Psychonomic Bulletin & Review (2024) 31:373–379
https://doi.org/10.3758/s13423-023-02360-9

BRIEF REPORT

# The prod eff: Partially producing items moderates the production effect

Megan O. Kelly[1] · Tyler M. Ensor[2] · Colin M. MacLeod[1] · Evan F. Risko[1]

Accepted: 3 August 2023 / Published online: 24 August 2023


**Abstract**
Current accounts of the production effect suggest that production leads to the encoding of additional production-associated features and/or better feature encoding. Thus, if it is the act of production that leads to the storage and/or enhanced encoding of these features, then less of this act should reduce the resulting production effect. In two experiments, we provide a direct test of this idea by manipulating how much of a given item is produced within a single mode of production (typing). Results demonstrate that such partial production can yield a significant production effect that is smaller than the effect that emerges from producing the entire item. These results suggest that how much of an item is produced can moderate the size of the production effect and are considered in the context of recent modelling efforts.

**Keywords** Memory · Production effect · Recognition

## Introduction

Active engagement with study material (e.g., generating, drawing, and enacting; for reviews, see Bertsch et al., 2007; Fernandes et al., 2018; Roberts et al., 2022) generally benefits memory performance compared to more passive engagement (e.g., reading silently). One reliable form of active engagement is producing study information (e.g., via reading aloud, spelling, writing, mouthing, whispering, singing, or typing; Forrin et al., 2012; Quinlan & Taylor, 2013, 2019; MacLeod et al., 2010; Zhou & MacLeod, 2021; for a brief review, see MacLeod & Bodner, 2017). When individuals produce to-be-remembered information, they remember that information better than information that is not produced (e.g., only read silently) – this is the *production effect* (MacLeod et al., 2010).

Multiple theoretical accounts have been proposed to explain the production effect. Forrin et al. (2012 ; also supported by MacLeod et al., 2010) proposed that the act of producing a study item leads to the formation of an item-specific and "distinctive" record in memory (see *relative distinctiveness*; Conway & Gathercole, 1987). Here, "distinctive" means that produced items have *additional* item-specific production-associated information stored in memory that differentiates them from items studied silently – and from other items studied via production. Production often consists of motoric (e.g., moving one's mouth, typing with one's fingers) and/or perceptual (e.g., hearing one's voice) features that silent reading does not entail. These aspects are thought to be encoded along with the unique item information (e.g., the meaning of the word), and stored in memory. Critically, these features are retrievable at test, leading to produced items generally being more retrievable than silent items. Computational instantiations of the production effect (e.g., Jamieson et al., 2016; Kelly et al., 2022; Saint-Aubin et al., 2021) have used this general "additional features" idea to explain the production effect. Another explanation of the production effect suggests that production *strengthens* item representations relative to silent reading (e.g., Bodner & Taikh, 2012; Fawcett, 2013). Within the computational models of the production effect, this is implemented by increasing the encoding quality for produced items, making them more accessible than non-produced items at retrieval.

Importantly, the *act* of production is central in each of these accounts. In "additional features" accounts, the act of production leads to the storage of the additional features. In the strength-based account, some putative requirement

✉ Megan O. Kelly
megan.kelly@uwaterloo.ca

1 Department of Psychology, University of Waterloo, Waterloo, ON N2L 3G1, Canada

2 Department of Psychology, California State University, Bakersfield, CA, USA

of the act of production (e.g., the need to devote more attention) could lead to a higher probability that a feature is accurately encoded and, thus, available at test. But what about the act of production leads to the storage of these "additional features" or this "strengthening"? Here we tested the idea that how much of the study information is produced via the production act determines the storage of these "additional features" or this "strengthening," and as such determines the magnitude of the production effect. Using typing as the mode of production, we tested this prediction by comparing the magnitude of the production effect as a function of the number of letters produced. To do so, we compared conditions where participants typed all the letters of the word to conditions where they typed only some of the letters.

