Mnemonics, Testing, and Creativity: Creative Thinking and Effectiveness of Learning Method

Sarah Krull graduated in May 2017 from Northern Kentucky University after earning a Bachelor of Science in Psychology with minors in English and Honors. She currently hopes to begin a PhD in Psychology in the fall of 2019. Her main research interests include morality, psychology of religion, and psychology and media.

Mnemonics, Testing, and Creativity: Creative Thinking and Effectiveness of Learning Method

Sarah R. A. Krull. Faculty mentor: Kalif E. Vaughn
Psychological Sciences

Abstract

Much research has been done on various techniques for increasing learning, and both testing and mnemonics such as the keyword method have been proven effective. We addressed whether people high and low in creativity benefit more or less from test practice as compared to a mnemonic strategy. We had participants study twenty Lithuanian-English word pairs. Then participants either restudied the words, received test practice by being given the cue and attempting to recall the target, generated a mnemonic linking the cue to the target, or received no extra practice with the words. After a final test over all twenty words, participants completed Remote Association Triad (RAT) problems (Bowden and Jung-Beeman, 2003) which required participants to find the relationship between three words (e.g., "Cold" is related to "Sore, Shoulder, Sweat"), and yields an index of creativity. Our hypothesis that the effectiveness of different study methods would differ across groups failed to achieve significance, and, surprisingly, so did the expected testing effect. One interesting finding did emerge: higher creativity, as measured by RAT performance, benefited participants regardless of the method employed in studying.

Keywords: retrieval practice, mnemonics, creativity

Introduction

For years, psychologists have conducted extensive research on learning, and with the study of learning comes the study of techniques to facilitate it. Over time, two techniques have emerged as particularly powerful—retrieval practice and the keyword method. Roediger and Karpicke (2006) note that scientists once considered testing simply a tool for assessment of knowledge; however, Tulving (1967) provided some initial evidence that test trials enhance learning. Tulving (1967) had participants study common nouns (study phases were denoted by "S") and freely recall them (recall phases were denoted by "R"). After the initial study phase ("S"), participants either received two additional study phases, followed by a free recall test (SSSR), two free recall tests alternated with two study phases (SRSR), or three free recall tests (SRRR). When Tulving compared final recall performance across groups, he found that recall was approximately the same across groups. The implication that testing could be used to improve, not simply to demonstrate, knowledge opened up a whole new field of research.

Research since Tulving (1967) suggests that testing enhances learning across a variety of situations. For instance, Pierce and Hawthorne (2016) asked college students to memorize a list of words, administered either by audio or visual means. Some of these participants were also required to take a free recall test shortly after completing each list. The researchers found that although modality impacted overall performance in cases where the items were non-categorizable, testing benefited recall for both groups. Furthermore, testing appears useful in situations that hamper study. Mulligan and Picklesimer (2016) first presented participants with a list of 60 unrelated word pairings to study. Then some of the participants reviewed a list of 20 word pairs. During this period, some were challenged to recall the second word (testing period) and some simply saw both words again (restudy). Additionally, the researchers required some participants to perform two tasks simultaneously (representing a divided attention task). In the divided attention condition, participants who restudied the words performed worse than those who attempted to remember them, suggesting that test practice may benefit learning even when attentional resources are scarce.

Research has also found evidence that the testing effect translates from the laboratory into scholastic life. Batsell, Perry, Hanley, and Hotstetter (2016) conducted a quasi-experiment, using two classes, both of which were assigned daily readings from sections of a textbook that were not covered in class. In the testing class, the teacher administered a brief quiz at the start of each class period. In the control class, the teacher simply encouraged the students to read the assigned material. Throughout the semester, both classes were tested on the material during scheduled exams. The test group outperformed their counterparts in the control group, even on questions that had been answered in the material but had not appeared on the daily quizzes. Clearly, testing benefits the process of learning.

However, interestingly, the testing effect appears to have a very significant weakness. Peterson and Mulligan (2013) suggested that under certain circumstances, testing may negatively impact recall by focusing attention on aspects of the memory items that will help cued recall, but not free recall. The researchers asked participants to study a set of 36 rhyming words (e.g., cork/fork) taken from six different categories (e.g., kitchen utensils). Afterwards, they either restudied the word pairs or tried to recall the second word when presented with the first

word. When the researchers asked the participants to recall as many words as they could, the restudy group emerged with better recall than the testing group. Apparently, the emphasis on cued recall prevented the test group from noticing the categories, which might have helped on the free recall test.

