

The production effect in paired-associate learning: Benefits for item and associative information

Adam L. Putnam · Jason D. Ozubko ·
Colin M. MacLeod · Henry L. Roediger III

Published online: 24 October 2013
© Psychonomic Society, Inc. 2013

Abstract In five experiments, we extended the production effect—better memory for items said aloud than for items read silently—to paired-associate learning, the goal being to explore whether production enhances associative information in addition to enhancing item information. In Experiments 1 and 2, we used a semantic-relatedness task in addition to the production manipulation and found no evidence of a production effect, whether the measure was cued recall or item recognition. Experiment 3 showed that the semantic-relatedness task had overshadowed the production effect; the effect was present when the semantic-relatedness task was removed, again whether cued recall or item recognition was the measure. Experiments 4 and 5 provided further evidence that production can enhance recall for word pairs and, using an associate recognition test with intact versus rearranged pairs, indicated that production may also enhance associative information. That production boosts memory for both types of information is considered in terms of distinctive encoding.

Keywords Memory · Recall · Recognition · Production

The *production effect* is the finding that saying a word aloud enhances memory as compared to reading a word silently

(MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). Although evidence of this effect appeared decades ago (Conway & Gathercole, 1987; Gathercole & Conway, 1988; Hopkins & Edwards, 1972), the phenomenon remained relatively unknown until 2010, when MacLeod and his colleagues (MacLeod et al., 2010) named the effect and brought it to wider attention. In a series of studies (Forrin, Jonker, & MacLeod, *in press*; Forrin, MacLeod, & Ozubko, 2012; Hourihan & MacLeod, 2008; Lin & MacLeod, 2012; MacLeod, 2011; Ozubko, Gopie, & MacLeod, 2011; Ozubko, Hourihan, & MacLeod, 2012; Ozubko & MacLeod, 2010; Ozubko, Major, & MacLeod, *in press*), they have suggested that various forms of production (such as speaking and writing) enhance future recall and recognition by creating a more distinctive encoding.

In the standard production effect experiment, subjects are asked to study a word list for an upcoming memory test by reading half of the words silently and the other half aloud. On a subsequent memory test, items that were read aloud (i.e., produced) typically show an advantage of 10 %–20 % over items that were read silently, whether the test is recognition (e.g., MacLeod et al., 2010) or recall (e.g., Lin & MacLeod, 2012). Variations have supported the robust nature of the effect and indicated that mouthing, writing, whispering, typing, and spelling also enhance memory, although typically not to the same degree as reading aloud (Castel, Rhodes, & Friedman, 2013; Conway & Gathercole, 1987; Forrin et al., 2012; Gathercole & Conway, 1988; MacLeod et al., 2010).

Certain boundary conditions do exist for the production effect, however. For example, the production task must require subjects to produce a unique response for each item; that is, saying “yes” in lieu of actually producing does not enhance future performance (MacLeod et al., 2010, Exp.4) although, interestingly, subjects think that it will (Castel et al., 2013). Critically, the production effect is largely limited to within-subjects designs in which the spoken items are intermixed

A. L. Putnam · H. L. Roediger III
Washington University, St. Louis, MO, USA

J. D. Ozubko · C. M. MacLeod
University of Waterloo, Waterloo, ON, Canada

A. L. Putnam (✉)
Department of Psychology, Washington University in St. Louis,
One Brookings Drive, Campus Box 1125, St. Louis,
MO 63130-4899, USA
e-mail: adam.putnam@wustl.edu

with the silent items (e.g., Hopkins & Edwards, 1972; Jones & Pyc, *in press*, MacLeod et al., 2010), although meta-analysis has suggested that a small effect may occur in between-subjects designs, as well (Fawcett, 2013).

This pattern of results has led several researchers (Conway & Gathercole, 1987; Ozubko & MacLeod, 2010) to suggest that producing a word does not enhance the overall strength of the word, but rather adds a distinctive “I said it aloud” record, relative to silent reading (Hunt, 2006, 2013; see also Ozubko et al., *in press*). MacLeod and colleagues (e.g., MacLeod et al., 2010; Ozubko & MacLeod, 2010) have advanced the argument that enhanced distinctiveness is responsible for the production effect. They suggest that producing a word results in a qualitatively different memory record than does silent reading and that, at retrieval, subjects are able to use this distinctive information as part of their decision process in determining whether an item is old or new: “I remember saying that item aloud, therefore I must have studied it.” Thus, the advantage of production is relative to reading silently. This account is consistent with the many studies (e.g., Dodson & Schacter, 2001; Hopkins & Edwards, 1972; MacLeod et al., 2010) showing that production manipulated as a between-subjects variable (with aloud and silent items in different pure lists) typically fails to enhance recall reliably. Furthermore, several studies have directly tested the distinctiveness hypothesis, with generally favorable results (Ozubko & MacLeod, 2010; Ozubko et al., *in press*; but see Bodner & Taikh, 2012, for a different interpretation).

In sum, the production effect appears to be a robust phenomenon resulting in a memory advantage for words spoken aloud as compared to words read silently. However, most published articles on the production effect have used old–new or two-alternative forced choice recognition tests (e.g., Hopkins & Edwards, 1972; MacLeod et al., 2010), or occasionally free recall tests (Conway & Gathercole, 1987; Lin & MacLeod, 2012; MacLeod, 2011), or even a fill-in-the-blank test (Ozubko et al., 2012). Thus far, no study has examined whether a production effect can occur in a standard cued recall setting, such as in the context of a paired-associate learning paradigm.

Ozubko et al. (2012, Exp.2) did show a production effect with word pairs, but their subjects simply studied intact pairs of words and then took an old–new recognition test with other word pairs as distractors; at no point was one member of a pair used to cue the other, as in the standard paired-associate learning paradigm. Given the claim that the production effect is a robust, reliable phenomenon, it is important to investigate whether it extends to paired-associate learning. But this issue is not just about empirical generalizability: Importantly, paired-associate learning would also allow us to examine whether production can enhance the learning not just of item information, but also of associative information (see Hockley, 1991; Hockley & Consoli, 1999). Does production enhance memory for connections between items, as well as memory for the component items themselves? And if production does

enhance paired-associate learning, what might the boundary conditions on that enhancement be?

