Production on hold: delaying vocal production enhances the production effect in free recall

Yaniv Mama & Michal Icht

To cite this article: Yaniv Mama & Michal Icht (2017): Production on hold: delaying vocal production enhances the production effect in free recall, Memory, DOI: 10.1080/09658211.2017.1384496

To link to this article: http://dx.doi.org/10.1080/09658211.2017.1384496
Production on hold: delaying vocal production enhances the production effect in free recall

Yaniv Mama a and Michal Icht a,b

aDepartment of Behavioral Science and Psychology, Ariel University, Ariel, Israel; bDepartment of Communication Disorders, Ariel University, Ariel, Israel

ABSTRACT
The Production Effect (PE) represents superior memory for produced (read aloud) relative to non-produced (silently read) items. Another method of improving memory is taking a test on the study material – the Testing Effect. We evaluated the combined influence of both effects on free recall memory, using delayed vocal production, in which study words were vocally produced only after their disappearance. Such procedure involves an initial instant test since participants had to vocally retrieve the words (rather than read them aloud). In five experiments, participants were presented with study words that they were instructed to learn by non-production (reading silently), immediate production (reading aloud), delayed reading aloud, or delayed vocal production (instant retrieval). The results showed superior recall for delayed production over all other conditions. We suggest that the source of this superiority is the desirable difficulty induced by the addition of the initial test (retrieval effort) to the vocal production. The novel delayed production condition forms a superior mnemonic.

The production effect (PE) represents an improvement in explicit memory (recall, recognition) for items that were read aloud (vocalised) relative to items that were read silently during study (Forrin, MacLeod, & Ozubko, 2012; MacLeod, 2011; MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; Mama & Icht, 2016b; Ozubko, Hourihan, & MacLeod, 2012). It occurs mainly when the produced items are presented in a single mixed-study list, with other, non-produced items (MacLeod et al., 2010; for a small effect in between-subjects experimental designs with recognition tests, see Fawcett, 2013). The PE has been documented with several types of study material, such as pictures (Icht & Mama, 2015), non-words (MacLeod et al., 2010) and text (Ozubko et al., 2012). It is evident across the lifespan, for children (Icht & Mama, 2015), younger and older individuals (Lin & MacLeod, 2012), and special populations (e.g., cochlear implant users – Taitelbaum-Swead, Icht, & Mama, 2017; dysarthric adults – Icht, Bergerzon-Biton, & Mama, 2016).

The PE is usually attributed to encoding distinctiveness (e.g., Forrin et al., 2012; MacLeod, 2011; MacLeod et al., 2010). According to this account, the PE generates from relatively rich encoding at study. As the number of distinct encoding processes involved in learning (e.g., visual processing of reading; motor processing of articulating; auditory processing of hearing) is larger, the memory benefit is bigger (Forrin et al., 2012; Mama & Icht, 2016a). Following this logic, saying a word aloud provides for additional processes of encoding relative to silent reading. The rich encoding is diagnostic with respect to the item’s status at study when testing memory, either by recognition or by free recall.

Another explanation for the benefit of vocal production emphasises the qualitative perspective (rather than the quantitative aspect of the number of unique encoding processes). MacLeod (2011) stresses that one’s own voice is personal and unique. Thus vocal self-production results in better memory relative to other types of productions involving a similar number of encoding processes (e.g., joint production with another person). In other words, “Recollec-
tion of one’s own production at the time of test is optimally distinctive” (p. 1201).

Reading aloud (vocalising) is considered a simple and easy-to-perform act, and usually results in a significant memory advantage. Hence, it was offered as a mnemonic device – a technique for improving retention at studying (Icht & Mama, 2015; Lin & MacLeod, 2012; Ozubko et al., 2012). MacLeod et al. (2010) noted that “production is so very simple, straightforward, and effortless that it should certainly be better recognized among manipulations that benefit memory” (p. 672). The current study addresses a pair of related issues: (a) from the theoretical perspective – a better understanding of the mechanisms underlying the memory benefit, and (b) from an applied perspective – examining the possibility to further enhance the PE. Addressing these points, we employed a minor alteration from the typical PE design in order to produce larger memory benefits. This change is based on another
simple technique that has a robust effect on learning, testing.

**The testing effect**

Taking a test on studied material enhances subsequent learning and retention of that material on a final test (Bartlett, 1977; Darley & Murdock, 1971; Karpicke & Roediger, 2008). In other words, repeated retrieval practice (multiple tests) was found to enhance long-term retention. It also has a greater effect on slowing forgetting than does a single test (Wheeler & Roediger, 1992). This mnemonic benefit of testing is referred to as the testing effect (Masson & McDaniel, 1981; McDaniel, Anderson, Derbish, & Morissette, 2007; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985; Whitten & Bjork, 1977).

Surprisingly, testing seems a more powerful learning device than additional (repeated) study of the target material (Carrier & Pashler, 1992; Karpicke & Roediger, 2008; Kuo & Hirshman, 1996; Roediger & Karpicke, 2006b). Immediate retrieval of once-studied target items was found to benefit performance on subsequent tests more than did another study presentation of the target material (Hanawalt & Tarr, 1961; Hogan & Kintsch, 1971; McDaniel & Masson, 1985).

The testing effect can be found across materials, and was observed with word lists (Hogan & Kintsch, 1971; McDaniel & Masson, 1985), paired associate lists (Allen, Mahler, & Estes, 1969; Carrier & Pashler, 1992), pictures (Wheeler & Roediger, 1992), and prose material (Glover, 1989), for both recall and recognition tests (McDaniel et al., 2007). The effectiveness of testing has been documented across a variety of study–initial test intervals, from seconds (an initial recovery; Bartlett, 1977; Kuo & Hirshman, 1996) to minutes (Karpicke & Roediger, 2008).