Fortunately, learners need not depend solely on testing to improve their knowledge. Another powerful technique, the keyword method, has also proved effective. The keyword method requires the learner to connect two often dissimilar words with a mental picture linked phonologically to the word to be learned (for instance, using the word "wing" to remember the Swahili-English word pair "wingu-cloud"). For instance, Atkinson (1975) assigned participants to learn a set of 120 Russian words. Participants listened to the Russian words through headphones while viewing the English equivalent on a CRT device. For participants in the keyword condition, a keyword (displayed in brackets) accompanied the English word. After each study period, participants took a test in which they listened to the Russian word and then had to type the English version. Participants also took a test over all 120 words after the three study/test trials and another comprehensive test six weeks later. Atkinson discovered that the keyword group recalled far more words than their counterparts in the control group (72% to 46% recall performance on the first comprehensive test and 43% to 28% recall performance on the second test occurring after six weeks for the keyword versus control group, respectively).

Piribabadi and Rahmany (2014) found that the benefit the keyword method produces in language learning translates to learning technical jargon. The researchers randomly selected 120 university students from two engineering classes. Using the Oxford placement test to assess the students' English vocabulary, the researchers sorted the sixty students taken from each class into upper-intermediate (30 highest scores) and lower-intermediate groups (remaining scores). Then one of the classes learned engineering terms using the word-list method and the other learned engineering terms using the keyword method for four weeks. Following this training both classes took a multiple-choice vocabulary test to ascertain learning. In both the upper- and lower-intermediate groups, students trained in the keyword method achieved mean scores higher than their counterparts trained in the word-list method.

Jenpatturakul (2012) found some interesting results particularly relevant to the current study. The researcher selected 40 Bangkok university students enrolled in an English course. During the first week, students learned new words in English. In the second week, they took the first of two vocabulary tests. The teacher introduced the students to the keyword method during the third week and had them practice the technique using new words. At the end of the class, the students took a test over some of the vocabulary learned using the keyword method. Interestingly, not only did the students perform better on the second test, but also 97.5% of them self-reported that the keyword method improved their imagination.

In planning this study, we became interested in the effect of creativity and verbal aptitude on both test practice and mnemonic techniques as study strategies. Testing requires little creative effort; learners are given a cue and attempt to recall the corresponding target. Indeed, Peterson and Mulligan's (2013) results could be interpreted as saying that testing sometimes prevents a person from coming up with a new, better way to solve a problem (that is, being creative). It may therefore prove more effective for learners who lack creativity in certain situations. The keyword method, by contrast, requires the learner to connect two often dissimilar words with a mental picture linked phonologically to the word to be learned. (For example, a learner using the keyword method for the Lithuanian word suo [dog] might create the image of a sumo-wrestler dog to remember the definition.) Thus, the keyword method may be more suited to individuals higher in creativity and verbal aptitude. Therefore, we hypothesized that individuals higher in verbal aptitude and creativity would learn better using the keyword method and individuals lower in creativity would benefit more from test practice. To assess creativity and verbal aptitude, we had participants solve Remote-Associate Triad problems that required them to think of one word related to three other words (e.g., "Cold" is related to "Sore, Shoulder, Sweat"; see Bowden and Jung-Beeman, 2003).

Methods

Participants

361 participants from Northern Kentucky University started the online experiment; however, participants were excluded if they did not finish (n = 61), indicated that they had completed the experiment previously (n = 16), or wished to be excluded (n = 18). After excluding these participants, the remaining sample size was 266 participants (182 females, 62 males, 1 transgendered male, 1 non-binary assigned male at birth, and 1 genderfluid; median age = 19.8 years, age range: 17-43 years). 61 participants did not enter their age or entered a nonsensical value (e.g., 0), and 19 participants did not indicate their gender or entered a nonsensical value (e.g., "vhvhj"). Regarding race, participants indicated whether they were "African-American, Non-Hispanic" (n = 14), "American Indian/Native Alaskan" (n = 1), "Asian/Pacific Islander" (n = 3), "Other" (n = 6), "White, Non-Hispanic" (n = 220), or did not indicate their race or had unusable data for this question (n = 14). Regarding classification, participants indicated that they were Freshman (n = 154), Sophomore (n = 48), Junior (n = 23), Senior (n = 16), Post-baccalaureate (n = 3), Non-degree seeking (n = 1), or did not answer or provided unusable data (n = 21).

Materials

Participants studied 20 Lithuanian-English word pairs (e.g., namas-house) during the experiment (see Appendix A for a complete list of these words). Additionally, participants attempted to solve 10 Remote-Associate Triad problems (reported in Appendix B; see Bowden and Jung-Beeman, 2003) to assess their creativity.

