An argument for a spaced repetition standard

Matt Chung
mchung63@gatech.edu

Abstract— This paper discusses to what extent spaced repetition learning is an effective response to combatting the "forgetting curve", and in which educational contexts it is a successful, or unsuccessful, tool. In order to ascertain the efficacy of spaced repetition learning, this paper will consider the four types of knowledge acquisition – factual, observational, procedural and metacognitive - and focus on learning in three particular academic contexts: medical learning and linguistic learning. This paper will ultimately conclude that while spaced repetition algorithms are highly effective learnings tools for learners in diverse fields of study, they have their limitations and therefore must be used in tandem with other educational tools, and for developing knowledge in three of the four areas of knowledge acquisition: factual, conceptual and procedural knowledge. A standard for spaced repetition would therefore be effective so far as it took these arguments into account.

1 QUALIFIER QUESTION

Spaced repetition algorithms are generally responses to the "forgetting curve" and can be applied to a wide variety of materials in a large number of contexts. What does the literature say about these different contexts and what commonalities / differences exist that would make a standard either effective or ineffective? What does the literature say regarding types of learning that spaced repetition is not effective at addressing?

2 INTRODUCTION

Since the 1880s, cognitive psychologists have run hundreds of experiments proving the advantage of spaced practice over mass practice (Kang 2016). But is spaced practice applicable to all contexts? Should spaced practice be used for all types of learning? Under what circumstances should it not be used? Are there any negative side effects to utilizing spaced repetition? To answer these questions, this paper begins by exploring psychological phenomena that embody (?)

spaced repetition. Following this, the paper then examines the different types of knowledge in which spaced repetition can be applied and subsequently explores literature on the application of spaced repetition across different subjects and disciplines, including second language acquisition and medicine.

Spaced repetition learning is a method used to efficiently memorize information to improve long-term retention. It is a technique where the learner repeatedly reviews content following a schedule determined by an algorithm (Tabibian et al. 2019). Although the definition was only formalized in the 1960s (Ausubel and Youssef 1965), the idea goes back as far as 1620, when Bacon stated:

"If you read a piece of text through twenty times, you will not learn it by heart so easily as if you read it ten times while attempting to recite from time to time and consulting the text when your memory fails" (Bacon 1620).

While spaced repetition is gaining popularity, it should be noted that there are certain types of knowledge acquisition to which this technique is best suited for. Furthermore, students utilizing this technique should be aware that spaced repetition can, in some instances, have a negative impact on learning (Roediger III and Karpicke 2006). Though spaced repetition learning is not without flaws, when used correctly and for the right type of knowledge acquisition, spaced repetition learning can be used effectively across most subjects and disciplines.

As described in more detail later in this paper, human learning can be divided three areas: cognitive (i.e. knowing), affective (i.e. emotions, feelings) and psychomotor (Wilson and Leslie 2016). This paper's focus will be on the application of spaced repetition as it applies to the levels of knowledge acquisition: factual, conceptual, procedural, and metacognition. However, due to the limitations of this paper, *metacognition* will not be discussed in great detail due to its abstract nature.

2.1 Three psychological phenomena

Space repetition consists of three psychological phenomena that facilitate learning. These three phenomenon are the *forgetting curve*, the *spacing effect* and the *testing effect*. (Teninbaum 2017).

Discovered in 1885 (Ebbinghaus 1885), the *forgetting curve*, reveals that in the context of memorization, an individual is most likely to forget information as soon as they learn it. What this curve also demonstrates is that the longer information is retained for, the less likely it will be forgotten in the future. Thus, it's important to review material early on, in order to keep these memories alive (Teninbaum 2017).

Not to be confused with the forgetting curve is the *spacing effect*. The *spacing effect* describes the phenomena wherein people tend to remember things more effectively if they use spaced repetition practice — short study periods spread out over time — as opposed to practicing en masse (i.e. "cramming"). Most people would agree that a single exposure to material is insufficient for long term retention (Kang 2016). The phenomenon was first documented by Ebbinghaus, who used himself as a subject in several experiments to memorize verbal utterances (Settles and Meeder 2016).