Common explanations of the testing effect are based on deep or effortful retrieval (the retrieval hypothesis, or retrieval effort – see a review in Roediger & Karpicke, 2006b). According to these models, depth of retrieval may operate similarly to depth of processing at encoding – the greater the retrieval effort, the stronger the memory. Hence, any procedure that increases retrieval efforts strengthens the memory trace and promotes better retention and long-term learning (Bjork & Bjork, 1992).

The positive effect of taking a test on long-term memory can be explained in term of desirable difficulties (Bjork, 1994, 1999). Several study methods that create difficulties at study (e.g., they may slow initial learning), were found to yield more durable and flexible learning, and to enhance long-term memory (retention). Such difficulties may include varying the conditions of learning, rather than keeping them constant and predictable, or using spaced practice rather than massed practice. Using tests as a learning event (rather than re-presentation of the material), fits this definition as well (Bjork & Bjork, 2011; Roediger & Karpicke, 2006b). In fact, an initial test (or retrieval attempt), even when no corrective feedback is given, can be more effective in the long term than repeated reading of the study material. According to the desirable difficulties account, as learning is more effortful and difficult, it triggers encoding and retrieval processes that support learning, comprehension, and remembering. Hence, learning outcomes are improved (Bjork & Bjork, 2011).

Overall, both the PE and the testing effect have a robust positive influence on memory. The current study aims to evaluate a novel learning condition, which convolutes both tactics – vocal production and testing – delayed vocal production.

**Introducing a new learning condition: delayed vocal production**

In the current study, a new learning condition is used, in which the participants are asked to briefly delay their vocal productions. Namely, during the study, the participants were transiently presented with each study word, and were required to say it aloud only after its disappearance (to recall the word from short-term memory following an “off time”, upon a given cue). Hence, in this learning condition, saying aloud serves as an instant initial test (in fact, many consecutive single-item initial recall tests. Note that testing can be effectively used even with short study–initial test intervals; Bartlett, 1977). Final recall test (from long-term memory) followed learning, in which we compared memory performance (recall rates) in three experimental conditions: (a) no-production – silent reading of the study words, (b) immediate production – reading aloud (the participants were asked to vocalise the study words as they appeared), and (c) delayed production. Conditions “a” and “b” are typically used in PE studies. The delayed production condition is the novelty of the current study. In a final experiment, we tested another learning condition, (d) delayed reading (in which the participants were required to re-read aloud the study words as they reappeared).

Using the delayed production condition is of theoretical importance in understanding the qualitative aspects of the encoding distinctiveness theory. This original condition entails a similar number of distinct encoding processes as the typical vocal production condition; visual (reading the words), motor (articulating), and auditory (hearing your voice). Thus, according to the quantitative perspective, both conditions will result in comparable memory performance.

However, we assume that delayed production is qualitatively different from immediate production, since the former is more effortful and demanding (i.e., the retrieval act as a desirable difficulty). Enhanced memory for delayed production relative to immediate production would support the notion that the PE carries a qualitative aspect (affected by the type and nature of the encoding processes), and is not only a quantitative-
based phenomenon (stemming from the number of distinct processes).

Evaluating the novel delayed production condition reveals that in addition to testing (retrieval), it entails another pair of features previously found to enhance memory performance: namely, response inhibition (Anderson, 2005), and a reflective action of refreshing (Johnson, 1992). The next sections will discuss their possible role in memory enhancement.

**Delayed vocal production and response inhibition**

The delayed production condition may shed light on another theoretical aspect related to learning – response inhibition. A key behavioural feature of the delayed production is the need to withhold the vocal response until a cue is given. Not only does the participant have to retrieve the word (as it is no longer present – the act of testing), but she must also wait to do so. Thus, besides the testing itself, this condition entails an inhibitory aspect (Anderson, 2005). Since a delay is inserted between a study item and the participant’s response (“off time”), the participant must inhibit the vocal production. Possibly, the act of withholding the vocalisation explains the memory benefit, rather than the act of testing (retrieving a word from short-term memory). Indeed, previous research has demonstrated that response withholding is quite challenging (Kahana & Loftus, 1999) and may improve performance (Koriat & Goldsmith, 1996). For example, Loft and Remington (2013) delayed participants’ task responses by presenting a signal at varying onsets (0–1600 m sec) following the stimulus presentation, indicating when participants could make their responses. The authors showed that delaying the participants’ responses improved prospective memory performance (for similar results, see Loft et al., 2014).

Another theoretical aspect lies within the new delayed production condition. Explicitly, when one delays the vocal response, one has to think of the no longer present word, keeping it active, in order to retrieve it successfully upon a given cue. This elementary maintenance activity was referred to as the Refreshing process (Johnson, 1992). Let us consider this alternative theoretical interpretation.

**Delayed vocal production as a refreshing activity**

Refreshing is the act of briefly thinking of a just-previous activated thought or percept (Chun & Johnson, 2011), a basic cognitive act of reflective attention directed to an item that is no longer perceptually present (Raye, Johnson, Mitchell, Greene, & Johnson, 2007). Refreshing is one of the executive processes described within the Multiple-Entry, Modular memory model (MEM, Johnson, 1992).

In brief, the MEM framework delineates memory in terms of processes or cognitive actions. These mental procedures operate simultaneously and determine long-term memory. They are organised in separate sub-systems, perceptual and reflective. Each consists of various sub-processes, and interfaces with the other (Johnson, Reeder, Raye, & Mitchell, 2002).

Reflection operations are executive, self-generated mental processes, which are post-perceptual. They allow one to maintain, organise, and recover information, and thus are critical in learning and long-term memory performance. Reflective operations strengthen relations among various elements that can cue each other, and combine several related aspects of a previous experience (content and context) to form a single complex memory trace (Yi, Turk-Browne, Chun, & Johnson, 2008).

