Production Effect in Complex Texts and Over Time

**Camryn S. Kline1**

1 Grace College, Winona Lake, USA

E-mail: klinecs@grace.edu

ORCID ID: https://orcid.org/0000-0001-6668-0326

Received: 12/12/2018

Accepted for publication: 6/17/2019

Published 11/15/2019

**Abstract**

This study was created to test the production effect in complex texts and the effect over time. Production effect arises when an individual reads something aloud and retains information more efficiently. 123 Grace College and Theological Seminary undergraduate students participated in the study over the course of three weeks. Participants were separated into four groups: reading silently with no delay before testing, reading aloud with no delay before testing, reading aloud with a 24-hour delay before testing, reading silently with a 24-hour delay before testing. Our study shows no significant differences between reading complex texts aloud or silently, as measured by both of our test results. Results could be impacted by our population and other limitations of our design.

Keywords: production effect, reading aloud, reading silently, reading, learning, complex texts, studying

**1. Introduction**

*1.1 Review of Literature*

It is common knowledge that studious individuals have long attempted to find optimal study habits for school, trades, and competitions. Studies have shown that reading aloud, otherwise known as the production effect (PE), can increase how much a person remembers (Bodner & Taikh, 2012; C. M. MacLeod & Bodner, 2017). Results have also been further researched to determine whether production increases recognition and/or recall. (Fawcett & Ozubko, 2016; Jones & Pyc, 2014; Lin & MacLeod, 2012; Mama & Icht, 2016b; J. Ozubko, Gopie, & MacLeod, 2012; Taikh & Bodner, 2016). Many tests have measured how it impacts recall or recognition of words. (Fawcett & Ozubko, 2016; Mama & Icht, 2016b). In contrast, there are not many studies that explore the extent of the effect when it comes to complex texts instead of individual words. The present study will further investigate that aspect of PE.

MacLeod defined the term “production effect” as the effect that takes place when people remember the words that they have read aloud more than the words that they have read silently (C. M. MacLeod & Bodner, 2017). The production effect can also ensue when words are written, typed, sung, or read aloud by another participant (C. MacLeod, 2011; Mama & Icht, 2016b; Quinlan & Taylor, 2013). There are a few different theories on why this effect takes place including distinctiveness, strength, list-discrimination bias, and strategy effect. While the PE can be utilized by all people, there has been research that it has a lesser effect on the elderly (Lin & MacLeod, 2012). This effect, if further explored, could help create better study habits for all people.

Some studies have attempted to find whether the distinctiveness account or the strength account is the most direct cause of the PE (Bodner, Jamieson, Cormack, McDonald, & Bernstein, 2016; Lin & MacLeod, 2012; Mama & Icht, 2016a; J. D. Ozubko, Major, & MacLeod, 2014; Taikh & Bodner, 2016). Distinctiveness supporters would say that the PE only comes into play when it is a mixed list, or when the words read aloud are unique compared to others (C. M. MacLeod & Bodner, 2017). A mixed list would have both words that would be read aloud and words that would be read silently. The findings have generally indicated that the distinctiveness account is the account that underlies most experiments (Bodner & Taikh, 2012; Mama & Icht, 2016a; J. D. Ozubko et al., 2014); however, results show for within-group experiments, the strength test has bearing (Taikh & Bodner, 2016).

The strength account claims that words said aloud are recalled and recognized more easily because the act of speaking them usually incorporates two or more physical senses (Taikh & Bodner, 2016). While not debating that producing words does increase recall and recognition of them, supporters of the strength account suggest that if there were participants who read the silent words multiple times, yet only read the aloud words once, people would be likely to assign the twice read silent words to the “aloud” category because they would have the same strength in their memory as read aloud words (Jones & Pyc, 2014).

Lin and MacLeod (2012), Mama and Icht (2016a), and Taikh and Bodner (2016) discovered list-discrimination bias as an impact factor (Bodner & Taikh, 2012). The idea behind this discriminatory bias is that those participating in a study would assign words they did not recognize to the “silent” category. This is because the PE seems to cause participants to remember words at an increased percentage, so any words that did not appear familiar would seem to belong to in the list that is not read aloud. The way researchers aware of this effect defended against it was to create a new words category that people could put unrecognized words in, instead of assuming the words must have been read silently (Bodner & Taikh, 2012).

