Chapter 2

REVIEW OF SELECTED LITERATURE

Introduction

This study is squarely focused upon an examination of aspects of the testing effect, a phenomenon long recognized by historical figures, cognitive researchers, and everyday individuals in various incarnations of the concept “Use it or lose it.” The contents of this literature review will substantially, but not exclusively, focus upon the various aspects of the testing effect. Looking at accounts of past research and a few other historical writings will place this study in context and will begin to lay a foundation for an understanding of this study’s results and conclusions.

Content Focus

Any quantitative, experimental research study that endeavors to garner new insight into the processes of human learning needs some type of targeted content. Strings of numbers, lists of Swahili words, sequences of letters, and various sections of prose have been familiar selections made by researchers. By no means a unique choice, a set of relatively rare and difficult dictionary words and their corresponding definitions have been selected by the researcher as the informational material upon which the methodologies of this study will be focused.

In The Information: A History, a Theory, a Flood, Gleick (2011) identifies the fundamental and earliest English-language appearance of the concepts of dictionary and definition; concepts which are basic to this study’s chosen content of focus. Currently, these ideas are generally taken for granted; for us, it is difficult to imagine a time when
these were not commonly recognized. These reference essentials, however, did certainly have originators and dates of origin. Statements from just a few pioneers in this realm indicate that being a student of the vocabulary of one’s language has long been considered important.

Robert Cawdrey’s 1604 Table Alphabetical (Lancashire, 1994) was an A to Z organized proto-dictionary created for the “..benefit & help of Ladies, Gentlewomen, or any other unskilfull persons.” He hoped that “..they may the more easilie and better understand many hard English wordes, which they shall heare or read in Scriptures, Sermons, or elswhere, and also be made able to use the same aptly themselves.”

In 1690, John Locke (Locke, 1995) embraces the curious task of defining the word definition. “For, definition being nothing but making another understand by words what idea the term defined stands for, a definition is best made by enumerating those simple ideas that are combined in the signification of the term defined: and if, instead of such an enumeration, men have accustomed themselves to use the next general term, it has not been out of necessity, or for greater clearness, but for quickness and dispatch sake.” In so doing, he highlights as the seminal purpose of vocabulary skills; the ability to communicate and make others understand the idea that one has in mind.

Each of these primary references lends a component to the rationale for selecting normal, albeit unusual, English dictionary words as the subject matter for this study. It is fundamentally important (as it has been important for centuries) for individuals to be able to understand difficult words when they are heard, and to be able use them skillfully during attempts to communicate with others. From these early vantage points
both men recognized the importance of being able to understand and skillfully use a substantial number of words.

*Early Cognitive Research*

Ebbinghaus’ *Memory* (1964), first published in 1885, is considered one of the earliest examples of quantitative cognitive research. This one-man research team saw him serving as both the researcher and the subject but he successfully identified early on the importance of the number of repetitions in developing long-term levels of retention. His grueling approach to research with careful implementation, controls, and record keeping set the tone for cognitive research that was to follow throughout the decades to come.

*The Testing Effect*

Gates (1917) asked if, while studying, it is more valuable to attempt to recite the target content or engage in another reading of it. As a result of his studies he concluded that while both strategies have their role, efforts including recitation could allow students to recall up to four times as much as reading alone in a delayed recall method. He qualified the results, however, by stressing the importance of the student spending enough time with the content initially as to avoid “distortion of the material”; that is, the errant development of inaccurate knowledge. He also concluded recitation not only helped students to know more but it also helped them develop confidence that they knew more.

Pyc and Rawson (2010) examined the effectiveness of mediator words in settings with re-studying only versus in settings with a test followed by re-studying. Transitional
words are words often used when individuals are making links between words in their native language and corresponding words in the target language. They concluded that successful retrieval effectively strengthened memory on transitional words. They also postulated that failures to retrieve successfully cued the student to adapt and choose a more effective mediator word.

McDaniel, Anderson, Derbish, and Morissette (2007) attempted to “Test the testing effect in the classroom” in response to their belief that significant research on the phenomenon was not being applied regularly in classroom instruction. They found that students were more successful on unit and final exams with the content that had been quizzed every week than with the content over which they had not been quizzed. They also observed that content that was quizzed via open-ended short answer questions was better handled by students than content that was quizzed via multiple-choice questions. They concluded that “Educational theory and practice would do well not to forget the use of testing as a tool to promote learning and retention.”