Refreshing is one of the reflective mental operations within the reflective sub-system. It is a simple process that operates during ongoing cognition, prolonging activation of just-activated representations. The refreshing information allows it to remain active, so one can easily shift back to it (Johnson, 1992). In the lab, a typical refreshing condition requires the participants to think of a previous event (a just-seen word or picture, no longer present; Yi et al., 2008), or to say it aloud (Chun & Johnson, 2011; Johnson et al., 2002). Contrary to a simple repetition, in the refresh condition, participants are asked to “mentally go back” to a perceptually absent item. Therefore, refreshing represents a simple form of imagery by which an active representation is briefly sustained. Johnson and colleagues have shown that refreshing items improve their retrieval from long-term memory, suggesting that this is due to the benefits of reflective attention, which brings back what is no longer perceptually present.

The new experimental condition of delayed vocal production is reminiscent of a refresh condition, as it involves thinking of and saying aloud a previously presented word, no longer present (Johnson et al., 2002). Just as a single refresh activity was found to be sufficient to improve long-term memory (compared to perceiving an item again, Johnson et al., 2002; Raye, Johnson, Mitchell, Reeder, & Greene, 2002), we assume that delayed vocal production will result in superior memory performance (relative to the other experimental conditions).

In sum, the delayed production condition holds the act of testing (desirable difficulty of retrieving), along with response inhibition, and the mental activity of refreshing (keeping active or foregrounding the no longer externally present information). All are known to improve memory. Attempting to determine which one underlies the expected memory benefit of the new condition is another theoretical goal of the current study.

**The present study**

In five experiments, we examined the benefit of vocalising and instant testing on memory performance. In each experiment, the participants were visually presented with to-be-remembered words. In Experiment 1A, these words were learned by two conditions – half were silently read, whereas the remaining half were read aloud (a typical PE
In Experiment 1B, half of the study words were learned by silent reading and the remaining half by delayed production. In Experiments 2 and 3, half of the study words were learned by reading aloud (immediate) and the remaining half by delayed production (3 sec delay in Experiment 2, and 1 sec delay in Experiment 3). Following each study phase, memory was probed by a final free recall test. Results were compared across conditions.

A direct examination of immediate vocal production and the delayed production condition enables a comparison of two mechanisms for memory improvement: testing and increased presentation duration (McDermott & Watson, 2001). Namely, in the immediate production condition, testing is absent (the words are perceptually present and read aloud), but presentation duration is longer (3 sec). However, in the delayed production condition, testing occurs (the words are perceptually absent), but the study words are visible for shorter periods (1 sec). Thus, comparing immediate and delayed production allows us to assess the advantage of testing relative to the gain of increased presentation duration.

According to the quantitative interpretation of encoding distinctiveness theory of the PE (MacLeod et al., 2010), recall rates of silent words (a single visual encoding process) would be lower relative to aloud words (three encoding processes). However, recall rates of immediately produced words and of delayed produced words would be comparable, since both involve a similar number of encoding processes. The desirable difficulties account (Bjork, 1994) predicts different results (in accord with a qualitative interpretation of the encoding distinctiveness theory). Since delayed production (retrieving) is more difficult relative to immediate production (reading aloud), recall rates for the former would be higher than for the latter.

As detailed above, the delayed production condition involves two additional mechanisms that may enhance memory (besides testing): response inhibition and refreshing. Both are related to the presence of an “off time” that occurs in each trial. Experiment 4 was designed in order to distinguish between these factors, directly tapping the source of the expected memory benefit. We compared memory performance between two conditions that involved similar “off times”: delayed reading (in which the study word reappeared, thus no testing ensued), and delayed production (in which only a retrieval cue appeared, thus testing was required). Comparable memory performance in these conditions will imply that the presence of an “off time” (refreshing, response inhibition) is critical, rather than the act of retrieving. However, superior performance in the delayed production condition will support the importance of testing in boosting memory.

**Experiment 1**

In Experiment 1, we evaluated the benefit of delayed production, exploring whether words learned under this novel condition were more memorable than were silently read words. The participants produced words by reading aloud (Experiment 1A) or by delayed vocal production (Experiment 1B), while the remaining words were not produced (silently read). Since any unique, item-specific type of production is relatively more distinct than silent reading, we predicted a PE in both experiments. Yet, since delayed production holds both the benefit of vocalisation (high relative distinctiveness) as well as the advantage of testing, we assumed that this condition would result in superior recall relative to aloud reading, hence larger PE in Experiment 1B than in Experiment 1A.

**Experiment 1A: silent vs. aloud reading (immediate production)**

In this baseline experiment, we used a typical PE paradigm, comparing memory performance for words read silently relative to words read aloud.

**Method**

**Participants**

Forty undergraduate students (29 females, mean age: 24 years) from Ariel University received course credit for participating in the experiment. One participant was excluded since he failed to follow the study instructions, resulting in a final sample of 39 participants. All were native Hebrew speakers, who gave their informed consent to taking part in the study. The study was approved by the local ethics committee.

**Apparatus and stimuli**

The pool of items consisted of 80 common Hebrew words, bi-syllabic nouns, three to five letters long (taken from Icht, Mama, & Algom, 2014). From this pool, 40 words were selected for study, a random sample for each participant. At study, each word was visually presented at the centre of a 19-inch colour monitor of a personal computer, under control of DirectRT programme. The words were presented in black (28-point David), against a white background. In each trial, a small icon (2 cm²) of an eye or a microphone appeared about 5 cm above the study word. The icon indicated the suitable learning condition for that word: Silent reading was indicated by an eye and vocal production by a microphone.