The strategy effect was considered possibly confounding to the production effect research in Taikh and Bodner’s (2016) study with a between-group design. Strategy effect would imply that participants would have prior set beliefs about whether they would remember produced or silently read words more (Taikh & Bodner, 2016). This would lead to a kind of list-discrimination bias, where the participant would place words they remember most into the list they believed to be the most effective, and the words they are unsure about would be assigned to the category they would assume they would not remember as well (Lin & MacLeod, 2012; Taikh & Bodner, 2016).

A physiological study was conducted to determine which theory is closest to the cause of production (Hassall, Turk, Krigolson, Quinlan, & Taylor, 2016). Researchers Hassall et al. (2016) used electroencephalography to map how the brain was working during the memory exercises. The final brain scans recorded while recreating the Quinlan & Taylor (2013) study found a visible difference and indication of the PE when comparing reading silently and reading aloud or singing (Hassall et al., 2016). However, they found differing results than Quinlan & Taylor (2013) as there was no significant difference or indication of the PE between reading aloud and singing the words, negating, at least in that study, the distinctiveness effect (Hassall et al., 2016). The study does admit that some differences in how they administered treatment may be the cause of the discrepancies (Hassall et al., 2016).

Another important distinction to note is between recognition and free recall. Recognition is simply remembering having seen the shown object or word before. Free recall is the ability to name what was read with no cue given by the confederate. Studies have shown that the effect that production has on the two is not connected and depends on text and technique (Fawcett & Ozubko, 2016; Jones & Pyc, 2014; Lin & MacLeod, 2012; Mama & Icht, 2016b; J. Ozubko et al., 2012; Taikh & Bodner, 2016).

One of the few studies on age and the PE is by Lin & MacLeod (2012). They found the PE was indeed applicable to all ages; however, the effect was significantly smaller in the elderly in a study involving 17-25 year olds and 67-88 year olds. That leaves a large portion of the public unaccounted for and reduces the study’s external validity. Research utilizing a wider range of ages should be conducted.

To align with previous studies, the population for the current study is 17-25 years old, the ages of most undergraduates enrolled at universities (Bodner & MacLeod, 2016; J. D. Ozubko et al., 2014; J. Ozubko et al., 2012; Lin & MacLeod, 2012). The decision to limit the population is to reduce any threat in replicity. However, more studies should be done in relation to how the PE impacts different age groups, such as middle-aged people. The current researcher also believes that the strength effect is the most plausible reason for the PE to occur.

*1.2 Independent Variable*

In the present study, there were four levels of independent variable. The first two levels of the manipulated variable were reading aloud, or the PE, with half of the participants reading an article aloud and the other half reading the same article silently. These two levels should be an acceptable addition to the miniscule research involving the PE and its occurrence in large text reading.

One variable that may confound the studies of the PE on complex texts is that in the studies mentioned, with the exception of J. D. Ozubko, Hourihan, & MacLeod (2012), experimenters administered their test of recognition or recall immediately following the treatment. Besides leaving the question of the longevity of the effect unanswered, if the length of a passage causes more time to pass before the participant would be tested, it could explain some results found in place of just the PE (J. D. Ozubko et al., 2012).

The PE has already been shown to be effective when studying mixed and, arguably, pure lists of words (Bodner et al., 2016; Jones & Pyc, 2014; Taikh & Bodner, 2016). In a study by J. D. Ozubko et al. (2012), they focused on how the PE would assist students in memorizing sentences and essays and how the effect lasts when there is a delay in testing. They discovered that the PE does in fact apply to longer and more complex pieces of text and remains even if there is a delay between the reading and the tests of recognition and recall (J. D. Ozubko et al., 2012).

To prevent the extraneous variable from causing threat to the internal validity of the present study, groups with a delay between treatment and testing were added. Staying consistent with J.D. Ozubko et al. (2012), who studied delay in lists and larger text, the delay groups that read their text aloud had a twenty-four-hour delay between their reading and their test. This helps answer the question of the PE’s longevity.

A second possible confounding variable is background noise (Mama, Fostick, & Icht, 2018). Mama et al. (2018) found that when there was noise of changing rhythm or volume, such as talking, even if what was being spoken was babble, it interfered with the production effect. To prevent this potential obstacle, all students in the present study were put into separate rooms where they could not hear one another regardless of assigned group.