The present series of experiments had their origin when two of the authors (A. L. P and H. L. R.) began to explore how different response modes during retrieval affected recall, both during a current test (can subjects recall more information by typing or speaking?) and on future tests (will typing or speaking on a current test lead to better recall in the future?) We began this research with an experiment (the present Exp. 1A) with the goal of demonstrating the production effect in cued recall before varying response mode at test (overt production vs. silent reading). We had wanted to determine whether production during study and during testing would produce additive effects on a final recall test. However, to our surprise, we failed to show the production effect. Consultation with the other two authors (J. D. O. and C. M. M.) led to the present series of experiments, to determine whether production could enhance recall in paired-associate learning and, if so, what conditions might be necessary for such an effect to emerge. Previous theorizing led to two different possibilities.

The first was that the production effect might not extend to cued recall, as it may only enhance item-specific information. Models of paired-associate learning suggest that a critical part of encoding pairs is strengthening the association between the cue and the target (McGuire, 1961). However, in a standard production-effect experiment, items are processed individually and require a unique response, both of which suggest that production may only enhance item-specific information, leaving the association between the cue and the target untouched (Humphreys, 1976). Furthermore, the recollective-distinctiveness account put forward to explain the production enhancement in recognition (MacLeod et al. 2010; Ozubko & MacLeod, 2010) suggests that production makes previously spoken items distinctive in the context of silently read items and lures, neither of which having been produced aloud. Indeed, distinctiveness theories typically suggest that item-specific processing enhances discrimination, which is consistent with this view of production (Hockley & Consoli, 1999; Hunt, 2006; Hunt & McDaniel, 1993; Nairne, 2006). This view suggests that production may only enhance item-specific information and may not assist in paired-associate learning, in which the association between a cue and a target must be strengthened.

A case can be made, however, for a second possibility—that production should enhance cued recall—because a case can be made that production encourages relational processing. First, research on the generation effect has indicated that although generation may disrupt listwide relational processing (see McDaniel & Bugg, 2008, for a detailed analysis of this issue), generation does enhance the associative link between a cue and a target (Hirshman & Bjork, 1988). The similarities between the production and generation effects (cf. MacLeod et al., 2010)

suggest that production should enhance the association between a cue and a target. Second, Yonelinas (2002) argued that subjects must use recollection to detect new associative pairings on a recognition test. Previous research has indicated that production enhances both recollection and familiarity (Ozubko et al., 2011), so production may also support associative learning. Finally, the production effect has been demonstrated in free recall (Conway & Gathercole, 1987; Lin & MacLeod, 2012; MacLeod, 2011), and success in free recall may be due to a combination of both item-specific *and* organizational processing, the latter of which entails encoding information about the relations among specific list items (McDaniel & Bugg, 2008). Granted, listwide relational processing is not identical to strengthening the association between a cue and a target (Mulligan & Lozito, 2004); however, the finding that production enhances free recall is an indication that production may support some relational processing, and thus may help strengthen the associative link between a cue and a target.

The goal of the present article is to examine whether the production effect extends to paired-associate learning and, if so, what the boundary conditions are. Beyond being a simple extension of the paradigm to cued recall, finding a production effect in paired-associate learning would also provide further evidence that production enhances recollection, and would further show that production can enhance the processing not just of item information, but also of relational information.

Experiment 1A was undertaken with the assumption that the production effect would extend to paired-associate learning; the failure to find an effect revealed a boundary condition of the production effect, and served as a starting point of our experiments. In Experiment 1A, subjects learned word pairs by speaking, typing, or reading them silently during study and then taking a cued recall test in which the left member of the pair was presented as a cue for recall of the right member. As noted, we found no production effect. In Experiments 1B, 2, and 3, we examined whether several procedures used in Experiment 1A prevented the production effect from emerging in that experiment, or whether production simply did not boost performance in paired-associate learning. We did find a production effect in paired-associate learning under somewhat different conditions than in the early experiments. Finally, in Experiments 4 and 5, we used a streamlined design to show a strong production effect in both cued recall and free recall. The addition of an associative recognition test after each recall test in these last experiments allowed us to examine directly whether production facilitated the learning of new associative information.

Experiments 1A and 1B

Experiment 1 was our first attempt to demonstrate the production effect in paired-associate learning. Subjects studied word pairs by reading each pair silently, by typing both members of

each pair, or by saying both members of each pair aloud. Both typing and speaking the pairs constituted production. After studying each pair, subjects made a rating of the semantic relatedness of the word pair. This rating was included to ensure that subjects were encoding each word pair, as we were worried that they might not pay full attention in the silent-reading condition if no overt response was required. The two main questions were whether production could occur with paired-associate learning and whether speaking or typing would lead to a stronger production effect. On the basis of Forrin et al. (2012), if we were able to observe a production effect in paired-associate cued recall, we expected that spoken production would be more beneficial than typed production. In Experiment 1A, we used weakly related word pairs but, after seeing the results, we hypothesized that the relatedness of the words combined with the semantic-rating procedure might have resulted in too strong an encoding. We therefore switched to unrelated word pairs in Experiment 1B. Other than this difference, the two experiments were identical in procedure, and thus are reported together.

Method

Subjects In experiment 1A 24 subjects from Washington University in St. Louis participated for course credit or US\$5. In addition, 20 subjects from the University of Waterloo participated in Experiment 1B for course credit or CAN\$5. The target sample sizes for all experiments reported here were selected after observing fairly large effect sizes in previous studies on the production effect (e.g., MacLeod et al., 2010).

Materials and design For Experiment 1A, we used 45 weakly related word pairs generated from the University of South Florida Word Association, Rhyme, and Word Fragment Norms (Nelson, McEvoy, & Schreiber, 1998). Each word pair had a forward cue-to-target strength and backward target-to-cue strength of between .01 and .02, and each word was three to nine letters long (e.g., “sailor–anchor”). For Experiment 1B, we used 45 unrelated word pairs taken from the MRC Psycholinguistic Database (Wilson, 1988). These words were matched for length with the words used in Experiment 1A. An example unrelated pair is “approach–record.”

Each set of 45 word pairs was divided into three groups of 15 pairs. The three groups of items were assigned to the different conditions (aloud, type, and read) and were counterbalanced across subjects. During the study phase, the word pairs were presented in red, blue, or green font, with each font color signaling a different condition. The colors and corresponding conditions were also counterbalanced across subjects.

Procedure Subjects were informed about the nature of the experiment and gave their informed consent. Before beginning the study phase, subjects completed a practice phase in

which they learned the task. Word pairs were presented in red, blue, or green font on a white background, and subjects were asked to respond by reading each pair silently, reading each pair aloud, or typing both members of each pair. The font color cued subjects how to respond: for example, red= read silently, green= type, and blue= read aloud. Throughout the experiment, subjects had an index card listing the color-to-condition pairings, to remind them of how to respond to each color. After completing the practice session, subjects began the study phase.