**Design and procedure**

The participants were tested individually in a quiet room. A research assistant was present along the experimental session, verifying that the participants followed the study instructions. Upon arrival, they were told that they would be presented with a list of words to be learned by silent or aloud reading (mixed-study list). They were informed...
that a memory test would follow the presentation of the words. Each participant signed the informed consent form, and then was seated at a distance of 60 cm from the centre of the computer screen.

**Study phase.** The 40 study words were randomly divided into two equal subsets defined by the learning condition – 20 to be read silently and 20 to be read aloud – randomly intermixed and presented individually.

In each experimental trial, a visually presented study word accompanied with an appropriate icon (eye, microphone) appeared for 4 sec. A blank screen followed for 1 sec. Hence, the interval between successive words was 5 sec. A visual depiction of the study phase is presented in **Figure 1**, part A (upper section).

**Filler task.** Four arithmetic problems (multiplication of two-digit numbers) printed on an A4 paper were prepared for a filler task. Participants were given 4 min to complete these problems.

**Memory:** A free recall test was performed, whereby the participants were asked to write down from memory as many study words as they could recall, with no time limitation. An empty sheet of paper and a pencil were provided by the experimenter. The whole experimental session lasted no more than 20 min.

**Results and discussion**

The left hand part of **Figure 2** gives the results of free recall for Experiment 1A, the baseline experiment. Plotted are the proportions of words recalled for the two learning conditions, reading silently and reading aloud. Visual inspection reveals the superiority of vocal production \( (M = .30, SE = .02) \) over silent reading \( (M = .23, SE = .02) \), a classic PE. A paired-sample \( t \)-test (one-tailed) comparing vocal production and silent reading supports this conclusion, \( t(38) = 2.99, p = .003 \). Experiment 1A confirmed the well-known PE, demonstrating the advantage of vocal production relative to silent reading.

**Experiment 1B: silent reading vs. 3 sec delayed production**

Experiment 1B tested the novel condition of delayed production, which combines the advantage of vocal production and the testing effect.

**Method**

**Participants**

The same group of 39 students who participated in Experiment 1A performed the present experiment. The order of the experiments was random across participants.

**Apparatus and stimuli**

These were the same as in Experiment 1A (note that for each participant, the study words selected for the present experiment were different from the study words in Experiment 1A).

**Design and procedure**

These were similar to Experiment 1A, except that vocal production was not immediate (as in Experiment 1A), but delayed for 3 sec following the disappearance of the appropriate study word. Again, such delayed production actually serves as an initial instant recall test of the aloud words.

**Study phase.** In each experimental trial, each study word appeared for 1 sec and disappeared. A blank screen for 3 sec followed, and then the appropriate icon solely appeared for 1 sec (without the study word), signalling the study condition. In the delayed production condition, the microphone icon appeared, indicating (delayed) vocal production (note that since the study word did not reappear, and only the icon was presented, this condition served as an initial instant test. Performance on the initial test was 100% for all participants, across all experiments, and thus will not be further discussed). In the silent condition the eye icon appeared, hence no-production was indicated. A blank screen followed for 1 sec. Therefore the interval between successive words was 5 sec. **Figure 1**, part B, shows the study phase (learning conditions) of Experiment 1B.

**Filler task and free recall memory** were the same as in Experiment 1A.

**Results**

The right panel of **Figure 2** shows the results of free recall for Experiment 1B. The proportions of words recalled for the two learning conditions (silent reading and delayed production) are plotted. **Figure 2** clearly depicts the superiority of delayed vocal production \( (M = .35, SE = .02) \) over silent reading \( (M = .21, SE = .03) \), a sizable PE. A paired-sample \( t \)-test (one-tailed) comparing delayed production and silent reading supports this conclusion, \( t(38) = 6.54, p < .001 \).

A repeated-measures ANOVA with Experiment (1A – immediate vocal production, and 1B – delayed production) and learning condition (aloud or silent reading) as within-subject variables, revealed a significant main effect for learning condition, \( F(1, 38) = 39.39, p < .001, \eta^2 = .51 \). This main effect results from the expected benefit of aloud over silent words, in both experiments.

In addition, an interaction between these two variables was found, \( F(1, 38) = 4.32, p = .044, \eta^2 = .10 \), indicating the different PE size between the experiments – namely, larger effect in Experiment 1B relative to 1A. No main effect was found for the experiment, \( F(1, 38) = .48, p > .25 \), indicating that overall memory performance (recall of aloud and silent words) was similar in both experiments.

Following the significance of the interaction, we conducted two paired-sample \( t \)-tests (one-tailed), comparing the recall rates of aloud and silent words in both
experiments, separately. Comparing the aloud conditions (immediate – Experiment 1A, and delayed – Experiment 1B) revealed a significant difference, in favour of the delayed vocal production, $t(38) = 1.98, p = .028$. No such difference was found comparing the pair of silent conditions of both experiments, $t(38) = 0.98, p > .25$. 

Figure 1. Visual depiction of the study phase (e.g., the different learning conditions) of each Experiment.
Discussion

The goal of Experiment 1 was to determine whether delayed vocal production would yield larger PE than the typical immediate production. We assumed that this novel condition would be superior to silent reading as well as to immediate production, since it holds high relative distinctiveness in addition to the advantage of initial testing. The results of Experiments 1A and 1B support this prediction, and demonstrate the advantage of delayed production on memory. A gradient of recall rates across learning conditions was observed. Silent reading resulted in low recall rates, aloud reading resulted in moderate performance, and delayed vocal production yielded the highest recall rates. This indicates that a minute departure from the standard (immediate) vocal production, namely postponing the vocalisation by 3 sec, produced a larger PE (although both conditions involve a similar number of encoding processes, i.e., they are quantitatively comparable).