Procedure

On NKU's Sona psychology experiment systems page, participants viewed the study name and a brief abstract. We provided a link to the actual experiment, which was hosted on another website. Once participants clicked the link, they saw a consent form, which briefly described the study and informed them that participation was voluntary and that their data would be anonymous. Once the participant gave consent, they were directed to the experiment and randomly assigned to groups.

During the first phase of the experiment, participants studied the Lithuanian-English word pairs one at a time. On each trial, a Lithuanian-English word pair appeared. Below the Lithuanian-English word pair, the Lithuanian word was repeated with a textbox next to it for the participant to copy in the corresponding English word. After all 20 word pairs had been presented during the copy trials, the second phase of the experiment began.

During the second phase of the experiment, participants completed different tasks depending upon group assignment. In the retrieval group, the participants were presented with the Lithuanian word and asked to recall the corresponding English word. In the keyword group, the participants were instructed to generate a mnemonic for each of the word pairs, which we instructed them to type into a textbox on the screen. In the restudy group, they restudied the word pairs in the same way as during the first phase. The control group did not receive additional exposure to the word pairs and were instead directed immediately to the distractor task (see below).

After the second phase, all four groups played Tetris for two minutes as a distractor task. After the distractor task was finished, the third phase of the experiment began.

During the third phase of the experiment, all of the groups took a final test over the word pairs. During each final test trial, we presented one Lithuanian word at a time and instructed participants to recall the English translation. Each final test trial was self-paced. After the final test, the fourth phase of the experiment began.

During the fourth phase of the experiment, participants completed 10 Remote-Associate Triad problems (see Bowden and Jung-Beeman, 2003). During each RAT problem, we presented three words to the participant (e.g., "Sore, Shoulder, Sweat") and asked them to generate the related fourth word (e.g., "Cold").

Statistical Analysis

To evaluate whether the learning strategy influenced final test performance, an ANOVA was conducted comparing group (keyword, retrieval, restudy, or control) and final test performance.

To determine whether creativity influenced final test performance, participants were first divided into "high" or "low" creativity based upon their RAT performance. Then, an ANOVA was conducted comparing group (keyword, retrieval, restudy, or control), final test performance and RAT performance (high versus low).

Results

Copy Performance (Phase 1)

All participants were given an initial study phase to the word pairs, during which they were instructed to copy the target word (the English word) into a textbox next to the Lithuanian word. The proportion of items that were copied correctly during the study phase was high ($M = 0.96$, $SD = 0.17$), suggesting that participants successfully encoded most of the items. Importantly, out of the 266 participants remaining in the study, there were some participants who failed to copy any items correctly ($n = 4$) or copied 25% or less of the words correctly ($n = 5$). To prevent possible skewing of the results, these nine participants with poor or zero copy performance were excluded from subsequent analyses. Thus, the remaining analyses only included participants with high copy performance (greater than or equal to 75%; $n = 257$). The 257 participants were distributed relatively evenly across the keyword group ($n = 61$), restudy group ($n = 68$), test group ($n = 64$), and control group ($n = 64$). For these 257 participants, the proportion of words copied correctly during the study phase increased ($M = 0.99$, $SD = 0.04$). Importantly, any items which were not correctly copied during the study phase ($68/5140 = 1.32\%$) were excluded from subsequent analyses.

Recopy and Test Performance (Phase 2)

Participants in phase 2 either did nothing (control group), recopied the word pairs (restudy group), attempted to recall the English word when presented with the Lithuanian word (e.g., "namas - ???"; retrieval group), or generated a keyword mnemonic to associate the Lithuanian word with the English word (e.g., one participant wrote "mama's house" to remember namas - house; keyword group).

Participants in the restudy group copied items at a high rate during the restudy phase ($M = 0.97$, $SD = 0.13$). Participants in the retrieval group occasionally recalled the corresponding English word when presented with the Lithuanian cue ($M = 0.30$, $SD = 0.30$). Given that participants in the control group received no further exposure to the material before the final test, and

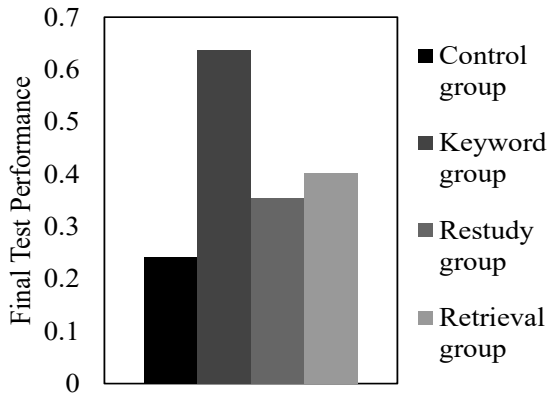

Figure 1. Final test performance (reported as a proportion correct) as a function of group (control, keyword, restudy, and retrieval).

participants in the keyword group generated their own mnemonics that cannot be assessed for accuracy, no statistics are reported for the control or mnemonic group in Phase 2.