During the study phase, word pairs appeared on the screen for 4 s each. Subjects were asked (1) to read both the cue and the target of the pair silently and to continue reading silently until the pair left the screen (the read condition), or (2) to read both members of the pair aloud and to continue saying them aloud until the pair left the screen (the read-aloud condition), or (3) to type the pair into a text box once (the typed condition). The repeated-reading instruction departs from the standard production-effect instructions (MacLeod et al., 2010), but it was included to ensure that processing times for the aloud and silent items would be equivalent to those for the typed items. A second departure from the standard production-effect procedure was that, after processing each word pair, a semantic-relatedness prompt appeared, asking the subjects to indicate how related they thought the two words were, on a scale from 1 to 5. Subjects made their responses with the keypad, and the computer advanced after they had pressed the Enter key. A 500-ms interstimulus interval (a blank screen) appeared between trials.

After studying all of the 45-word pairs, subjects spent 5 min working on an unrelated distractor task (listing US presidents in Exp. 1A, or naming countries of the world in Exp. 1B). Finally, subjects took a cued recall test in which they were presented with a cue word (the left member of a pair) accompanied by a series of question marks, both presented in black font to avoid any confound with the initial presentation color. Subjects typed their responses—intended to be the corresponding right member of the pair—and pressed the Enter key to submit their answers, after which they moved on to the next trial. When subjects could not recall an item, they just pressed the Enter key to move on. No time limit was imposed. The experiment lasted about 15 min; after finishing the cued recall test, subjects were thanked and debriefed.

Results and discussion

For all of the experiments reported, answers were coded as correct if they were obvious misspellings of the target word—for example, “anchr” for “anchor.” Any target word recalled in response to a different cue was counted as incorrect.

Performance on the cued recall tests in Experiments 1A and 1B, which failed to show a production effect, is

presented in Table 1. Unsurprisingly, two separate one-way repeated measures analyses of variance (ANOVAs) confirmed that no significant differences emerged between the conditions for Experiment 1A, $F(2, 46)=0.06$, $p=.945$, $\eta^2=.003$, or for Experiment 1B, $F(2, 38)=1.31$, $p=.28$, $\eta^2=.064$, suggesting that neither typing nor reading pairs aloud resulted in a production effect on a cued recall test. Post-hoc tests were not conducted.

No production effect emerged in this paradigm. Typing, speaking, and reading silently led to equivalent performance on a cued recall test, regardless of whether the materials were unrelated or weakly related word pairs. However, as noted, the design of these experiments departed from the standard production-effect paradigm. Could one of these modifications have prevented the production effect from occurring? First, the standard production effect only pits reading aloud against reading silently, whereas in the present experiments we used reading aloud, typing, and reading silently. It is possible that having three levels of production rather than two made the produced items less distinctive (i.e., two-thirds of studied items were produced in one form or the other). This consideration seems unlikely to account for our failure to obtain the production effect, however, because recent research by Forrin et al. (2012) has shown a production effect with multiple forms of production in the same experiment.

Second, including the semantic-rating task after encoding may have overshadowed any influence of production. The benefit to recall from the semantic-relatedness task may have been so powerful that it masked or eliminated the effect of production. This hypothesis may seem unlikely, in that performance on the test was hardly at ceiling in either experiment; in fact, performance was quite low in Experiment 1B, with the unrelated pairs (9 % to 14 % across conditions). The hypothesis is also in contrast to other results showing that production enhances memory above and beyond generation or a semantic-processing task (Forrin et al., *in press*; MacLeod et al., 2010, Exps. 7 and 8). However, those previous experiments had used single words, whereas in the present experiments we used paired associates. Given that cued recall requires a strong link between the cue and target members of pairs, it is possible that the semantic-relatedness task directly enhanced the associative link between cue and target more

Table 1 Experiments 1A and 1B: Proportions of correct cued recall as a function of study condition

Condition	Aloud		Typed		Silent	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
1A: Weakly Related Words	.69	.04	.69	.04	.68	.03
1B: Unrelated Words	.14	.03	.09	.02	.13	.03

than any possible influence of production could. Experiment 3 directly addressed this issue.

Finally, a third possibility was that production might simply not benefit cued recall in paired-associate learning. As we mentioned earlier, one line of thinking suggests that production may enhance only item-specific information, and not the associative or relational information necessary for cued recall. If so, then producing word pairs should still enhance memory for the individual cue or target words—the item information. Thus, production should enhance recognition for the individual members of the word pair, even if it does not benefit associatively cued recall. Examining this possibility was the goal of Experiment 2.

Experiments 2A and 2B

The study phases of Experiments 2A and 2B were identical to those of Experiment 1, but the cued recall test was replaced with an old–new recognition test for the target words. If production enhances recognition for the targets when they are tested as individual items, this outcome would provide evidence that production enhances retention for the component members of word pairs if the items are tested individually. If this outcome were to occur, it would appear that production does not influence the associative link between the cue and the target; rather, it serves only to make the individual components of the pair—the items—more distinctive. The materials were the same as in Experiments 1A and 1B; Experiment 2A featured weakly related word pairs, and Experiment 2B featured unrelated word pairs.

Method

Subjects There were 18 subjects from Washington University in St. Louis in Experiment 2A, and 24 subjects from the University of Waterloo in Experiment 2B.

Materials, design, and procedure The same stimuli from Experiment 1 were used: weakly related word pairs for Experiment 2A, and unrelated word pairs for Experiment 2B. In addition, 45 new words were taken from each pool to serve as lures during the respective recognition tests.

The procedure was identical to that of Experiment 1, except that the cued recall test was replaced with an old–new recognition test for target members of the pairs (i.e., those studied on the right side of each pair). Subjects were not aware that only the target items would be tested. During the recognition test, an individual word appeared on the screen, and subjects were to decide whether they had seen that word during study. They responded by pressing the “z” key, if they thought that the word was old, or the

“m” key, if they thought that the word was new. No time limit was imposed for each response.