The different presentation durations of the study words between Experiments 1A and 1B (4 sec vs. 1 sec, respectively) cannot explain the advantage of the delayed condition relative to immediate production. Previous findings showed that recall increases with longer exposure (presentation) durations ("on times", Ballardini, Yamashita, & Wallace, 2008; McDermott & Watson, 2001). If increased presentation time improves memory performance, one should expect better memory in Experiment 1A than 1B, since it involved longer exposure time to the study words. However, the current results indicate no such difference.

Alternatively, can the memory benefit of delayed production be the result of the increased "off time" (the blank interval following the word, which allows refreshing activities)? In Experiment 1A, no such blank interval appeared, but in Experiment 1B, each trial included a 3 sec blank screen. Ruling out this seemingly straightforward explanation is possible by comparing the recall rates of the non-produced (silent) items. If longer "off times" result in better memory, then silent words' recall should be higher in Experiment 1B than in Experiment 1A. However, memory performance was similar for the silently read words across experiments (although "off times" durations were different). Thus, differential "off times" cannot explain the different memory performance for the silent words, nor the difference between aloud words' recall. Indeed, Shaffer and Shiffrin (1972) found no improvement in memory performance from increasing the length of the blank interval following an item, where "off time" varied from 1 to 4 sec.

Taken together, it seems that the advantage of delayed relative to immediate production may be attributed to the effort of retrieval, a qualitatively based difference. The initial recovery of the study words enhances their memory relative to saying them aloud. In order to evaluate further the memory benefit of delayed vocal production and its source, Experiment 2 directly compared the typical (immediate) vocal production, and the novel delayed condition.
Experiment 2: immediate vs. 3 sec delayed vocal production

Experiment 1 demonstrated the advantage of delayed vocal production relative to silent reading and to reading aloud. Hitherto, reading aloud was considered the most beneficial production, and it was consistently found to enhance memory, relative to non-auditory (e.g., writing, mouthing), as well as auditory (e.g., whispering) productions (Forrin et al., 2012). Our results show that adding an initial instant test to this condition further enhances memory.

In Experiment 2, we compared reading aloud and the novel condition of delayed production in a single mixed-study list. It is important to note that all of the trials in this Experiment were of a similar duration (5 sec). Accordingly, in the immediate production condition, the study words were visible for longer periods relative to the delayed production condition (3 sec vs. 1 sec). This enabled us to directly compare the benefit of testing with the advantage of longer presentation duration (McDermott & Watson, 2001).

Method

Participants

A sample of 30 undergraduate students (who had not participated in the previous Experiments) from the same source as in Experiment 1 participated for course credit, with similar inclusion criteria.

Apparatus and stimuli

These were the same as in Experiment 1.

Design and procedure

These were similar to Experiment 1, except that the study conditions were immediate vocal production (reading aloud) and 3 sec delayed vocal production (similar to those used in Experiments 1A and 1B, respectively).

Study phase. All the 40 study words were vocally produced (20 words read aloud, and 20 words vocally retrieved). The participants were instructed to vocalise each word as soon as a microphone icon appeared. In the experimental trials of immediate production, each of the 20 study words appeared for 3 sec, alongside a microphone icon. The participant immediately read the word aloud (a typical “aloud” condition). Two seconds blank screen followed. In the experimental trials of delayed production, each of the 20 study words appeared for 1 sec, followed by 3 sec blank screen. Next, the microphone icon appeared for 1 sec (without the study word), indicating vocal production of the word (i.e., initial test, since the participant had to retrieve the word from memory). In both conditions, each experimental trial lasted 5 sec. Figure 1, part C, visualises the study phase of Experiment 2. Filler task and free recall memory were the same as in Experiment 1.

Results and discussion

The left panel of Figure 3 gives the results of free recall for Experiment 2. Plotted are the proportions of words recalled for the two learning conditions, immediate and 3 sec delayed vocal production. Observing Figure 3 shows the superiority of delayed vocal production ($M = .30, SE = .02$) over immediate production ($M = .25, SE = .02$). A paired-sample $t$-test (one-tailed) comparing the two learning conditions supports this conclusion, $t_{(29)} = 1.90, p = .034$. 
The superiority of delayed vocal production relative to immediate production is interesting in light of the fact that both conditions involve the same three encoding processes: visual, motor, and auditory. According to the distinctiveness account, their relative distinctiveness is similar. Thus, the quantitative interpretation of encoding distinctiveness cannot explain the superiority of delayed production relative to immediate production.

Note that the total duration of all experimental trials was equal (5 sec). However, in the delayed condition, the study items were visible only for 1 sec, while in the immediate condition items were shown for 3 sec. Again, if the longer presentation durations (“on time”) result in better recall (Ballardini et al., 2008; McDermott & Watson, 2001), then the immediate condition should have been superior to the delayed condition. The results were opposite. The difference in memory performance is thus attributed to the delay of vocal production.

The current results are in accordance with previous findings reported by Kuo and Hirshman (1996; Experiment 2). The authors compared free recall performance in two learning conditions: (a) study trial (initial presentation), in which the item was read aloud, followed by another study trial, in which the item was read aloud again, and (b) study trial followed by a test trial (in which the item was recalled aloud). The results showed superior memory performance in the study-test trials relative to the study-study trials (Note that Kuo and Hirshman used two successive vocal productions for each item. Yet in the current experiment, only a single vocal production was performed, either immediate or delayed).

It appears that adding a difficulty to the study phase by integrating the PE and the testing effect, increased memory performance. We will discuss this explanation in detail in the General Discussion section. Let us now describe Experiment 3, which was conducted in order to re-confirm our findings.

Experiment 3: immediate vs. 1 sec delayed vocal production

Experiment 2 documented the advantage of 3 sec delayed vocal production relative to immediate production (reading aloud). In the present Experiment, we replicated this comparison, with even a shorter delay, of 1 sec.