Glover (1989) looked at the previous body of evidence indicating the significance of the testing effect and sought to investigate two different hypotheses as to what was producing it. The first hypothesis was the *amount of processing hypothesis*. That is, if more mental processing is dedicated to the content through recall attempts, performance on a final measure would be better. The *number of complete retrieval events* hypothesis postulates that the count of completed retrieval events is the factor that strengthens performance on a final memory test. After conducting four experiments, the researchers concluded that the *number of complete retrieval events* hypothesis was the best fit for the
results. The researchers cautioned, however, that although they controlled the amount of
time allowed during the trials it is difficult to be certain that that it was always a direct
measure of processing time. Glover also laments that while many laboratory studies
focusing upon this concept have been conducted over the years, little research examining
it in the context of education has been conducted.

In 1992 Wheeler and Roediger attempted to reach a conclusion as to why P.B.
Ballard’s 1913 results (Ballard, 1913) and F.C. Bartlett’s 1932 results (Bartlett, 1932)
were so contradictory in spite of the fact that their methodologies were quite similar.
These more modern researchers concluded that the difference in results was not due to
content differences between the two studies. The difference was actually a result of the
different intervals that the two earlier researchers used between their tests. Bartlett used
testing intervals that were significantly longer than Ballard, which ultimately resulted in
poorer performance. Wheeler and Roediger also concluded as a side note that multiple
tests taken a short time after study has completed improved recall greatly on measures
applied one week later.

Nunes and Weinstein (2012) conducted three experiments to investigate the
concern that retrieval practice may produce an increase in instances of false recall on a
cumulative test. The results from these experiments pointed to beneficial results of
repeated retrieval attempts will no increase in the instances of false recall.

A brief article in Science News (Cevallos, 2010) recounts the earlier cited Pyc and
Rawson study and puts it the context of students’ daily study activities. The article
quotes both Pyc and Washington University’s Henry Roediger who each imply that it is
common for students to fool themselves into thinking that they know a particular sphere of content simply by reading it. Both researchers emphasize the importance of informal, personal testing as necessary formative practice for successful performance on a summative assessment.

As the result of three unique experiments Kornell, Bjork, and Garcia (2011) conclude that the successful retrieval of an item on a trial greatly strengthens the memory of that item while an unsuccessful retrieval attempt does not strengthen the memory of an item. They also conclude that the rate of degradation of memory on all items is relatively uniform. This uniform “drop” continually takes items in memory from higher states of accessibility toward lower states of accessibility. Successful retrieval attempts, however, work to move those items much higher above this threshold of forgetting. This makes those items have a greater “distance” to fall before reaching a point of inaccessibility; even though all the items are heading toward that threshold at the same rate.

In a publication different from their previously cited work, Roediger and Karpicke (2006) provide a valuable review of research relating to the testing effect. In their introduction they liken the testing effect to a psychological version of the Heisenberg uncertainty principle; which basically states that it is impossible to measure something without changing it as well. Attempts to measure knowledge in individuals invariably alter those individuals’ knowledge as well. They summarize the significant contributions of significant researchers such as Gates, Tulving, and others and conclude that in a wide variety of research implementations the testing effect is a powerful tool for learning in
education. They also intimate that the application of its power to education is one that is long overdue.

McDaniel, Roediger, and McDermott (2007) do take steps to investigate the application of the testing effect into the realm of education. The sought to go beyond the memorization of word lists that are the steady diet of laboratory cognitive research. They sought to apply the testing effect in the context of articles, lectures, and other materials from a college class and in various classroom settings. They relate that their findings in regards to the testing effect in an instructional context are as robust as those observed in experimental settings that are not as real world.

Halamish and Bjork (2011) were determined to examine boundary issues in relation to testing events as learning events. They examined the roles of variations in final test formats and their resultant difficulty. Of most significance to this proposed study is the indication that retrieval-rich experiences did not show superior results over study-rich experiences when the final assessment was a cued-recall test. The advantage was obvious only when a free-recall test was the final assessment. Each of the experiments in this study was conducted in the absence of feedback for participants.

Serial Position Effect

When conducting research that involves exposing participants to lists or groups of words over which they are later tested upon, it should be realized that the order in which the words were presented can be very influential to the results. Murdock (1962) conducted an experiment that showed slightly higher levels of probability of recall for items that were presented sequentially first to the participants compared to items more
centrally presented in sequence. The latest items in the presented sequence showed even more drastically high levels of recall probability.

Conclusion

The background information gained by the review of these previous works and studies informs the methodology and goals of this present study. It is hoped that the research study described in the following pages will advance the existing understanding of the testing effect portrayed by these earlier studies. This foray into the new context of a population of high school students attempting to gain long-term use of difficult vocabulary words was conducted in an attempt to extend the extant knowledge on these related subjects.