Final Test Performance

Final test performance as a function of group is reported depicted in Figure 1. A one-way univariate ANOVA indicated a significant difference in final test performance as a function of group, $F(3, 253) = 21.03$, $MSE = 0.08$, $p < 0.001$, $\eta^2 = 0.20$. Post-hoc Tukey t-tests indicated that participants in the keyword group recalled significantly more word pairs on the final test compared to participants in the retrieval, restudy, or control group (all p-values < 0.001). Additionally, participants in the retrieval group recalled significantly more words than participants in the control group ($p = 0.009$). The control group and the restudy group did not significantly differ in terms of final recall performance ($p = 0.111$). Finally, the retrieval group did not significantly differ from the restudy group ($p = 0.761$).

RAT Performance

RAT performance indicated that participants could solve approximately one-quarter of the problems correctly ($M = 0.27$, $SD = 0.20$). A one-way univariate ANOVA found no significant difference in RAT performance as a function of group, $F(3, 253) = 0.83$, $MSE = 0.04$, $p = 0.479$, $\eta^2 = 0.01$. Given that participants were randomly assigned to groups and that this is a proxy for creativity (which should be evenly distributed amongst our groups), this is unsurprising.

We treated participants who were above the average RAT performance score as "high creativity" individuals ($n = 118$) and individuals who were below the average RAT performance score as "low creativity" ($n = 139$). We then analyzed final test performance as a function of RAT performance (high versus low) and group (control, retrieval practice, restudy, or keyword) (see Figure 2). There was a significant main effect of RAT performance on final test performance, suggesting that individuals who scored higher on the RAT creativity task tended to score higher on the final test, $F(1, 249) = 8.00$, $MSE = 0.08$, $p = 0.005$, $\eta^2 = 0.03$. There was also a main effect of group on final test performance, $F(3, 249) = 21.80$, $MSE = 0.08$, $p < 0.001$, $\eta^2 = 0.21$. Most importantly, there was not a significant interaction between RAT performance and group on final test performance, $F(3, 249) = 0.18$, $MSE = 0.08$, $p = 0.909$, $\eta^2 = 0.00$.

These analyses highlight several key findings. First, higher RAT performance is associated with better final test performance. Second, final test performance changed as a function of what participants did during practice, with the keyword group performing the best overall. Third, the lack of a significant interaction between RAT performance and group on final test performance suggests that the learning strategy used does not necessarily depend upon the creativity of the individual. Rather, more creative individuals tend to perform better on the final test regardless of which learning strategy they use during practice. Additionally, the keyword strategy appears to be effective regardless of RAT performance.

Discussion

Given the finding that no significant interaction existed between group and RAT performance on the final test, our hypothesis that people with different levels of creativity would benefit more from different methods appears incorrect. Instead, creativity, measured by an ability to see a connection between disparate things, acts as a major factor no matter what method of study the person uses. However, a future study might explore the same question using an alternate measure of creativity. This study used the Remote Associations Triad test to

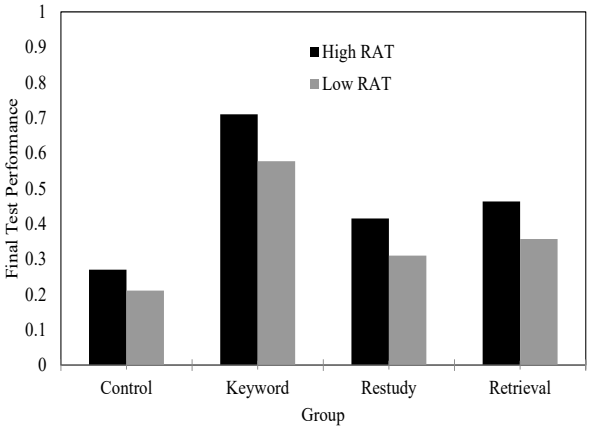

Figure 2. Final test performance (reported as a proportion correct) as a function of RAT performance (High versus Low) and group (control, keyword, restudy, and retrieval).

ascertain the creativity of participants. Perhaps this does not provide the best measure of creativity, and, if the study were attempted with a better measure, we would receive a better idea of our hypothesis' accuracy.