Method

Participants

A sample of 20 undergraduate students from the same source as in Experiments 1 and 2 participated for course bonus (these participants did not participate in the previous experiments). Inclusion criteria were the same as in the previous experiments.

Apparatus and stimuli

These were the same as in Experiment 2.

Design and procedure

These were similar to Experiment 2, except that the study conditions were immediate vocal production (reading aloud) and 1 sec delayed vocal production (rather than 3 sec in Experiment 2). Therefore, during the study, all the 40 words were vocally produced once the microphone icon appeared. In the experimental trials of the immediate production, each of the 20 study words appeared for 3 sec, accompanied by the icon. Hence, the participant immediately read the word aloud. A 1 sec blank screen followed. In the experimental trials of the delayed production, each of the 20 study words appeared for 1 sec, followed by 1 sec blank screen. Next, the icon appeared for 2 sec, indicating vocal production (retrieval) of the word. The interval between successive words was 4 sec (see Figure 1, part C). The filler and free recall memory tasks were the same as in Experiments 1 and 2.

Results and discussion

The proportions of words recalled for the two learning conditions (immediate production and 1 sec delayed vocal production) of Experiment 3 are shown in the middle panel of Figure 3. This figure demonstrates the superiority of delayed vocal production ($M = .29$, $SE = .04$) over immediate production ($M = .22$, $SE = .03$). A paired-sample t-test (one-tailed) that compared these conditions supports this pattern, $t(19) = 3.27$, $p = .002$.

Finally, we conducted a repeated-measures ANOVA with a type of vocal production (immediate, delayed) as a within-subject variable, and study–initial test interval (3 sec – Experiment 2, 1 sec – Experiment 3) as a between-subjects variable. This analysis revealed a significant main effect for type of vocal production, confirming that delaying the vocal production results in superior memory relative to immediate production, $F(1, 48) = 10.98$, $p = .002$, $\eta^2 = .19$. No main effect for Experiment (i.e., the study–initial test interval) was found, $F(1, 48) = .29$, $p > .25$, nor an interaction of these two variables, $F(1, 48) = .37$, $p > .25$. This pattern suggests that the impact of 1 and of 3 sec delay on recall performance was similar.

To recap, it appears that the effect of instant testing is robust, as the delayed conditions (regardless of the interval between the study and the initial test) resulted in superior recall relative to the immediate production conditions. Even a very short delay of 1 sec is beneficial to memory, comparable to the benefit reaped from a 3 sec delay. The similar recall pattern obtained in these delayed conditions suggests that the “off time” duration (the blank interval following the study words), did not affect memory performance. The decisive factor seems to be the act of retrieving (initial test) itself.
Although this conclusion seems quite convincing, one should note (as was detailed above) that the delayed production condition involves another aspect that may underlie the memory benefit – response inhibition. In each delayed production trial, a word is briefly presented, and vocal production is indicated (by the microphone icon) only after its disappearance. Obviously, delayed production involves testing since the participants are required to retrieve the word. However, they also have to withhold their vocal response, and retrieve the word only when the appropriate icon appears. In other words, since the vocal response takes place following the “off time”, response inhibition inevitably occurs. Both mechanisms, testing and response inhibition, were found to improve memory (Karpicke & Roediger, 2008; Koriat & Goldsmith, 1996; Loft & Remington, 2013; Masson & McDaniel, 1981; McDaniel & Masson, 1985; McDaniel et al., 2007).

Initial evidence that response inhibition does not underlie the memory benefit of the delayed production condition can be unveiled by examining the results of Experiments 1A and 1B. In Experiment 1A, no response inhibition was involved. Rather, the responses (vocal production, no production) were immediate. In contrast, in Experiment 1B, both learning conditions (delayed production, no production) involved (3 sec) response inhibition. The comparable memory performance found in the two silent (no production) conditions indicates that the act of withholding the study word (Experiment 1B) itself did not improve memory. Note, this rationale also rules out the act of refreshing as the source of the memory benefit observed in the delayed production condition.

Experiment 4 was conducted in order to directly test the role of the “off times” (response inhibition or refreshing) vs. retrieval (testing) in a single experimental design, tapping the source of the memory improvement of delayed production.

**Experiment 4: “Off times” vs. retrieving as the source of memory benefit of delayed production**

To clarify the possible sources that underlie the memory benefit of the delayed production condition, we directly compared two experimental conditions that involve “off times”: the delayed production condition (which also involves testing) vs. delayed reading. In both learning conditions, the study words were visually presented for a short duration and disappeared (note, the “off time” durations were similar for these two learning conditions, thus rehearsal or refreshing periods were equated). Following the “off time”, either vocal production (retrieval, indicated by a microphone icon) or reading aloud (indicated by the microphone icon along with the study word) were performed.

In light of our previous results (specifically, the comparison of Experiments 1A and 1B), we assumed that the memory advantage of the delayed production condition is due to the act of testing (retrieval), rather than response inhibition or refreshing. Thus, we expected that recall rates would be higher for the delayed production condition (which comprises testing) relative to the delayed reading condition.

**Method**

**Participants**

A fresh sample of 27 undergraduate students from Ariel University, who did not participate in the previous experiments (16 females, mean age: 23 years) received course credit for participating in this experiment. Inclusion criteria were the same as in the previous experiments.

**Apparatus and stimuli**

The pool of items was similar to that used in Experiments 1–3. From this pool, 40 words were selected for study (a random sample for each participant). The 40 study words were randomly divided into two learning conditions of 20 words each: delayed production and delayed reading. At study, each word was visually presented at the centre of the monitor in black against a white background (under the control of DirectRT programme) and disappeared. Following a brief “off time”, on each trial, a small icon of a microphone appeared with or without the study word, indicating the vocal response, as will be detailed next.