Results and discussion

The hit rates and false alarm rate for Experiments 2A and 2B can be found in Table 2. Once again, no production effect emerged: Reading aloud and typing led to recognition performance equivalent to that of reading silently, both for the weakly related word pairs and for the unrelated word pairs. Two repeated measures one-way ANOVAs confirmed that study condition did not influence recognition, for either the weakly related word pairs, $F(2, 34) = 0.17, p = .848, \eta^2 = .01$, or the unrelated word pairs, $F(2, 46) = 1.15, p = .33, \eta^2 = .05$.

Thus, it appears that producing a word pair does not enhance future recognition of the individual target words, discrediting the hypothesis that the absence of a production effect in cued recall in Experiment 1 was due to production only influencing item-specific information in pairs. The outcome does not support the idea that production makes individual items within a pair distinctive while leaving interitem associative information untouched. It should be noted, though, that these results contrast with those of Ozubko et al. (2012, Exp. 2), who did show a production effect for studied pairs. Using a recognition test containing studied pairs and unstudied pairs, they showed better recognition of pairs that had been studied by reading both words aloud. We suspect that the difference between the present result and that of Ozubko et al. (2012) is the presence of the semantic-rating task in our experiments but not in the prior experiments. We will document this point in Experiment 3.

We also conducted a set of experiments like the ones just reported, in which the final test was a free recall test. However, subjects who studied unrelated word pairs recalled only 1%–2% of the items, and subjects who studied weakly related word pairs recalled only 3%–7% of the word pairs. Obviously, no conclusions were possible due to floor effects. We decided not to pursue free recall, because bringing performance up to moderate levels would have necessitated making large changes to the study phase, and we wanted to keep the experiments as similar as possible.

Following Experiment 1, we discussed three hypotheses for why the production effect might not have occurred in cued recall: Typing and speaking in the same experiment might not allow for distinctive records to emerge, the semantic-relatedness task might somehow overshadow the production manipulation, and production might only enhance item-specific information and not associative information. The results of Experiment 2 discredit the third hypothesis. Given that previous research (Forrin et al., 2012) discounted the first hypothesis, Experiment 3 was designed to directly examine the role of the semantic-relatedness task in influencing

Table 2 Experiments 2A and 2B: Proportions of “yes” responses (hits for studied items; false alarms for new items) as a function of study condition

Condition	Aloud		Typed		Silent		New	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
2A: Weakly Related Words	.85	.03	.83	.04	.83	.03	.11	.01
2B: Unrelated Words	.76	.03	.71	.04	.73	.04	.14	.02

performance. We introduced a between-subjects manipulation: One group of subjects completed the semantic-rating task after encoding each word pair, whereas the other group simply moved on to the next study item without the semantic-rating task.

Experiments 3A and 3B

The goal of Experiment 3 was to determine whether the semantic-relatedness task included after encoding each word pair could have overshadowed any influence of production. To do so, we used two groups of subjects: One group rated the relatedness of the cue and target (as in Exps. 1 and 2), whereas the other group did not. We tested performance on both a cued recall test and a recognition test. (We also collected data with a free recall test, but, as mentioned above, floor effects prevented us from drawing any conclusions.) In both experiments, subjects studied weakly related word pairs, as in Experiments 1A and 2A. In fact, Experiment 3A was exactly the same as Experiment 1A (cued recall), with the addition of the between-groups manipulation. Experiment 3B was exactly the same as Experiment 2A (item recognition), with the addition of the between-groups manipulation. This design permitted us to replicate Experiments 1A and 2A.

We hypothesized that we would observe a Condition (read aloud, type, or read silently) \times Task (semantic rating or no semantic rating) interaction. If the semantic-rating task overshadowed any influence of production in the preceding experiments, we should see that typed or read-aloud items would be recalled and recognized better than items read silently without the semantic-rating task, but that items from the three conditions would be recalled or recognized equally well in the semantic-rating condition, as in the earlier experiments.

Method

Subjects All of the subjects in Experiment 3 came from Washington University in St. Louis. There were 48 subjects (24 in each condition) in Experiment 3A (with the cued recall test), and 48 subjects (24 in each condition) in Experiment 3B (with the recognition test). The subjects were randomly assigned either to the group with the semantic-rating task or to the group without that task.

Design, materials, and procedure Each experiment used a 3 (encoding condition: read aloud, type, read silently) \times 2 (semantic rating: present vs. absent) design. Encoding condition was varied within subjects, and semantic rating was varied between subjects. The 45 weakly related word pairs from Experiment 1A were used. The experimental procedure was identical to that of Experiments 1A and 2A, except that half of the subjects were not asked to complete the semantic-rating task after each pair; instead of making a judgment after studying each pair, the subjects in this group simply moved on to the next trial. In Experiment 3A, a cued recall task was used as the final test; in Experiment 3B, a recognition test was used instead.

Results and discussion

Experiment 3A Figure 1 shows the results of Experiment 3A. Subjects who performed the semantic-rating task recalled significantly more items than did those without the rating task, showing the power of this manipulation. More importantly, and as predicted, performance was equivalent across the encoding conditions for the group that performed the semantic-rating task, whereas performance differed across the study conditions for the group that did not perform the semantic-rating task: For the first time in this series of experiments, we obtained the production effect. A 3 \times 2 mixed model ANOVA revealed a main effect of encoding condition, $F(2, 92) = 3.26, p = .043, \eta^2 = .06$, a main effect of semantic rating, $F(1, 46) = 59.95, p < .001, \eta^2 = .57$, and a marginally significant interaction, $F(2, 92) = 2.65, p = .08, \eta^2 = .05$. Given the specificity of our prediction, we also conducted two repeated measures one-way ANOVAs, one for each semantic-rating group, essentially treating each as a separate experiment. As expected, the ANOVA for the semantic-rating condition did not yield any significant variation among groups, $F(2, 46) = 0.06, p = .95, \eta^2 = .002$, replicating Experiment 1A. The results for the group without semantic ratings, however, revealed significant variation among the conditions, $F(2, 46) = 5.62, p = .007, \eta^2 = .20$. The read-aloud condition yielded significantly better cued recall than did the read-silently condition, $t(23) = 3.45, p = .002, d = 0.57$. However, the type condition did not differ significantly from either the read-silently condition, $t(23) = 1.72, p = .098, d = 0.31$, or the read-aloud condition, $t(23) = 1.63, p = .117, d = 0.11$. We

conclude that speaking pairs led to a production effect in cued recall when the semantic-rating task was removed from the procedure.