The importance of creativity, regardless of method, may help explain Jenpatturakul's (2012) findings. The participants claimed the keyword method (as opposed to the list method) enhanced their creativity. It is possible the participants were correct, or perhaps, rather than enhancing creativity, it only stimulated the creativity already present. Nonetheless, with creativity serving an important role in test performance, it would still be worthwhile to know if certain methods stimulate it. Perhaps a future study could explore whether the keyword method indeed enhances or activates creativity.

The lack of significant difference between the recall and restudy groups is unexpected. The general body of research states that retrieval practice is superior to restudying in promoting performance. Roediger and Karpicke (2006) conducted a literature review and found that retrieval practice has proven effective many different times and in many different scenarios. Yet our study failed to replicate this expected "testing effect." It seems unlikely that the blame rests on our study's sample size. Even after excluding participants who had already taken part in the study, had not finished, or had requested exclusion, the number left, divided among the experimental groups, resulted in approximately 66 participants per group. It is uncertain why the study came up with this surprising finding.

On an unrelated note, the participants' mnemonic methods in the keyword condition did not strictly match the traditional keyword method. The participants more commonly used phrases than single words to assist their memory; we also cannot be certain if the participants used imagery to link the words, which is considered (according to at least some researchers) to be part of the keyword method. A future study could attempt to implement something closer to the classic definition of the keyword method or simply change the focus of the study from the keyword method to mnemonics in general.

References

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. American Psychologist, 30(8), 821-828. http://dx.doi.org/10.1037/h0077029 Accessed 8/7/2017.

Batsell, Perry, Hanley, and Hotstetter (2016) Ecological validity of the testing effect: The use of daily quizzes in Introductory Psychology. Teaching of Psychology, 44(1), 18-23. https://doi.org/10.1177/0098628316677492

Bowden, E. M., and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. Behavior Research Methods, Instruments, & Computers, 35(4), 634-639.

Jenpatturakul, W. (2012) The impact of keyword technique on the students' vocabulary retention ability in an EFL class. Mediterranean Journal of Social Sciences, 3(3), 565-573. Doi:10.5901/mjss.2012.v3n3p565

Mulligan, N. W., and Picklesimer, M. (2016). Attention and the testing effect. Journal Of Experimental Psychology: Learning, Memory, And Cognition, 42(6), 938-950. doi:10.1037/xlm0000227.

Peterson, D. J., and Mulligan, N. W. (2013). The Negative Testing Effect and Multifactor Account. Journal Of Experimental Psychology: Learning, Memory, And Cognition, 39(4), 1287-1293.

Pierce, B. H., and Hawthorne, M. J. (2016). Does the testing effect depend on presentation modality?. Journal Of Applied Research In Memory And Cognition, 5(1), 52-58. doi:10.1016/j.jarmac.2016.01.001.

Piribabadi, A., and Rahmany, R. (2014) The effect of the keyword method and word-list method instruction on ESP vocabulary learning. Journal of Language Teaching and Research, 5(5), 1110-1115.

Roediger, H.L., and Karpicke, J.D. (2006). The power of testing memory: Basic research and implications for educational practices. Perspectives on Psychological Science, 1(3), 181-210.

Tulving (1967). The effects of presentation and recall of material in free-recall learning. Journal of Verbal Learning and Verbal Behavior, 6(2), 175-184. https://doi.org/10.1016/S0022-5371(67)80092-6

Van Gog, T., and Kester, L. (2012) A test of the testing effect. Cognitive Science, 36(8), 1532-1541. doi.org/10.1111/cogs.12002

Appendix A

Lithuanian-English word pairs	
smuikas	violin
bugnas	drum
kunigas	priest
plaukas	hair
kambarys	room
arbata	tea
mygtukas	button
ugnis	fire
pyragas	cake
durys	door
stogas	roof
pastatas	building
vejas	wind
medis	tree
tinklas	net
batas	shoe
daina	song
sesuo	sister
namas	house
upe	river

Appendix B

Remote Association Triad (RAT) problems	
Stick Light Birthday	Candle
Note Dive Chair	High
Rock Times Steel	Hard
Barrel Root Belly	Beer
Notch Flight Spin	Top
Salt Deep Foam	Sea
Playing Credit Report	Card
Broken Clear Eye	Glass
Gold Stool Tender	Bar
Falling Actor Dust	Star