**Design and procedure**

The participants were tested individually in the lab. Upon arrival, they were given a brief explanation regarding the experimental procedure.

**Study phase.** The 40 study words were randomly divided into two equal subsets defined by the learning condition – 20 to be vocally produced following a delay, and 20 to be read aloud following a delay – randomly intermixed and presented individually.

The presentation of the words was similar across conditions: Each study word appeared for 1 sec, followed by an “off time” of 3 sec. The presentation of the icons differed across the conditions: (1) Delayed production – following the blank screen, the microphone icon appeared for 1 sec, indicating vocal production (retrieval). (2) Delayed reading condition – following the blank screen, the microphone icon appeared along with the study word, and both were presented for additional 1 sec, indicating reading aloud. Figure 1, part D (lower section), depicts the study phase of Experiment 4. The filler and free recall memory tasks were the same as in Experiments 1–3. The whole experimental session lasted no more than 20 min.

**Results and discussion**

The right hand part of Figure 3 gives the free recall results of Experiment 4, with the proportions of words recalled for the two learning conditions: delayed production and
delayed reading. Examining Figure 3 reveals the superiority of delayed vocal production ($M = .26$, $SE = .03$) relative to the delayed reading condition ($M = .21$, $SE = .02$). A paired-sample t-test (one-tailed) that compared these conditions revealed a significant difference, $t(25) = 2.29$, $p = .016$.

The previous experiments unveiled the superiority of delayed vocal production relative to silent reading and to reading aloud (immediate production). This advantage was attributed to the difficulty of testing. However, delayed production also entails "off time" (i.e., withholding the vocal responses or refreshing the perceptually absent material), which may serve as a confound to the desirable difficulty explanation.

The goal of Experiment 4 was to directly examine the effect of testing (word retrieval) in two learning conditions that involved "off times": delayed production (in which the participants were required to retrieve the word following a delay), and delayed reading (in which the participants reread the word following a delay). Note, in both conditions the presentation duration and the "off times" were identical, as well as the overall duration of each experimental trial.

The current results show an advantage of delayed production over delayed reading, affirming the important role of testing in the memory benefit of delayed production. Since "off times" (response inhibition, and refreshing, or thinking of the preceding item) were similar in both conditions, these factors cannot explain the superiority of delayed production. This conclusion correlates with the previous results of Experiment 1A and 1B (i.e., the similar recall rates of the no-production conditions, which entailed "off time").

**General discussion**

In recent years, vocal production was found to be a reliable method for improving memory (both recall and recognition), relative to other types of unique productions (e.g., whispering, mouthing, writing; MacLeod et al., 2010). Forrin et al. (2012) noted that: "… reading out loud facilitates memory more than these other productions because the act of speaking is more likely to be incorporated into the processing record, and hence is more distinct … more pronounced" (p. 1054). The underlying mechanism of the PE is encoding distinctiveness. Words learned by saying aloud (vocalising) are more distinct and unique relative to words learned by silent reading, and thus are better remembered.

Another powerful method to enhance memory is testing (Roediger & Karpicke, 2006b). Taking an initial test improves the performance in a consecutive (final) test. Each effect, production as well as testing, can serve as a mnemonic. From the applied perspective, the present study aimed to test the combined effect of these techniques. Our goal was to improve the well-documented PE, by adding low-scale testing – an initial instant recall test that followed each study word. In this novel delayed production condition, participants were presented with a study word. Only after the word disappeared, vocal production (saying the word aloud) was performed. Since at production the word was no longer seen, saying it aloud served as an initial recall test (in other words, the word had to be retrieved rather than read aloud). Since both immediate and delayed production are quantitatively comparable (both include visual, motor, and auditory encoding processes), comparing them may shed light on the qualitative aspect of the encoding distinctiveness theory as a possible source of the memory benefit.

The results reveal that delaying the vocal production yields a larger PE (the difference between aloud and silent words) relative to the typical (immediate) vocal production (Experiment 1). A direct comparison showed that delayed production was superior to immediate production, with 3 sec (Experiment 2) and even 1 sec (Experiment 3) study–initial test intervals. In addition, delayed vocal production was found to be more effective than delayed reading aloud. Both have an "off time" (response inhibition and equal refreshing activities), but only the former entails testing (Experiment 4).

Attempting to explain these results, we first focused on the encoding distinctiveness account (MacLeod et al., 2010), the dominant theoretical explanation of the PE. As described above, this account emphasises uniqueness at the item level, relative to other items in the study list, enhancing memory for the unique (and distinct) items. Ozubko and MacLeod (2010) suggested that vocalising at study makes the aloud words distinctive. Recollection of this information (distinctiveness heuristic) is used by the participants at the test, enhancing memory.

A key feature of the encoding distinctiveness account is relativity (relative distinctiveness). As noted by Forrin et al. (2012), “the production that includes a greater number of distinct, nonoverlapping processes should provide a larger boost to explicit memory” (p. 1047). Indeed, in most PE studies, as well as in the current Experiment 1, saying aloud is quantitatively distinct (since it involves three encoding processes: visual, motor, and auditory) relative to reading silently (which involves only a single visual encoding process).

However, in Experiments 2 and 3, such relative distinctiveness was eliminated, since all study words were vocally produced, either immediately or after a brief delay. Since both study conditions involved a similar number of encoding processes (i.e., three), this cannot explain the benefit of the delayed production condition. This pattern of results suggest that delayed vocal production is qualitatively different from immediate vocal production. Attempting to explain the source of the memory benefit, two alternative accounts were ruled out:

(a) **Presentation duration ("on time")**

A possible explanation of the advantage of delayed vocal production relative to immediate production (reading
aloud) may be the differential presentation durations of the study items. Previous studies found that longer presentation durations (“on times”) or longer encoding times during initial processing yield stronger memory representations and better memory performance (Ballardini et al., 2008; McDermott & Watson, 2001; Schulman & Lovelace, 1970). However, in the present study, longer presentation durations characterised the immediate production conditions, and the delayed conditions involved shorter “on times”. Thus, this explanation was ruled out.