A different idea is that the storage of "additional features" or "strengthening" caused by production is driven not by how *much* of a given item is produced, but by the number of distinct or unique features (e.g., auditory, motoric) in the production act itself. A similar idea has been suggested in the context of the drawing effect, wherein memory is proposed to be enhanced by drawing largely due to the integration of unique motoric, pictorial, and elaborative components putatively combined within the act of drawing (Fernandes et al., 2018). On this type of account, the magnitude of the production effect would be modulated by the number of unique dimensions of the production act, not necessarily by the amount of production within a given dimension (e.g., how much of an item was typed).

One potentially relevant empirical observation is the modulation of the production effect by the production method (Jamieson et al., 2016; Forrin et al., 2012). Although a production effect results from mouthing or whispering the items at study, these production effects are smaller than those resulting from reading aloud (Forrin et al., 2012). Jamieson et al. accounted for this modulation by production method using their computational instantiation of an additional features model. Forrin and colleagues suggest that certain modes of production allow for the processing of information along unique dimensions not shared with other modes of production or through qualities inherent to the production modes being compared that affect production distinctiveness (e.g., differences in strength of auditory signal, degree of active encoding required, variety of vocal responses available vs. spectral flatness). For example, they suggested that speaking allows for the processing of information along extra dimensions compared with other production modes (e.g., mouthing and whispering), thereby giving rise to production effects for speaking that are more pronounced than for others. Critically, the production mode effect always involved producing the entire item, leaving open the question of whether how much of the item is produced modulates the magnitude of the production effect.

## Experiment 1

Experiment 1 (E1) was originally two experiments, Experiments 1a (preregistered at osf.io/k3h8j) and 1b (preregistered at osf.io/g2jpd), the latter being a replication of the former. Because they were identical, we deviate from the preregistered analyses that treated them separately and instead present them together as one experiment. The original analyses in which they are separately examined are available in the Online Supplementary Materials (OSM). We manipulated production within-participants across four levels: Participants typed ALL of the word (i.e., standard production), PART of the word (i.e., partial production; either the FIRST-3 or LAST-3 letters), or NONE of the word (i.e., no production). Data and analysis code are available at osf.io/rsgpe.

### Method

#### Participants

We collected and analyzed data from $N = 250$ participants on PROLIFIC who were paid GBP 3.75 (USD ~4.80) for their time and who were current residents of Canada, the USA, or the UK. This was double the number needed to detect a small-to-medium main effect of Cohen's $f = .29$ at 99% power within a repeated-measures one-way ANOVA with four levels, and ~80% power to detect paired-effect sizes as small as .25; using the Superpower *ANOVA_exact* shiny app, two-tailed, alpha = .05, correlation among repeated measures = .50 (Lakens & Caldwell, 2021).

#### Stimuli

Two 60-word lists were created by sampling all six-letter words with word frequencies of 51–54,159 (FREQcount from the Open Lexicon Project; Brysbaert & New, 2009). Because each item had an equal chance of having each of the four typing cues assigned, no items had the same first three or final three letters, thereby ensuring unique productions in all production conditions. With equal probability across participants, one list was assigned as the list of targets and the other as the list of foils. For each word on the randomly assigned target list, the four types of cues (i.e., to type ALL, FIRST 3, LAST 3, or NONE) were assigned with equal probability such that each type of study cue appeared 15 times.

An additional consideration is the potential influence of the production on item distinctiveness. Here, we used an item set in which all partial (and full) productions were unique from all others in the item set (e.g., the partial productions CHO and PLA are distinct from each other just as the full productions CHORE and PLANT are distinct from each other). This provides a test of a strong distinctiveness account wherein the production effect requires only a distinct production – where "distinct" is defined within the experimental context. If all that is required is a distinct production, then, in the present stimulus set, the production effect should be independent of the amount of production (i.e., partial vs. full). Finally, we kept target length constant (i.e., six letters) and manipulated how much of an item was produced (when it was). How the absolute amount of production (e.g., producing a three- vs. a six-letter word) might influence the resulting production effect is addressed in the *General discussion*.