Going hand in hand with the *spacing effect* is *the testing effect*. Studies show that taking a test on material that has been studied, increases and promotes learning and retention (McDaniel et al. 2007). Examples and studies of this phenomenon will be presented in depth in the language acquisition section further down in this paper.

Collectively, these three concepts — *the forgetting curve, spacing effect and the testing effect* —make up spaced repetition. Before discussing the different contexts in which spaced repetition is well suited for, it is important to first understand the four different types of knowledge.

2.1.1 *Types of Knowledge*

As mentioned earlier, human learning can be divided into four categories (Wilson and Leslie 2016) and that this paper focuses on the cognition. Anderson and Krathwohl explain that the cognitive domain consists of different types and levels of knowledge: factual knowledge, conceptual knowledge, procedural knowledge and metacognitive. (Krathwohl 2002)

Factual knowledge is the collection of basic elements that an individual must be familiarized with in order to gain familiarity with a discipline, or in order to solve basic problems within that particular domain. For example, if an aspiring computer scientist wants to build a compiler that translates a programming language into machine code, they first must understand the fundamentals of finite automaton and regular expressions (Cooper and Torczon 2011).

Once the basic elements are understood, an individual can then draw connections between them, allowing them to function together. These connections are what's known as *conceptual knowledge*, the second level of knowledge. Continuing the example of compilers, the student can then unify the two topics to understand that together, they make up what is called a *scanner* (Aho, Sethi, and Ullman 1986).

These first two types of knowledge deal with learning about a particular domain. It is important to note that while one might know about something, does not mean one can yet apply that knowledge. The application of knowledge is known as *procedural* knowledge. Returning once more to the example of a computer scientist attempting to build a compiler, one might understand the concept of a *scanner* but may not be able build one (using a programming language) oneself.

Finally, the fourth type of knowledge is *metacognition*, which deals with one's own awareness of cognition (Wilson and Leslie 2016). However, due to the limitations of this paper, *metacognition* will not be discussed in greater depth due to its abstract nature.

2.1.2 *Arguments against spaced repetition*

TODO: add transitional paragraph or phrase between this section (i..e arguments for spaced repetition) and the previous section (i.e. types of knowledge)

As mentioned earlier, the *testing effect* proves that individuals will recall learned information more readily after testing themselves on the information. However, what happens when the information being rehearsed contains incorrect information, like in the case of multiple choice, or true or false questions? Similarly, what effect does recalling the incorrect answer have — does this promote memorization of incorrect information? According to one study (Hasher, Goldstein, and Toppino 1977), that's exactly what happens: students judged statements they had previous encountered as more true than new statements, even in the cases where those statements were incorrect (Roediger III and Karpicke 2006). This phenomenon is known as the *negative suggestion effect* (Remmers and Remmers 1926).

Another negative effect of testing is known as *interfering effects of recall* (Roediger III and Karpicke 2006). In a nutshell, this means that although the act of recall increases the probability of later recall for some material, this act can impair the recall of some other material (Roediger III and Karpicke 2006). Put differently, this means that if a student is applying testing and recall for English, their ability to recall another subject (e.g. Math) is negatively impacted.

2.2 Arguments for spaced repetition

Despite the negative effects of spaced repetition, the positive effects are often so large that in most circumstances, they will overwhelm the relatively modest interference effects (Roediger III and Karpicke 2006). This next section begins with exploring how spaced repetition is applied to language acquisition.

2.2.1 *Language Acquisition*

Spaced repetition as it relates to second language acquisition has been studied in depth by Atkinson, who evaluated four different learning strategies against 120 undergraduate Stanford students that were presented a large set of German-English vocabulary (Atkinson 1972). Four groups were created, and the 120 students were divided evenly into the groups. The students then participated in two instructional sessions, the first session lasting two hours, and then a second *testing* session administered one week later. Each group was taught using a different one of the four learning strategies.