When given textbook chapters to read and remember for a class, students may read silently in their dorms, thinking that maximum retention will take place if they can sit and force themselves to focus on the text. However, the PE suggests that reading silently holds the smallest effect when it comes to memory, even in comparison to words read by others and typed words (C. M. MacLeod & Bodner, 2017). Bodner & MacLeod (2016), J. D. Ozubko et al. (2014), and J. Ozubko et al. (2012) found students who took a fill in the blank test after reading textbook passages aloud the previous day did better on the test than students that had not.

Mama & Icht (2016b) claimed that the more encoding processes, such as reading, hearing, or speaking, that take place when studying a word, the deeper the memory will bury itself in the brain, explaining why production would have an effect and supporting the strength effect. They found that hearing words could possibly be even more advantageous than seeing words that are on a screen (Mama & Icht, 2016b). They claimed that regardless of the recall or recognition test, words that are heard are remembered more. Mama & Icht (2016b) and Quinlan & Taylor (2013) suggest this happens because the information is triggering multiple parts of the brain; the more sections of the brain that are involved, the more information is retained. This idea is partly inconsistent with the previously mentioned study by Hassall et al. (2016) because there was no significant difference or indication of the PE between reading aloud and singing the words.

Fleck et al. (2017) tested active reading compared to silent reading. The study did not involve reading aloud, but instead studied implementing techniques, such as active reading questions placed in intervals throughout the textbook chapters (Fleck et al., 2017). The results supported Mama and Icht’s (2016b) statements that active reading techniques have superior encoding processes, so they increase the volume of information that is retained.

*1.3 Dependent Variable*

There are some conflicting views on whether the PE assists in free recall, recognition, or both. Jones and Pyc (2014) found that there is little to no evidence for the PE in studies of recollection, and in some cases, inconsistent results in the effect on recognition. Fawcett and Ozubko (2016) insisted there is no effect that occurs in recollection, especially in between-subject designs. C. MacLeod (2011) stated there is strong support for the occurrence of the PE in recognition. However, Lin and MacLeod (2012) and J. Ozubko et al. (2012) claimed there is support for the PE in recognition and recollection.

The scoring of the free recall test in the present study was similar to the procedures used by J. Ozubko et al. (2012). A test of 14 fill in the blank questions was administered either immediately after reading or after a delay. Correct answers received one point, and answers that demonstrate the general idea of the correct answer received half of a point (J. D. Ozubko et al., 2012). The present study also added a simple yes/no test of recognition that mirrors the test used by Hassall et al. (2016). The current study did not include a “new” category in the recognition test, unlike Bodner & Taikh (2012) and J. D. Ozubko et al. (2014). This is because no new information was created, and participants were asked to judge whether or not they read the sentences at all.

Recollection tests, in most cases, have been done by distributing a fill-in-the-blank test that requires a participant to write what they can remember without cues from experimenters (Lin & MacLeod, 2012; J. Ozubko et al., 2012). Some cases give partial scoring for partially correct or generally correct answers (J. D. Ozubko et al., 2012; J. Ozubko et al., 2012). The goal of recollection tests is not to find if the participant recognizes the material but to find if they have memorized it.

Recognition tests typically will show words to participants who must indicate whether they have studied the revealed word or not (Lin & MacLeod, 2012). Some studies incorporate a “new” category to negate the list-discrimination or strategy effect (Bodner & Taikh, 2012; J. D. Ozubko et al., 2014). On average, more silently read words were misplaced than those read aloud (Bodner & Taikh, 2012).

A confounding variable for both tests could be if the colors for the words that are written are not counterbalanced. If text is harder to read against the background color, the need to focus more on the words to read them could be the explanation for the results, instead of just the PE (J. D. Ozubko et al., 2012). Certain colors may be more mentally stimulating than others. This was a problem for Lin and MacLeod (2012).

To prevent that confounding variable, texts were the same for each group in the present study. Instead of colors representing the task, instructions from the researcher were given vocally to read the passages aloud or silently. Tests and passages were also given on the same color paper and in the same font.

