A common conception of “testing” by both educators and laypeople is one of students sitting in rows, pencils in hand, completing an assessment for a unit that was recently studied in class, an entrance exam for a school or university, or a midterm or final exam for a course. It might also include students cramming for a test the night before, rereading passages from books, or memorizing vocabulary words. Tests are traditionally thought of as assessment instruments which measure what knowledge students possess and provide an instructor’s evaluation of a student’s academic performance, or to summarize the attainment of student educational objectives, often for the purposes of external accountability. Tests are also used to rank and divide learners for the purposes of classroom placement or entrance into an educational institution. These uses of testing are familiar to educators and laypeople alike, and share the common function of assessing student knowledge. However, the technique of “retrieval practice” can also serve an important function that is often given scant attention—utilizing tests not only to measure knowledge, but also increase and enhance it, often significantly improving retention of the tested knowledge (Carpenter et al., 2008; Karpicke & Blunt, 2011; Karpicke & Roediger, 2007; Roediger & Karpicke, 2006).

“Retrieval practice” is when the learner “retrieves” specific information from memory. By doing so frequently, especially in a variety of different contexts, the learner strengthens the capacity to remember this information, assisting long-term retention and transfer of knowledge from one situation to another (Rohrer, Taylor, & Sholar, 2010). Thus, testing not only serves as a means of assessment, but as an educational tool that can directly and positively affect student learning. Several

* Instructor of English as a Foreign Language, Department of Economics, Kwansei Gakuin University
studies (e.g., Cull, 2000; Glover, 1989; Wheeler & Roediger, 1992) show that retrieval practice has a positive effect on learning. While this “testing effect” has been receiving renewed interest as of late, the concept that testing directly improves student learning is not a recent one. Two classic studies by Gates (1917) and Spitzer (1939) showed strong positive effects of retrieval practice on retention. In his study of elementary and middle school children, Gates found that the optimal amount of testing seemed to be about 60% of the total learning period. However, with prose passages, he found that the positive effects leveled off and even appeared to drop when the amount of time spent on testing exceeded 60%. Thus, the data suggest that a certain amount of study may be necessary before recitation can begin to benefit learning. Spitzer conducted a landmark study involving over 3,600 sixth-grade students in which they studied prose passages and took tests over various time intervals during the course of the following 63 days. Spitzer found that the more time passed between students studying the passage and taking the initial test, the poorer the student performance was on that test. In addition, he found that if students were tested again at a later time following the first test, the performance of the students did not decrease, and even in some circumstances, increased. Furthermore, the time interval between studying the passage and taking the initial test also affected the results of later tests; the shorter the time between study and assessment, the better the performance was on the following assessments. Thus, Spitzer concluded that a test should be conducted fairly soon after study—while the student still retains the ability to recall the material—if there is to be a continued positive effect on later tests.

Tulving (1967) also conducted an important study which challenged the common assumption that learning occurs only through studying. He had subjects learn a set of words presented in a different random order in each study session. With one group of subjects, he alternated study and test sessions equally (STST) for a total of 12 study sessions and 12 testing sessions. With another group of subjects, he conducted three study sessions to one testing session (SSST) for a total of 18 study sessions and six testing sessions. The third group had one study session to three testing sessions (STTT) for a total of six study sessions and 18 testing sessions. Tulving found that all three groups performed almost equally as well in their recall. In other words, it made little difference whether students had more study sessions or more testing sessions—their learning of the words was nearly the same.

Karpicke and Roediger (2007) conducted a similar study with some minor modifications, including the conducting of a delayed test one week following the experiment in order to analyze long-term memory effects. They found that recurrent
testing during learning increased recall in relation to repeated studying, although interchanging study and test sessions produced the best results. They also conducted a second experiment in which words that had been recalled by the subjects were eliminated from continued studying or testing in order to examine how this affected retention. The dropping of words to be studied that had been recalled earlier by the subjects did not affect their retention; in other words, studying words that had been remembered earlier did not improve their future recall of those words. However, repeated testing of those previously recalled words did indeed improve retention by more than 100%. The researchers concluded that while additional studying of previously learned information has little effect on retention, repeated testing or retrieval practice has significant positive effects, and are critical to long-term retention. In summary, testing not only assesses learning, but enhances it, and a mixture of study and tests is more effective than expending the same amount of time only studying the material, such as rereading passages or vocabulary lists. Numerous other studies (e.g., Cull, 2000; Glover, 1989; Wheeler & Roediger, 1992) have also shown that retrieval practice has a positive effect on learning.