The first strategy used presented students with material in a random order; accordingly, the next group of students were taught using the second strategy, wherein the subjects themselves controlled the order of the material, a strategy known as "*learner control*"; and finally, the third and fourth strategies followed a mathematical model of the learning process based off of the *forgetting curve* and *spacing effect*. The difference between the third and fourth strategies is that the third strategy assumed that the two vocabulary lists were of equal difficulty, whereas the fourth assumed that the vocabulary lists were of differing difficulty. These four strategies were presented in the order of their effectiveness.

The study shows that compared to the first strategy, the second strategy and its students outperformed the first by 53% on the final test. Similarly, the third and fourth models yielded an improvement of 108%, more than twice the improvement.

These findings on the benefit of testing are consistent with another study on language acquisition that was conducted by Hogan and Kinstch in 1971 (Hogan and

Kintsch 1971). In this study, students were divided into two groups, and each group presented a list of 40 words. The first group studied this list four times with short breaks in between. In contrast, the second group studied the list only once but took three recall tests. Two days later, both groups were tested against the list and the results revealed that the first group recalled only about 15% of the words, while the second group recalled about 20 percent of the worlds.

In much the same way, another randomized controlled study (Larsen, Butler, and Roediger III 2009) revealed the importance of spaced repetition for residents in medical school. In this study, two counterbalanced groups were formed and students were randomly placed in one of the two groups — one group studied a review sheet on epilepticus and took tests on myasthenia gravis, and the other group studying myasthenia gravis and taking tests on epilepticus. For both groups, a lecture was delivered and then immediately following this lecture, the students took tests and were provided feedback on their results. This combination of lesson, followed by a test and feedback, occurred two times for specific intervals for roughly two weeks. Then, at six months, all the students across both groups took a final test on both topics. The end result was that the two groups in this study performed 13% higher than other residents who only studied (i.e. no testing, no feedback). This demonstrates that students using repeated testing improved their long term retention better than students that only used spaced studying.

Another impressive example is a randomized control study conducted to determine the impact of spaced repetition on teaching surgical skills (Moulton et al. 2006). In this study, 38 junior residents were divided into two groups. Although both groups spent an equal amount of time practicing these surgical skills — general surgery, urology, orthopedics, plastic surgery, otolaryngology, neurosurgery and cardiac surgery — one group's practice occurred all on a single day whereas the second group's practice was spread out across multiple days. The students in the groups were all tested prior to receiving the training, immediately after receiving the training, and then again one month post training. The result was that the group with the distributed, spread out practice performed significantly better in most measurements including time, number of hand movements, and expert global ratings.

3 GENERATION EFFECT

Another important aspect of spaced repetition learning that must be taken into consideration is the generation effect: how the material that you use is created. In addition to testing oneself at repeated, spaced out intervals, the material is most effectively committed to memory when the flash cards (or testing material) are created by the student themselves. This is known as the generation effect (Smith 2012). This effect was not only studied by Smith, but also by Slamecka and Graf, two scientists who discovered that, during their experiments, students will remember words that they themselves generated (Slamecka and Graf 1978).

3.1 Conclusion

Spaced repetition algorithms are generally responses to the "forgetting curve" and can be applied to a wide variety of materials in a large number of contexts. What does the literature say about these different contexts and what commonalities / differences exist that would make a standard either effective or ineffective? What does the literature say regarding types of learning that spaced repetition is not effective at addressing?

In conclusion, the research shows that a standard for spaced repetition is recommended as it is a highly effective memorization and knowledge acquisition tool that is effective for diverse areas of study; from fact memorization, to language learning, to the development of surgical skills. The literature shows that while these areas of study present different contexts for learning, demanding the me-memorization of diverse material, they are alike in that they all fall under one of the first three categories of human learning: factual, conceptual, or procedural knowledge. This similarity makes them suitable candidates for use with a spaced-repetition standard. However, spaced repetition is not a silver bullet if used in isolation, without the support of other educational tools (Wilson and Leslie 2016). In particular, when acquiring a second language, one should employ a combination of techniques including reading, contextual guessing, and paying attention to word formation (Gu and Johnson 1996). Equally important to employing multiple strategies when learning, a student should be involved in creating the learning material itself since material that is self-generated is more likely to be remembered (Slamecka and Graf 1978). Finally, the positive effects of testing are often so large that in most circumstances, they will overwhelm the relatively modest interference effects (Roediger III and Karpicke 2006).