Experiment 3B A similar pattern of results emerged when an item recognition test replaced the cued recall test (see the hit rates in Fig. 2). Subjects in the semantic-rating group had good recognition memory, with hit rates above .80 and a false alarm rate of .10, but little variation among the three encoding conditions. Subjects in the group without semantic ratings, however, did show a production effect, with better recognition of items read aloud than of those typed or read silently; their false alarm rate was .16. A 3×2 mixed-model ANOVA revealed significant main effects of encoding condition, $F(2, 92) = 4.49, p = .014, \eta^2 = .09$, and of semantic rating, $F(1, 46) = 18.92, p < .001, \eta^2 = .29$, together with a significant interaction, $F(2, 92) = 3.80, p = .026, \eta^2 = .08$. In brief, a production effect occurred when there was no semantic-rating task, but not when the semantic-rating task was included. Two separate one-way repeated measures ANOVAs revealed no differences in recognition performance for the semantic-rating group, $F(2, 46) = 0.37, p = .695, \eta^2 = .02$, but the conditions did differ for the group without the semantic-rating task, $F(2, 46) = 6.20, p = .004, \eta^2 = .21$. Further comparisons for the group with no rating task revealed that items read aloud were recognized better than items read silently, $t(23) = 3.58, p = .002, d = 0.62$. Typed items were recognized marginally worse than items read aloud, $t(23) = 1.96, p = .062, d = 0.29$, but were not recognized differently from silently read items, $t(23) = 1.65, p = .114, d = 0.29$. Thus, items read aloud were recognized better than typed items or silently read items when no semantic-rating task was included.

In both Experiments 3A and 3B, we observed a numerical trend in the no-semantic-rating condition for the typed items to be remembered better than the silent items, but these

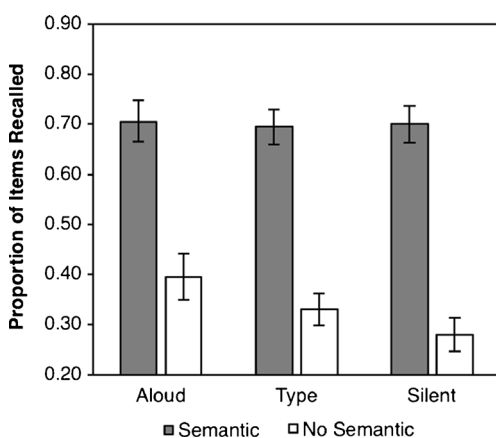


Fig. 1 Experiment 3A: Cued recall performance in terms of correct recall by the semantic-rating and no-semantic-rating groups as a function of study condition. Error bars represent standard errors

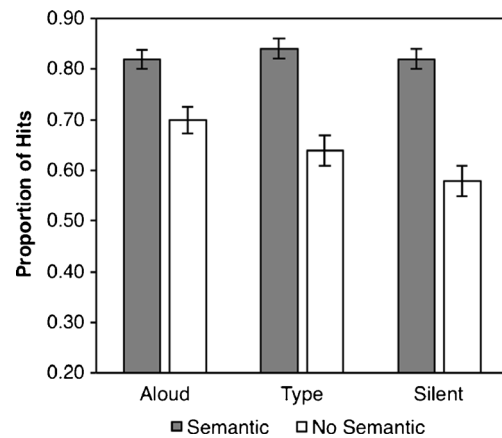


Fig. 2 Experiment 3B: Recognition performance in terms of hit rates by the semantic-rating and no-semantic-rating groups as a function of study condition. Error bars represent standard errors

comparisons did not reach significance. To gain more power to determine whether a production effect due to typing (relative to silent reading) occurred, we combined the cued recall results from Experiment 3A and the hit rates from Experiment 3B, with appropriate weighting of the means, which led to the following results: aloud $M = .55, SEM = .04$; typed $M = .49, SEM = .04$; and silent $M = .43, SEM = .04$. A 2 (experiment) \times 3 (condition) ANOVA revealed a main effect of type of final test, $F(1, 46) = 36.73, p < .001, \eta^2 = .044$, with recognition of course being superior, and a main effect of production, $F(2, 92) = 11.80, p < .001, \eta^2 = .204$, but no interaction, $F(2, 92) = 0.05, p = .950, \eta^2 = .001$. Follow-up comparisons among the three conditions revealed that the items spoken aloud produced better performance than did the typed items, $t(47) = 2.52, p = .015, d = 0.24$, and that the typed items led to better performance than did the items read silently, $t(47) = 2.37, p = .022, d = 0.24$. Thus, with these combined data, we conclude that both speaking and typing are effective forms of production, relative to silent reading, although speaking confers a greater benefit.

The results of Experiments 3A and 3B provide strong evidence that production can enhance the learning of paired associates and that the null effects of the earlier experiments were due to the inclusion of the semantic-relatedness task. This outcome is somewhat inconsistent with the findings of MacLeod et al. (2010) and Forrin et al. (in press), who showed that production enhances retention after a semantic judgment task or a generation task, although of course we used a somewhat different task. Why does a production effect occur when semantic-orienting tasks are included in a single-word-list learning paradigm, but not when a semantic-rating task is used with paired associates? We think that the answer lies in the nature of the semantic task—whether it emphasizes item-specific or relational (associative) information. In the present experiments, rating how related the two words were encouraged the encoding of relational information or the building of

associative links between the cue and the target. The fact that the groups that provided semantic ratings had better cued recall and recognition performance than did the groups that did not perform semantic ratings during study (.70 vs. .34 for cued recall, and .83 vs. .64 for recognition hit rates) fits with the idea that the semantic-rating task provided a much stronger mnemonic effect than did production.

Cued recall requires associative information, so making an explicit judgment about the semantic relatedness of two words is exactly the sort of task that should facilitate performance on that type of test (in line with transfer-appropriate processing; Morris, Bransford, & Franks, 1977). The lack of a production effect in the item recognition test is still puzzling, however. One possibility is that, in the recognition test, when subjects do recollect the item, they are more likely to recollect some aspect of the semantic rating rather than whether the item was produced. The semantic-rating task may have provided sufficient evidence for subjects to recollect that the item is old, and thus they may never have used cues related to production.

Regardless of why the semantic-rating task eliminates the production effect, it is clear from Experiment 3 that production—especially reading the pair aloud—can enhance paired-associate learning. In Experiments 4 and 5, we used a streamlined design to provide further evidence that this is so. To evaluate the relational-encoding possibility, we also introduced an associative recognition test to investigate whether production enhances item-specific information, relational information, or both.

