led to greater accuracy (relative to restudying the information) on a final test inquiring how a military aircraft modeled after a bat wing would differ from traditional aircrafts. Benefits of testing on performance in answering later inference questions have also been recently reported by Karpicke and Blunt (2011).

Finally, at least one study has shown that tests can facilitate learning of new material that is unrelated to the previously tested material. Wissman, Rawson, and Pyc (2011) had participants read a passage of text that was organized into three sections. After finishing each of the first two sections, some participants attempted to recall what they had just read, and some did not; however, all participants attempted to recall the third and final section after reading it. Even though all participants attempted recall of the third section, those who had attempted recall of the previous two sections recalled more of the third section than did those who had not (see also Szpunar, McDermott, & Roediger, 2008). Although it is not clear whether this interim-test effect persists beyond passages that immediately follow the tested passages (see Wissman et al., 2011, Experiment 4), the benefits of testing on memory for subsequently encountered material might suggest that tests improve metacognitive awareness or encourage the adoption of more effective encoding strategies (see also Pyc & Rawson, 2010). Some support for this possibility may be offered by the fact that testing seems to reduce the ubiquitous tendency for participants to be overconfident in their own learning (e.g., Carpenter & Olson, 2012; Finn & Metcalfe, 2007).

Conclusion

In this brief review, I have summarized recent evidence that tests can promote transfer of learning across temporal contexts, test formats, and knowledge domains. Future research should continue exploring the potential of tests to promote various forms of transfer. A specific question to be addressed is whether testing enhances metacognitive transfer—the enhanced awareness and regulation of one’s own learning—and whether these effects are long-lasting and independent of the specific information being learned. There is great potential for further exploration of whether and how testing consistently enhances what may be considered the ultimate goal of education: the successful application of effective learning tools, strategies, and habits outside of the classroom.

Recommended Reading

retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General*, 140, 535–551. Provides evidence that testing can enhance transfer of learning across different social contexts.