(b) Blank interval duration (“off time”)

Next, we considered the different durations of the “off times” between the learning conditions as a possible source for the delayed production benefit. Presumably, participants may use the blank interval following a study word for its rehearsal or for a refreshing activity. Indeed, this interval was longer in the delayed conditions relative to the immediate conditions. If so, then the silent words should benefit from the longer “off times” as well. However, the comparable silent words’ recall ratios in Experiments 1A and 1B negates this explanation (see also Shaffer & Shiffrin, 1972).

Additional support for the insignificant role of the blank intervals durations can be found by comparing memory performance for the delayed production conditions of Experiments 2 and 3. The “off times” differences (3 vs. 1 sec, respectively) did not affect memory as both conditions resulted in similar recall rates.

Note that Roediger and Karpicke (2006b) suggested that when tests are spaced, the placement of the first test might be an important factor in memory performance. In order to enhance later retention, one should increase the delay before an initial test. A longer interval before the first retrieval attempt demands more retrieval effort and leads to better retention (for a similar rationale, see Bartlett, 1977). The current results showed no difference in memory benefit between the pair of selected study–initial test intervals (1 and 3 sec). Future studies should examine other study–initial test intervals, in order to determine the optimal interval, resulting in maximum gain (final test performance) with minimum “pain” (prolonged learning phase and increased learning effort).

(B.1.) The reflection operation of refreshing

According to the MEM model (Johnson, 1992), refreshing (thinking of-, or keeping active) information is a basic reflective process (Raye et al., 2007). This simple maintenance activity can be recruited to prolong activation of (no longer present) representations, and is associated with better performance on later memory tests (Raye et al., 2002). When asking participants to delay their vocal response, they must maintain its activation, in order to retrieve it later on. Following this reasoning, the delayed production condition must include a refreshing mental procedure.

However, the aforementioned explanations undermine the role of refreshing in the memory benefit of delayed production (namely, comparing Experiments 1A and 1B, and Experiments 2 and 3). Furthermore, Experiment 4 is revealing. In both learning conditions, refreshing occurred, since the participants had to maintain the previously presented word, as they did not know whether it would reappear (delayed reading) or not (delayed production). The results showed the superiority of delayed production compared to delayed reading, suggesting that the critical factor was the act of retrieving.

(B.2.) Response inhibition

Finally, the role of response inhibition explaining the memory benefit of delayed production was rejected. As detailed above, delayed production involved response withholding (Koriat & Goldsmith, 1996), since the participants were required to retrieve the word only when a cue was given. The first evidence that allowed this explanation to be excluded was the similar recall rates obtained in the no-production conditions of Experiments 1A (no response inhibition) and 1B (response inhibition). Further evidence was provided through the results of Experiment 4, in which we directly confronted two learning conditions that hold response inhibition, delayed production and delayed reading. The results showed enhanced memory for delayed production relative to delayed reading. Since response inhibition equally characterised both conditions, it cannot explain the memory difference.

Ruling out the abovementioned accounts, the current results lead to a straightforward explanation – the difficulty that was induced by the initial instant tests (delaying the vocal production) relative to reading aloud (immediately or delayed), resulted in enhanced performance. Indeed, more difficult initial processing leads to stronger or more complete representations in memory (Pazzaglia, Staub, & Rotello, 2014). These difficult learning conditions are desirable since they improve long-term retention (Bjork, 1994; Bjork, Little, & Storm, 2014).

In a related study, Icht et al. (2016) found that vocalising improved memory relative to silent reading in a group of dysarthric adults. This clinical population is characterised by reduced speech abilities (associated with neurological disorders), relatively unintelligible and “slurred” speech. For these individuals, saying aloud is “more challenging and effortful than it is for healthy individuals; hence, it provides greater memory benefit” (Icht et al., 2016, p. 9). Indeed, the PE was found to be larger for this group relative to healthy controls. Similar results were found among cochlear implant users (Taitelbaum-Swead et al., 2017), with larger PEs as the learning task was more difficult (auditory vs. visual). Possibly, the desirable difficulties account may explain the classic PE in non-clinical populations as
well, since vocalising is slightly more demanding than silent reading.

Interestingly, the MEM rationale explains different levels of cognitive effort in terms of the number of different component processes (e.g., refreshing, rehearsing, noting, etc.) that are involved in a given task, and how they are coordinated during encoding and remembering (Mitchell & Johnson, 2009). Presumably, delay of the vocal production holds an additional encoding procedure: reflective (the act of retrieval), in addition to visual (reading a word), motor (articulating the word), and auditory (hearing the self-vocal production). Thus, it results in enhanced memory relative to no-production (a single visual process) or to immediate production (three encoding processes: visual, motor, and auditory).

In summary, delaying the vocal production is more demanding relative to the typical immediate vocal production, since it convolutes a pair of memory tools—vocalising and testing. Indeed, it has a significant impact on memory performance, and it results in larger recall rates relative to saying aloud, although it involves only a minute change from the familiar immediate vocal production. The relatively insignificant procedural alteration results in a significant change in performance. Thus, this new learning condition can serve as a powerful mnemonic device, more beneficial than vocalisation.

Disclosure statement

No potential conflict of interest was reported by the authors.

References


Mama, Y., & Ich, M. (2016a). Auditioning the distinctiveness account: Expanding the production effect to the auditory modality reveals...