Several classroom-based studies have also shown that retrieval practice produces significant learning benefits for students. Researchers conducted an experiment in combination with a university course and found that students performed better on their final test if their chapter readings were followed by questions which reviewed the students’ knowledge rather than statements which did not require information retrieval (McDaniel et al., 2007). Butler and Roediger (2007) found that taking an initial short-answer test following a lecture each day improved final recall the most. Research by Karpicke and Blunt (2011) found that even elaborate studying techniques such as concept mapping, while good for learning, are not as effective as retrieval practice (in this case, self-testing by students). However, when students were asked to predict which technique would be more effective for their long-term learning, the majority believed that the opposite was true. Testing also decreases forgetting of material that has been recently studied, and multiple tests have a larger effect in reducing forgetting than does a single test (Wheeler & Roediger, 1992). Even with test delays as long as 42 days, testing can slow the rate of forgetting (Carpenter et al., 2008). Retrieval practice can even be effectively used to remember student names (Morris & Fritz, 2000). This positive effect of retrieval practice is also shaped by the “spacing effect”, which refers to the phenomenon that when retrieval practice/testing sessions are distributed, or spread further apart over time, they are more effective in improving learning over the long-term than when such sessions take place closer together. In addition, students were found to learn more from free-recall tests compared to cued-recall tests (Glover, 1989), which indicates
that retrieval practice that employs the use of minimal cues can increase its positive effect on learning.

Feedback after testing is also effective, particularly feedback after incorrect responses. Guthrie (1971) showed subjects sentences with a word missing, and the subjects had to guess the missing word. Some subjects were given no feedback, while others were given the correct answer. At the end of the learning session, an assessment was administered to determine how learning was affected by the feedback. Guthrie found that when subjects made an error on an item, learning was significantly aided by feedback; however, when subjects did not make an error on an item, providing feedback did not facilitate learning. Other researchers (Pashler et al., 2005) have found that providing the correct answer after an incorrect response not only improved performance during the preliminary learning session, but also increased final recall by nearly 500%. The results also indicated that feedback provided after correct responses had little effect either immediately following or after a delay. Recent research (Hays, Kornell, & Bjork, 2010) also found that providing feedback after correct responses was generally not effective. Furthermore, students giving incorrect answers on tests does not harm student learning, but enhances it, given that adequate feedback is provided (Kornell, Hays, & Bjork, 2009).

In addition, teachers and students should not be apprehensive about introducing difficult material when using tests as tools to enhance learning, even if error rates increase at first (Kornell et al., 2009). Engaging in difficult assignments and taking challenging tests, instead of attempting to avoid chances for error, may be one important aspect of effective learning. The University of Memphis (2008) also recommends the following applications for educational settings: (1) align lessons, assignments and assessments, so that information will have to be remembered at various times throughout the course, improving long-term retention; (2) during lessons, ask students questions that will allow students to reflect upon their understanding of course material discussed in prior lessons; (3) design classroom and homework assignments so that students retrieve important information from lessons, readings, or other materials; (4) facilitate group studying and group tasks in which students must discuss and negotiate meaning; and (5) administer frequent tests, quizzes, or other assessments which improve knowledge and understanding of the material.

The practical implication of these results is that retrieval practice should be a regular part of the classroom as well as among students themselves. Yet surveys of
American university students indicate that most students do not utilize self-testing when studying on their own, and instead study almost entirely by rereading (Carrier, 2003; Karpicke, Butler, & Roediger, 2009). Teachers and students alike should frequently utilize retrieval practice—not only to assess knowledge, but to enhance learning.

References