Experiment 4

The goal of Experiment 4 was to show a strong production effect in a paired-associate learning paradigm involving just reading aloud versus reading silently, like the single-word production-effect experiments. Subjects studied paired associates by reading aloud or reading silently before taking a cued recall test. Notably, subjects did not complete a semantic-rating task during the study phase. To investigate the hypothesis that production could enhance the learning of associative information as well as item-specific information, after the cued recall test, subjects also took an associative recognition test in which the studied word pairs were presented as intact studied pairs or as rearranged pairs consisting of the studied cues and targets randomized into new word pairs. The associative recognition test was included to test whether production could enhance recognition for the pair as a unit, or whether it only enhanced recall for the individual target when given its cue. On the basis of the hypothesis that production makes events more distinctive, we predicted that production would enhance both cued recall and associative recognition.

Method

Subjects There were 62 subjects in Experiment 4. Of these, 40 were from Washington University in St. Louis, and 22 were from the University of Waterloo. Both sites ran identical experiments, so the two samples of subjects were combined.

Materials The subjects studied unrelated word pairs. The unrelated words from Experiment 1B and just the cue items from Experiment 1A were combined into one pool. From this pool, 20 random word pairs were created for each subject.

Procedure In Experiment 4, we used the same general procedure as we had used in the earlier experiments. However, only the read-aloud and read-silently conditions were used (i.e., the typed condition was omitted), with no semantic-rating task after encoding. In line with earlier work on the production effect (MacLeod et al., 2010), the stimuli appeared in a blue or white font on a black background. Subjects read the blue pairs aloud and the white pairs silently. After studying all of the word pairs, subjects took a cued recall test similar to the test in Experiment 1.

Then subjects took an associative recognition test. In this test, half of the word pairs (ten total: five silent and five aloud) were presented exactly as studied. The other ten were rearranged pairs: The cue from one pair was combined with the target from a different pair. When rearranging pairs, words were always combined with a word that had been processed in the same way at study (i.e., both words were from aloud pairs or both were from silent pairs). Thus, subjects were exposed to ten intact pairs and ten rearranged pairs (five rearranged aloud pairs and five rearranged silent pairs), one pair at a time. Subjects were asked to identify whether they thought each pair was intact or rearranged by a keypress—“z” for “old” and “m” for “new.” Subjects were not informed that both members of rearranged pairs had always been processed in the same way at study. No time limit was imposed for the associative recognition test. The entire experiment took about 20 min.

Results and discussion

The results reported are for the combined data sets from the University of Waterloo and Washington University. Analyzing each data set individually yielded the same pattern of results.

Cued recall Reading a pair aloud ($M = .29$, $SEM = .03$) yielded better cued recall performance than did reading silently ($M = .18$, $SEM = .02$). A paired-samples t test confirmed that this difference was significant, $t(61) = 5.25$, $p < .001$, $d = 0.54$. Production enhanced cued recall, replicating the

finding that production enhances paired-associate learning—as long as no semantic-relatedness rating task follows encoding.

Associative recognition The first row of Table 3 presents the proportions of “old” responses on the associative recognition test. Responding “old” to an intact pair constituted a hit, and responding “old” to a rearranged pair represented a false alarm. Production helped in correctly identifying intact pairs: Saying a pair aloud yielded more hits than did reading a pair silently, $t(61) = 3.55$, $p = .001$, $d = 0.62$. Production did not, however, reduce false alarms for the rearranged pairs: Saying a pair aloud yielded a false alarm rate—.32—identical to that for reading a pair silently, $t(61) = 0.08$, $p = .940$, $d = 0.25$. Further analyses revealed that d' was higher for read-aloud items ($M = 1.07$, $SE = 0.04$) than for silently read items ($M = 0.79$, $SE = 0.05$), $t(61) = 2.20$, $p = .032$, $d = 0.38$, confirming the conclusion that production enhanced discriminability.

Of course, the results of the associative recognition test were likely influenced by performance on the prior cued recall test. Retrieving an item should make it easier to remember that item again in the future (see Roediger & Karpicke, 2006, for a review of testing effects): Because subjects recalled more aloud items than silent items on the cued recall test, the enhanced discrimination seen for the intact aloud pairs over the intact silent pairs could be due to retrieval practice effects. One way to overcome this concern would be to examine performance on the associative recognition test for pairs that were not recalled. These data are shown in Table 4, although they must be interpreted with caution, due to the low numbers of observations (3.88 observations per item type per subject). For the intact pairs not correctly recalled, recognition of read-aloud pairs remained better than recognition of read-silently pairs, $t(60) = 1.94$, $p = .057$, $d = 0.36$. For the rearranged pairs not correctly recalled, subjects still committed false alarms at the same rate for aloud and silent pairs, $t(59) = 0.25$, $p = .804$, $d = 0.03$. Analyses using d' revealed that, consistent with the overall analysis, the aloud pairs ($M = 0.75$, $SE = 0.13$) led to numerically greater discrimination than did the silent pairs ($M = 0.59$, $SE = 0.13$), although the difference was not significant, $t(59) = 1.13$, $p = .265$, $d = 0.21$. Given that the same general pattern of results occurred with both the full and conditional analyses, and given the small number of observations for the conditional analyses, at least some benefit of production seems to have occurred for the intact pairs, but not for the rearranged pairs.¹

The outcome of Experiment 4 indicates that production enhances the cue-to-target associative link. If production only enhanced item-specific information, performance should have

been equivalent regardless of whether the pair was intact or rearranged, because subjects would be responding only on the basis of the distinctiveness of the individual items. However, this was not the case. In saying the pair aloud, subjects appeared to have created a more distinctive record for that pair; this conclusion is supported by the finding that when studied pairs were broken up (with both words in the rearranged pairs being from the same original encoding condition, read aloud or read silently), false alarm rates were identical for aloud and silent pairs.

Experiment 5

Experiment 5 was identical to Experiment 4, except that the initial test was free recall (of cues and targets) instead of cued recall. We expected to replicate the pattern seen in Experiment 4.

Method

Subjects There were 64 subjects in Experiment 5: 40 from Washington University in St. Louis, and 24 from the University of Waterloo.

Materials and procedure The materials and procedure were identical to those in Experiment 4, except that the cued recall test was replaced with a free recall test. Subjects were told to recall as many words as they could, both cues and targets. Subjects typed each word that they remembered into a text box and pressed Enter to save their response. When they could not recall any more words, subjects took the associative recognition test.