### Procedure

Participants were instructed that they would be presented a series of words to study for a later memory test. They were told that each word, presented one at a time, would have a typing prompt simultaneously right below it in smaller font to which they were to adhere. In the instructions, they were presented with each possible typing prompt (*ALL*, *FIRST 3*, *LAST 3*, and *NONE)* and an example of what to type given each prompt if the study word was *answer* (type "answer", type "ans", type "wer", type *nothing*). Participants were instructed to remember all of the words as best as they could, regardless of the typing prompt. When they were done with the instructions, they responded to a multiple-choice comprehension check about how to treat the typing prompts and were always provided the correct response as feedback: "[*That is correct/incorrect*]. *It is important that you follow the typing prompts and try to remember the words.*" After this check, they advanced to the study phase in which words were presented for 5,000 ms each, separated by 100-ms intervals. For each word during the study phase, participants typed the to-be-remembered word according to the given prompt into a text field directly below the study item and its prompt, thus showing what they were typing as they typed it. Participants had the full 5,000 ms to type, including correcting typos.

When the study list was completed, participants were given instructions for the recognition test and, on the same screen, given a 20-s countdown to the memory test. In the recognition test, participants were presented with the 60 targets and 60 distractors one at a time in random order, to which they responded either with "old" (i.e., the item had been studied) or "new" (i.e., the item had not been studied). Each test word stayed on the screen until the participant responded, and a 300-ms blank screen separated successive items. Upon test completion, participants were asked the following questions to check data quality: (i) "Did you take any notes or write anything down while completing the task?", (ii) "Were you doing anything else while completing this task? (e.g., watching Netflix)", and (iii) "Is there any reason we should or should not use your data? (It's okay if you think you weren't able to give it your best, just let us know)"

## Results

Figure 1 presents the hit rates by production condition and experiment. As preregistered, we excluded and replaced the data of individuals who did not follow production instructions on more than 10% of trials (20 participants), and excluded trials wherein participants did not follow production instructions (2.1% of the data).[1] Three participants were replaced due to reporting that they were doing something else while completing the task; one participant was replaced due to reporting that they took notes during the study.

In addition to collapsing across Experiments 1a (E1a) and 1b (E1b) to report the results of Experiment 1, we also deviate from the preregistered analyses of one-way ANOVA and paired-samples t-tests, instead reporting mixed effects regression with by-item and by-participant random intercepts, and random slopes for production conditions.[2] This allowed us to treat the production condition as continuous (i.e., such that the difference between partial and full is not treated the same as the difference between none and full). Any deviation from this random-effects structure is noted.

We used mixed effects regression to compare typing the *first* versus the *last* three letters on hits and found no significant difference [$B = 0.05$, $SE = 0.07$, $z = 0.80$, $p = .424$; random by-participant slopes removed to prevent singular fitting]. Therefore, we collapsed across these two conditions hereafter, referring to them collectively as the *partial* condition (i.e., typing three of six letters; see OSM for original analyses wherein we obtain the same results).

Mixed-effects regression revealed that producing (i.e., typing) more letters led to more hits [$B = 0.11$, $SE = 0.01$, $z = 11.52$, $p < .001$]. This was true for the ALL-NONE comparison [$B = 0.11$, $SE = 0.01$, $z = 11.42$, $p < .001$], the PARTIAL-NONE comparison [$B = 0.17$, $SE = 0.02$, $z = 9.34$, $p < .001$], and the ALL-PARTIAL comparison [$B = 0.05$, $SE = 0.02$, $z = 2.47$, $p = .014$].

[1] Note that our criterion for following production instructions was strict such that we did not excuse typos in any condition even if only one letter was incorrect. This was because it seemed less clear how to accommodate typos while maintaining the integrity of the partial productions. Despite this, production accuracy was near ceiling for each condition (e.g., across experiments, All: 96.5%; First-3: 97.6%; Last-3 96.8%; 2-Letter: 98.5%; None: 99.3%).

[2] In the original analyses separating Experiments 1a and 1b, the partial production effect was not significantly different from the full production effect in E1a (although numerically smaller) but was reduced in E1b. Further results and information on analyses are available in the OSM.