*1.4 Conclusion*

The research initiated by MacLeod and others has been further developed to find constraints and theories for the causes of PE. The main theories for the origin of the effect are the distinctiveness account, strength account, list-discrimination bias, and the strategy effect, with the distinctiveness account being the strongest. There is an argument that the PE occurs because of the weakening of silently read words, not the strengthening of produced words. They call this the reverse PE (C. M. MacLeod & Bodner, 2017).

Potentially confounding variables for experiments studying the PE include biases that lead to list-discrimination and strategy effect, which can be avoided with a category for new words, administering the test as soon as the treatment is over. This could explain some of the results instead of the studied effect, and not counterbalancing the cues given on how to read the images (Lin & MacLeod, 2012; J. D. Ozubko et al., 2012; Taikh & Bodner, 2016). More studies should be done to test the effects of delay, harder-to-read font, and biases on the PE results.

The four levels of independent variable that the present study explores include reading aloud, reading aloud with a delay until testing, reading silently, and reading silently with a delay until testing. The PE has proven significant in lists of words but has been underapplied to complex texts such as essays, articles, or even paragraphs (Bodner et al., 2016; Jones & Pyc, 2014; Taikh & Bodner, 2016). Each level of the independent variable in the current study was tested on two separate tests of recognition and recall.

The dependent variable was made up of a test of recognition and a test of recollection as there needs to be more evidence to support the PE’s occurrence in both levels. The present study was recreating the familiarity and free recall tests done by J. Ozubko et al. (2012), Bodner & Taikh (2012), J. D. Ozubko et al. (2014), and Hassall et al. (2016).

**2. Experiment 1: Production Effect in Complex Text**

*2.1 Method*

*2.1.1 Participants*

In the current experiments, undergraduate students (mostly made up of behavioral science majors) from Grace College and Theological Seminary were asked to volunteer in a short study. Participants were recruited over a one-week period by the researcher or by professors who informed them of the opportunity to earn extra credit in their behavioral science classes. They were incentivized by extra credit if they finished participation in the study; the amount was decided upon by the professors of those individual courses. There were some students who chose to participate with no incentive. A total of 164 students signed up to participate; however, only 123 did so. They were randomized into four groups: Reading Silently with no Delay (RSND), Reading Aloud with no Delay (RAND), Reading Silently with Delay (RSD), and Reading Aloud with Delay (RAD). The randomization process was done with a die and not every person who signed up participated, so each group is not necessarily equal (1=RSND, 2=RAND, 3=RSD, 4=RAD, 5 & 6=Reroll). This process leads to 34 students in the RSND group, 32 in the RAND group, 28 in the RSD group, and 23 in the RAD group

*2.1.2 Stimuli and Design*

The article “Feng-Shan Ho” (2017) was chosen for its interesting content and its clear language. It was printed onto white paper in black, Arial, 12-point font. The individual tests were on white paper in black, Calibri, 11-point font.In the initial portion of this between-group study participants were divided into two groups: RAND & RSND.

*2.1.3 Procedure*

Participants were directed into rooms in the Behavioral Science offices. Each individual was instructed to read the article “Feng-Shan Ho”(2017). Depending on which group the individual was randomly placed it, they were told to read the article aloud or silently. They were additionally instructed to read at a comfortable pace. They were told before reading the article they would receive one test when they were finished reading and another after completing the first test. Whenever a task was completed, the person administering the test would take the paper and provide the participant with the next task. Participants received instructions before reading and before each test on how to continue.

The first test administered was the free recall test to reduce any influence the tests may have on one another. One free recall question was chosen for every paragraph in the article, and efforts were made to avoid repeating subjects between the free recall and recognition test; this resulted in 14 questions.

The second test was given immediately following the completion of the first. It was a test of recognition made up of 10 questions. Students were instructed to circle “Yes” or “No” depending on whether they remember reading the sentences in front of them. The unread/incorrect sentences were slightly altered real excerpts from “Feng-Shan Ho” (2017). No sentences were completely made up by the researchers. Once finished, all participants were given a written debrief and were free to leave.