Results and discussion

As with Experiment 4, the results are for the combined data set; however, the same pattern emerged when the data from the University of Waterloo or from Washington University were analyzed. As anticipated, the general pattern of results was similar to that found with the cued recall test in Experiment 4.

Free recall A production effect emerged in free recall. As compared with reading a word pair silently at study ($M = .16$, $SEM = .02$), producing a word pair aloud ($M = .27$, $SEM = .02$) yielded better recall of the targets, $t(63) = 4.65$, $p < .001$, $d = 0.71$. The same pattern was found when just cues or both cues and targets were analyzed. This outcome replicates previous work showing that the production effect enhances free

¹ To gain more power, we combined the d' results for Experiments 4 and 5. A 2×2 mixed-model ANOVA yielded a significant main effect of production, $F(1, 122) = 4.29$, $p = .041$, $\eta^2 = .03$, but did not reveal a significant main effect of experiment or a significant interaction.

Table 3 Proportions of “old” responses on the associative recognition test following prior cued recall (Exp.4) or free recall (Exp.5), as a function of study condition

Condition	Aloud Intact		Silent Intact		Aloud Rearranged		Silent Rearranged	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
E4: After Cued Recall	.83	.02	.69	.03	.32	.03	.32	.03
E5: After Free Recall	.85	.02	.66	.03	.37	.04	.28	.03

Intact responses are equivalent to hits, and rearranged responses are equivalent to false alarms.

recall performance (Conway & Gathercole, 1987; Lin & MacLeod, 2012; MacLeod, 2011).

Associative recognition The second row of Table 3 presents the proportions of “old” responses on the associative recognition test. For intact pairs, read-aloud items yielded a higher hit rate than did silently read pairs, $t(63) = 5.33$, $p < .001$, $d = 0.88$. Unexpectedly, for the rearranged pairs, subjects committed more false alarms for the aloud pairs than for the silent pairs, $t(63) = -2.15$, $p = .036$, $d = 0.31$. However, we also examined d' , which revealed that production led to a higher d' ($M = 1.03$, $SE = 0.04$) than did reading silently ($M = 0.79$, $SE = 0.04$), $t(63) = 2.40$, $p = .018$, $d = 0.33$, which indicates that production did increase discriminability and did not just cause a criterion shift.

We were again concerned that performance on the free recall test could influence performance on the associative recognition test. Because subjects could recall both members of the pair, this issue is potentially more trouble than in Experiment 4. Table 4 presents performance on the associative recognition test for only those items that were not recalled on the free recall test. As before, the small number of observations in each cell (3.92 items per type) requires interpreting these results with caution. For the intact pairs not recalled, aloud items were recognized better than silent items, $t(63) = 5.58$, $p < .001$, $d = 0.90$, and for the rearranged pairs, subjects

committed more false alarms for the aloud items than the silent items, $t(63) = 2.05$, $p = .044$, $d = 0.36$. Comparing d' for the two conditions revealed that the aloud pairs ($M = 0.84$, $SE = 0.08$) were discriminated better than the silent pairs ($M = 0.64$, $SE = 0.08$), and that this difference approached significance, $t(63) = 1.93$, $p = .058$, $d = 0.29$. This pattern replicated the pattern found when all items were included in the associative recognition test analyses, providing further evidence that production increased both hits and false alarms.

The correct recognition of intact pairs replicates the results from Experiment 4, which indicated more hits for intact pairs after production than after silent reading. As before, we can interpret the production effect for the intact pairs as an indication that production helped to strengthen the associative link between the cue and the target. Unexpectedly, subjects were better at rejecting rearranged silent pairs than rearranged aloud pairs (a kind of “negative production effect”). Perhaps previously recalling some cues and some targets increases the familiarity of both members of studied pairs, so that subjects would false alarm to the rearranged pairs more after production than after silent reading, because both members of rearranged pairs seemed so familiar after production and recall (even though the cue and the target were produced separately). If so, this would explain why the same pattern did not occur in Experiment 4, where only targets, and not cues, were recalled when cued recall was the initial test format. Regardless of why increased false alarms may have occurred for the aloud items, however, the d' and conditional analyses supported the conclusion that discrimination is better for read-aloud pairs than for read-silently pairs.

Table 4 Proportions of “old” responses on the associative recognition test, given a prior failure to recall on cued recall (Exp.4) or free recall (Exp.5), as a function of study condition

Condition	Aloud Intact		Silent Intact		Aloud Rearranged		Silent Rearranged	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
E4: After Cued Recall	.74	.04	.64	.03	.35	.04	.34	.04
E5: After Free Recall	.84	.02	.62	.03	.40	.04	.30	.03

Intact responses are equivalent to hits, and rearranged responses are equivalent to false alarms. These data must be interpreted with caution, as few observations were available for each item type (averages of 3.88 items per cell for Exp.4 and 3.92 items per cell for Exp.5).

General discussion

The two main questions motivating these experiments were (1) whether the production effect could occur in paired-associate learning, and (2) if so, whether production would enhance item-specific information, relational information, or both. We also examined the effect of a semantic-rating task on recall and whether multiple forms of production at study

(reading aloud and typing) would each enhance recall (and if so, whether reading aloud or typing would lead to stronger effects).

We reported several interesting new results. First and foremost, we showed a production effect in paired-associate recall in Experiments 3A and 4. Experiment 3 indicated that the failure to find any benefit of production in Experiments 1 and 2 was due to the inclusion of a semantic-rating task in the earlier experiments that had overshadowed the value of production. Second, Experiments 3A and 3B, which included both typing and reading aloud as forms of production, yielded results consistent with those of other published studies (Forrin et al., 2012) indicating that reading aloud yields a larger production effect than does typing. Finally, and most important, production benefited both paired-associate recall and pair recognition, indicating that production enhances the associative link between the two elements of a pair, a component of paired-associate learning that is essential for successful cued recall (McGuire, 1961).

Experiment 3 showed that the semantic-rating task in Experiments 1 and 2 prevented the production effect from emerging. One explanation is that semantic rating of pairs is so powerful a manipulation relative to production that it swamped any effect of production, as was discussed in a previous section. On the other hand, MacLeod et al. (2010) and Forrin et al. (2012) have shown that production can enhance retention above and beyond a generation task, which suggests that the way in which production enhances associative information is different from how it enhances item-specific information. One difference between our research and prior work that had also used deep-encoding tasks (Craik & Lockhart, 1972) is that production occurred before the “deep” manipulation in our research, but occurred after the manipulation in the prior work. Whether this difference is critical or whether semantic-relatedness judgments in paired-associate learning simply render production ineffective must await future research.