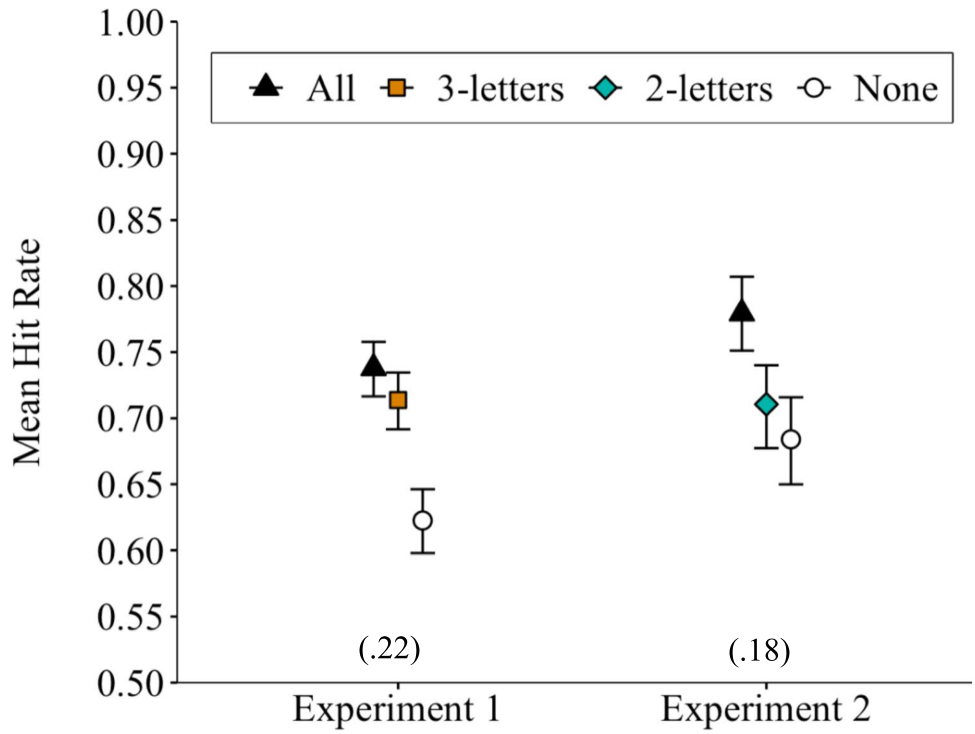


**Fig. 1** Mean proportion of targets recognized by production condition and experiment. *Note.* Error bars are bias-corrected accelerated bootstrap 95% confidence intervals using 10,000 replications. Values in parentheses are the false alarm rates for each experiment

### Discussion

We examined how amount of item produced influenced memory performance and found a significant production effect when comparing a standard typing production (i.e., typing the full word) with a silent reading condition, thus replicating the typing production effect (Bodner et al., 2016; Forrin et al., 2012; Jamieson & Spear, 2014). Interestingly, we also found that only *partially* producing an item (i.e., typing out the first or last three letters of a six-letter item) led to a production advantage (producing the first and last three letters also did not differ) that was significantly smaller than when typing the full item. Thus, production can lead to a sizeable influence on memory even when that production is partial. We tested this further in Experiment 2 by reducing the number of letters typed in the partial condition.

## Experiment 2

In Experiment 2 (E2), we reduced the partial typing production condition from three letters to two letters. Partial productions were of the first two letters of the six-letter item and all productions were unique. E2 was preregistered on the Open Science Framework at osf.io/evm56 and the data and analyses code are available at osf.io/rsgpe.

### Method

The method was largely the same as in Experiment 1. Data were collected and analyzed from $N = 125$ participants on PROLIFIC who were paid GBP 3.75 (USD ~4.80) for their time and were current residents of Canada, the USA, or the UK. This was more than the number needed to detect a small-medium main effect of $f = .35$ at 99% power within a repeated-measures one-way ANOVA with three levels, and ~80% power to detect paired effect sizes as small as d = 0.25. As before, this was determined using the Superpower ANOVA_exact shiny app (two-tailed, alpha = .05, correlation among repeated measures = .50).

The main change in E2 is that we altered the partial production condition from typing three letters to typing only the first two letters (of six-letter words), thus the three type prompts presented equally often were *ALL, FIRST 2,* and *NONE*. To ensure that the two-letter productions of each item were distinct from other items, the items of E2 were redrawn from the same pool as before (exceptions to this are highlighted in the materials available at osf.io/rsgpe; 55 words overlapped between experiments). Because there were fewer unique combinations of two-letter items compared to three-letter items, we used two 57-item lists, thus 114 items were presented at test.