*2.1.4 Results and Discussion*

A two-sample T-test was performed to determine whether or not there were significant differences in academic performance between reading aloud and reading silently with no delay. Within the recognition, there were no statistically significant differences between the data collected from the two groups. While the RAND group had a higher mean (9.72) than the RSND group (9.32), the P-Value was .389. Our hypothesis was not supported by results in the free recall test, either. Again, the mean on the RAND group (6.34) was higher than the RSND group (5.71) but not significantly so (P-Value = .405). Combining the results of the tests still yielded no significance, with a P-Value of .335. Our hypothesis is not supported by the current evidence, but neither can it be disproved. The results may be representative of a small sample size. There are still conclusions to be drawn from our results. First, it does not appear that the PE occurs when reading complex texts. Neither free recall nor recognition was statistically impacted by reading aloud.

**3. Experiment 2: Production Effect Over Time**

*3.1 Method*

*3.1.1 Stimuli and Design*

The stimuli and testing material were identical for the second experiment, with the only difference being the time of administration. In this experiment, participants were separated into two groups: RAD & RSD. Apart from the 24-hour delay between reading and testing, there were no differences between these groups and their no-delay counterparts.

*3.1.2 Procedure*

Participants were again divided amongst rooms. To maintain consistency, when they returned for testing, students were put in the same room they had read in. Each individual was instructed to read the same article as the no-delay groups. They were told privately, before receiving the article, that they were in a delay group. If they could not participate the following day as well, they were given the option to come back on another day that testing was taking place. Those who indicated they could return the next day were given the article to read and told they were free to leave once finished. They were given the same instructions for the two consecutive tests as the no-delay groups.

*3.1.3 Results and Discussion*

A two-sample T-test was used to analyze the significance of our results. We found no significant differences in test performance between the reading aloud and the reading silently groups after a delay. Surprisingly, the reading silently group had a higher mean (9.04) in the recognition test than the reading aloud group (8.65), yet there was still no significant difference between these two means (P-Value = .389). The two groups had near identical means in the free recall test (RAD = 3.7, RSD = 3.8, P-Value = .887). Combining results still yielded no significance with a P-Value of .614. Our hypothesis was not supported; however, the results may still be impacted by the study’s limited sample size and were not significant enough to disprove our hypothesis. It is not surprising there were no relevant findings in this experiment, as we found no significance in the prior experiment’s data.

**4. General Discussion**

*4.1 Limitations*

We were not able to replicate the findings in J. D. Ozubko et al. (2012). This could be explained by a number of factors, however. J. D. Ozubko et al. (2012) based their study at the University of Waterloo in Canada; they also used a within-group design and formed their tests with the assumption the distinctiveness account underlied PE. We had an overall smaller sample size, which was made up of predominantly white, Christian, undergraduate students. A study with a more diverse group of participants may see different results. It would be beneficial to increase diversity in race, ethnicity, and age. Results from a study with such a limited sample group are difficult to apply to a larger, more general population. There was a discrepancy in fonts between the article and the tests. This was an oversight, and most likely had little to no effect, but it could have negatively affected recognition. Our study assumed that reading aloud would strengthen the memory of the text in participants’ minds without any comparison. However, there are other explanations for the production effect, which, when utilized, may produce an effect in complex text.

*4.2 Application*

Even with these limitations, production effect has been shown to be effective in lists of words. These techniques could be implemented in other areas of studying such as vocabulary lists, the memorization of anatomy parts, or other groups of single-word information. In elementary education, active reading strategies are already being used to increase vocabulary retention (Gallagher & Anderson, 2016). Understanding the best ways to incorporate reading aloud into classroom activities can assist teachers in better equipping their students.

Secondary school students may also find it useful to read aloud if they must memorize groups of words. This information matters because it could help students and other individuals memorize information more quickly and effectively. College students are under pressure to do well in their area of study and a new study tool could assist in relieving this stress.

While our hypothesis was not supported, nor was it disproved. Further research may indicate that there is such an effect in complex texts, and these findings would increase the relevancy of PE to those in higher education. Finding PE effective in complex texts would allow students to study for less time as they may have to read over material fewer times.

The study of PE is relatively new (Bodner & MacLeod, 2016). It is also important that the reasons behind the PE be further explored in order to eliminate any false causes. Narrowing what causes it can allow us to structure studies in a more effective way. As we learn more about what causes it and what does not, we will come closer to enhancing learning experiences for all students.