Until now, production has only been shown to benefit item memory. In that context, the distinctiveness account (Conway & Gathercole, 1987; MacLeod et al., 2010) provides a clear mechanism through which such an advantage would occur. Namely, if a subject can remember having said a word aloud, that word was likely to be studied and could be recognized as “old.” The findings of the present experiments reveal that production also provides a benefit to associative memory, yet this benefit does not appear to be readily interpreted within current conceptualizations of distinctiveness that have emphasized item-specific distinctiveness. How, then, does a distinctiveness account explain the benefits of production on associative memory?

One possible explanation is that production may encourage distinctive associative encoding that assists later recollection, or at least that it may do so more than does reading a word pair silently (cf. Ozubko et al., 2012). Yonelinas (2002) argued that recollective processes are critical for successful performance on associative recognition tasks, so perhaps production of a

pair enhances recollection, which in turn assists with associative recognition. Under such an account, reading a word aloud encourages distinctive encoding by producing a more unitized event, owing to speaking both words aloud; stated differently, when spoken aloud, the two words form a greater associative bond than they do when they are both read silently, which in contrast may lead to greater processing of words individually (but not their associative relation). Furthermore, such encoding permits recollective retrieval. Thus, producing a word or word pair is more likely to produce a distinctive trace that, at test, can assist recollection in associative memory tasks. If subjects are also given semantic-relatedness tasks, as in Experiments 1–3, they form a powerful distinctive association, and perhaps that is why that production no longer adds any benefit.

In sum, the experiments reported here extend the production effect from a mnemonic benefit for item information alone to a benefit for both item and associative information. The results also show a boundary condition for the production effect in paired-associate learning: A semantic-relatedness task that follows speaking and silent reading of the pairs eliminates the effect. Although this manipulation provides a limit to the influence of production on associative memory, understanding how production influences associative recall and recognition may provide key insights into this phenomenon, as well as highlight new ways that production may be useful in educational settings. The data provided here are broadly consistent with a distinctiveness account of the production effect, although fully delineating the mechanism behind production’s benefit to associative memory will require further work. Regardless, the present research highlights that the production effect is a straightforward, robust, and effective mnemonic that can be used to enhance memory for both single items and paired associates, implicating roles in both item and associative learning.

Author note This research was supported by the James S. McDonnell Foundation and by National Science Foundation Grant No. DGE-1143954 at Washington University in St. Louis, and by Natural Sciences and Engineering Research Council Discovery Grant No. A7459 at the University of Waterloo. We thank Christian Gordon, Aleks Husik, and Elisabeth Martin for assisting with the data collection at Washington University, and Merrick Levene for assisting with the data collection at the University of Waterloo. J.D.O. is now at Rotman Research Institute, Baycrest Centre, Toronto, Ontario. Correspondence may be directed to any of the authors: adam.putnam@wustl.edu, jozubko@research.baycrest.org, cmacleod@uwaterloo.ca, or roediger@wustl.edu.

References

- Bodner, G. E., & Taikh, A. (2012). Reassessing the basis of the production effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1711–1719. doi:10.1037/a0028466
- Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition*, 41, 28–35. doi:10.3758/s13421-012-0249-6

- Conway, M., & Gathercole, S. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*, 341–361. doi:10.1016/0749-596X(87)90118-5
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. doi:10.1016/S0022-5371(72)80001-X
- Dodson, C. S., & Schacter, D. L. (2001). “If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161. doi:10.3758/BF03196152
- Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*, 1–5. doi:10.1016/j.actpsy.2012.10.001
- Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (in press). Production improves memory equivalently following elaborative vs. non-elaborative processing. *Memory*. doi:10.1080/09658211.2013.798417
- Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, *40*, 1046–1055. doi:10.3758/s13421-012-0210-8
- Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, *16*, 110–119. doi:10.3758/BF03213478
- Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 484–494. doi:10.1037/0278-7393.14.3.484
- Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennett B. Murdock* (pp. 227–248). Hillsdale: Erlbaum.
- Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition*, *27*, 657–664. doi:10.3758/BF03211559
- Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, *11*, 534–537. doi:10.1016/S0022-5371(72)80036-7
- Hourihaan, K. L., & MacLeod, C. M. (2008). Directed forgetting meets the production effect: Distinctive processing is resistant to intentional forgetting. *Canadian Journal of Experimental Psychology*, *62*, 242–246. doi:10.1037/1196-1961.62.4.242
- Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition*, *4*, 221–232. doi:10.3758/BF03213167
- Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 3–25). New York, NY: Oxford University Press.
- Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, *22*, 10–15. doi:10.1177/0963721412463228
- Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*, 421–445. doi:10.1006/jmla.1993.1023
- Jones, A. C., & Pyc, M. A. (in press). The production effect: Costs and benefits in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0033337
- Lin, O. Y. H., & MacLeod, C. M. (2012). Aging and the production effect: A test of the distinctiveness account. *Canadian Journal of Experimental Psychology*, *66*, 212–216.
- MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, *18*, 1197–1202. doi:10.3758/s13423-011-0168-8
- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685. doi:10.1037/a0018785
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, *15*, 237–255. doi:10.3758/PBR.15.2.237
- McGuire, W. J. (1961). A multiprocess model for paired-associate learning. *Journal of Experimental Psychology*, *62*, 335–347.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533. doi:10.1016/S0022-5371(77)80016-9
- Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 45, pp. 175–214). San Diego, CA: Elsevier.
- Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 27–46). New York, NY: Oxford University Press.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from <http://w3.usf.edu/FreeAssociation/>
- Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2011). Production benefits both recollection and familiarity. *Memory & Cognition*, *40*, 326–338. doi:10.3758/s13421-011-0165-1
- Ozubko, J. D., Hourihan, K. L., & MacLeod, C. M. (2012). Production benefits learning: The production effect endures and improves memory for text. *Memory*, *20*, 717–727. doi:10.1080/09658211.2012.699070
- Ozubko, J. D., & MacLeod, C. M. (2010). The production effect in memory: Evidence that distinctiveness underlies the benefit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1543–1547. doi:10.1037/a0020604
- Ozubko, J. D., Major, J., & MacLeod, C. M. (in press). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*. doi:10.1080/09658211.2013.800554
- Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. doi:10.1111/j.1745-6916.2006.00012.x
- Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*, 6–10. doi:10.3758/BF03202594
- Yonelinas, A. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517. doi:10.1006/jmla.2002.2864