References

Aho, Alfred V., Ravi Sethi, and Jeffrey D. Ullman. 1986. "Compilers, Principles, Techniques." *Addison Wesley* 7(8):9.

Atkinson, Richard C. 1972. "Optimizing the Learning of a Second-Language Vocabulary." *Journal of Experimental Psychology* 96(1):124.

Ausubel, David P., and Mohamed Youssef. 1965. "The Effect of Spaced Repetition on Meaningful Retention." *Journal of General Psychology; Provincetown, Mass., Etc.* 73:147–150.

Bacon, Francis. 1620. *Francis Bacon: The New Organon*. Cambridge University Press.

Cooper, Keith, and Linda Torczon. 2011. *Engineering a Compiler*. Elsevier.

Ebbinghaus, Hermann. 1885. "Memory: A Contribution to Experimental Psychology." *Annals of Neurosciences* 20(4):155.

Gu, Yongqi, and Robert Keith Johnson. 1996. "Vocabulary Learning Strategies and Language Learning Outcomes." *Language Learning* 46(4):643–679.

Hasher, Lynn, David Goldstein, and Thomas Toppino. 1977. "Frequency and the Conference of Referential Validity." *Journal of Verbal Learning and Verbal Behavior* 16(1):107–112.

Kang, Sean H. K. 2016. "Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction." *Policy Insights from the Behavioral and Brain Sciences* 3(1):12–19.

Krathwohl, David R. 2002. "A Revision of Bloom's Taxonomy: An Overview." *Theory Into Practice* 41(4):212–18.

McDaniel, Mark A., Janis L. Anderson, Mary H. Derbish, and Nova Morrisette. 2007. "Testing the Testing Effect in the Classroom." *European Journal of Cognitive Psychology* 19(4–5):494–513.

Moulton, Carol-Anne E., Adam Dubrowski, Helen MacRae, Brent Graham, Ethan Grober, and Richard Reznick. 2006. "Teaching Surgical Skills: What Kind of Practice Makes Perfect?: A Randomized, Controlled Trial." *Annals of Surgery* 244(3):400.

Remmers, H. H., and Edna M. Remmers. 1926. "The Negative Suggestion Effect on True-False Examination Questions." *Journal of Educational Psychology* 17(1):52.

Roediger III, Henry L., and Jeffrey D. Karpicke. 2006. "The Power of Testing Memory: Basic Research and Implications for Educational Practice." *Perspectives on Psychological Science* 1(3):181–210.

Settles, Burr, and Brendan Meeder. 2016. "A Trainable Spaced Repetition Model for Language Learning." Pp. 1848–1858 in *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*.

Slamecka, Norman J., and Peter Graf. 1978. "The Generation Effect: Delineation of a Phenomenon." *Journal of Experimental Psychology: Human Learning and Memory* 4(6):592.

Smith, S. Adam. 2012. "An Exploration of the Negative Effects of Repetition and Testing on Memory." *The Yale Review of Undergraduate Research in Psychology* 78.

Tabibian, Behzad, Utkarsh Upadhyay, Abir De, Ali Zarezade, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. 2019. "Enhancing Human Learning via Spaced Repetition Optimization." *Proceedings of the National Academy of Sciences* 116(10):3988–93.

Teninbaum, Gabriel H. 2017. "SPACED REPETITION: A METHOD FOR LEARNING MORE LAW IN LESS TIME." *The Journal of High Technology Law*, January, 273-.

Wilson, Leslie Owen, and Contact Leslie. 2016. "Anderson and Krathwohl Bloom's Taxonomy Revised." 7.