*4.3 Future Research*

We would hope to see this research repeated in the future with larger and more diverse sample sizes. All aspects of the effect cannot be understood if it is not applied to all types of people. We believe it would be beneficial for research on PE to be done with those who are middle-aged as the group has yet to be represented. If the intensity of the effect were to wane with age, the study technique would not work as well for those going back to school as adults, for teachers, etc. It would be interesting to test how volume and speed of projection would affect outcomes both in lists of words and full texts. Bodner & MacLeod (2016) emphasized the benefits of researching whether reading some of the material silently and some of it aloud would create an effect. If this was found to have an effect, what would be the balance best suited for retention (Bodner & MacLeod, 2016)?

**Acknowledgements**

This study could not have occurred without the opportunities created by Dr. Lisa Wooley at Grace College.

**References**

1. Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, D.-L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 93–98.
2. Bodner, G. E., & MacLeod, C. M. (2016). The benefits of studying by production . . . And of studying production: Introduction to the special issue on the production effect in memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 89–92. https://doi.org/10.1037/cep0000094
3. Bodner, G. E., & Taikh, A. (2012). Reassessing the Basis of the Production Effect in Memory. *Journal of Experimental Psychology. Learning, Memory & Cognition*, *38*(6), 1711–1719. https://doi.org/10.1037/a0028466
4. Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not Recollection, Supports the Between-Subject Production Effect in Recognition Memory. *Canadian Journal of Experimental Psychology*, *70*(2), 99–115. https://doi.org/10.1037/cep0000089
5. Feng-Shan Ho. (2017). In *Encyclopedia of World Biography* (Vol. 37). Detroit, MI: Gale. Retrieved from http://link.galegroup.com/apps/doc/K1631010630/BIC?u=palni\_grshcol&sid=BIC&xid=a300ce10
6. Fleck, B., Richmond, A. S., Rauer, H. M., Beckman, L., & Lee, A. (2017). Active reading questions as a strategy to support college students’ textbook reading. *Scholarship of Teaching and Learning in Psychology*, *3*(3), 220–232. <https://doi.org/10.1037/stl0000090>
7. Gallagher, M. A., & Anderson, B. E. (2016). Get All “Jazzed Up” for Vocabulary Instruction: Strategies That Engage. *Reading Teacher*, *70*(3), 273–282. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1118054&site=ehost-live
8. Hassall, C. D., Turk, D. J., Krigolson, O. E., Quinlan, C. K., & Taylor, T. L. (2016). A Preliminary Investigation Into the Neural Basis of the Production Effect. *Canadian Journal of Experimental Psychology*, *70*(2), 139–146. https://doi.org/10.1037/cep0000093
9. Jones, A. C., & Pyc, M. A. (2014). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 300–305. https://doi.org/10.1037/a0033337
10. Lin, O. Y. H., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(3), 212–216. https://doi.org/10.1037/a0028309
11. MacLeod, C. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*(6), 1197–1202. https://doi.org/10.3758/s13423-011-0168-8
12. MacLeod, C. M., & Bodner, G. E. (2017). The Production Effect in Memory. *Current Directions in Psychological Science*, *26*(4), 390–395. https://doi.org/10.1177/0963721417691356
13. Mama, Y., Fostick, L., & Icht, M. (2018). The impact of different background noises on the Production Effect. *Acta Psychologica*, *185*, 235–242. <https://doi.org/10.1016/j.actpsy.2018.03.002>
14. Mama, Y., & Icht, M. (2016a). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals the superiority of writing over vocalising. *Memory*, *24*(1), 98–113. https://doi.org/10.1080/09658211.2014.986135
15. Mama, Y., & Icht, M. (2016b). Influence of retrieval mode on effects of production: Evidence for costs in free recall. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 177–185. https://doi.org/10.1037/cep0000090
16. Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*(7), 717–727. https://doi.org/10.1080/09658211.2012.699070
17. Ozubko, J. D., Major, J., & MacLeod, C. M. (2014). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*, *22*(5), 509–524. https://doi.org/10.1080/09658211.2013.800554
18. Ozubko, J., Gopie, N., & MacLeod, C. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*(3), 326–338. https://doi.org/10.3758/s13421-011-0165-1
19. Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory*, *21*(8), 904–915. https://doi.org/10.1080/09658211.2013.766754
20. Taikh, A., & Bodner, G. E. (2016). Evaluating the basis of the between-group production effect in recognition. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *70*(2), 186–194. https://doi.org/10.1037/cep0000083