## Results

Again, we replaced the data of individuals who did not follow production instructions on more than 10% of trials (16 participants) and excluded trials wherein participants did not follow production instructions (2.3%). See Fig. 1 for the mean hit rate as a function of production type. As in Experiment 1, we deviated from the preregistration by reporting mixed-effects regression (following the same random-effects structure and noting any deviations).

Mixed-effects regression revealed that the more letters produced (i.e., typed), the higher the hits [$B = .11$, $SE = 0.02$, $z = 7.10$, $p < .001$]. Again, this was significant for the ALL-NONE comparison [$B = 0.10$, $SE = 0.02$, $z = 6.43$, $p < .001$], the ALL-PARTIAL comparison [$B = 0.11$, $SE = 0.02$, $z = 6.09$, $p < .001$; random by-participant slopes removed to prevent singular fitting], and marginally for the PARTIAL-NONE comparison [$B = 0.07$, $SE = 0.04$, $z = 1.71$, $p = .087$].

## Discussion

Consistent with Experiment 1, there was a significant production advantage when comparing a standard typing production to no production. With only two letters typed, the partial production effect was marginal, and performance was poorer than in the condition where participants typed all of the letters. Thus, partial production does not always lead to a production effect that is equivalent to the production effect that emerges when the entire item is produced, which is generally consistent with extant accounts of the production effect.

## Modelling

A final consideration is how the current results fit in relation to extant computational models of the production effect. As noted earlier, Jamieson et al. (2016) modelled the modulation of the production effect (in MINERVA2) by attributing more production-associated features as a function of production type (e.g., no production vs. mouthing vs. saying aloud). Thus, the model can account qualitatively for the present results if we assume "more" production leads to the storage of more production-associated features.

The Kelly et al. (2022) model using REM (REM.1 specifically, the most common version of REM; Shiffrin & Steyvers, 1997) is like the model by Jamieson et al. (2016) in that production leads to additional stored features. However, in the Kelly et al. model, production-associated features are included in the probe vectors and "presented" to memory at test. This was to capture the idea that individuals can use the production to probe memory (e.g., "Did I say this aloud?"). Using the assumption that "more" production leads to the storage of more production-associated features, we attempted to simulate the present results in the Kelly et al. model (see OSM for details). Critically, each item had 20 base features and ten possible production-associated features. In the full production condition, all ten features were included; in the silent condition, no features were included. We ran 1,000 simulations with partial conditions ranging from one to nine features and included the theoretical endpoints of zero and ten features to illustrate the changes from zero to one and nine to ten features, thus, running 11,000 simulations (model code and data available on osf.io/rsgpe). As with Kelly et al., probe vectors are "presented" to memory at test with all the potential production-associated features. The results presented in Fig. 2 appear to capture the qualitative result that the production effect increases as the number of production-associated features increases. While both the Jamieson et al. and Kelly et al. models seemingly capture the empirical pattern reported here, the (strong) assumptions made, for example, about how "more" production is translated into representations in memory, warrant further consideration.

# General discussion

We examined how partial production influences memory performance. Across all experiments, a relative memory advantage was conferred by standard production at study (i.e., typing the full study word). Critically, this production effect was modulated by how much of the target item was produced (ALL vs. PARTIAL vs. NONE). When producing 3/6 letters (E1), a significant production effect was observed that was significantly smaller than the standard production effect. When typing 2/6 letters (E2), the partial production effect was only marginally significant, and clearly smaller than the standard effect. These results demonstrate that the production effect is moderated by how much of the item is produced.

Against the background of extant accounts of the production effect, these results are consistent with the idea that the storage of production-associated features or strengthening is modulated by how much "production" is engaged in within a given mode of production. While it also remains possible that engaging in additional modes of production would also confer an advantage, the present results establish that this is not necessary – more of the same mode of production leads to a larger production effect.

We also demonstrated that one model of the production effect, based on the idea that production adds features, can capture the pattern reported here (and that another model based on a similar idea would also likely capture the present results). The model captured the pattern by assuming that production of each "letter" translated into additional

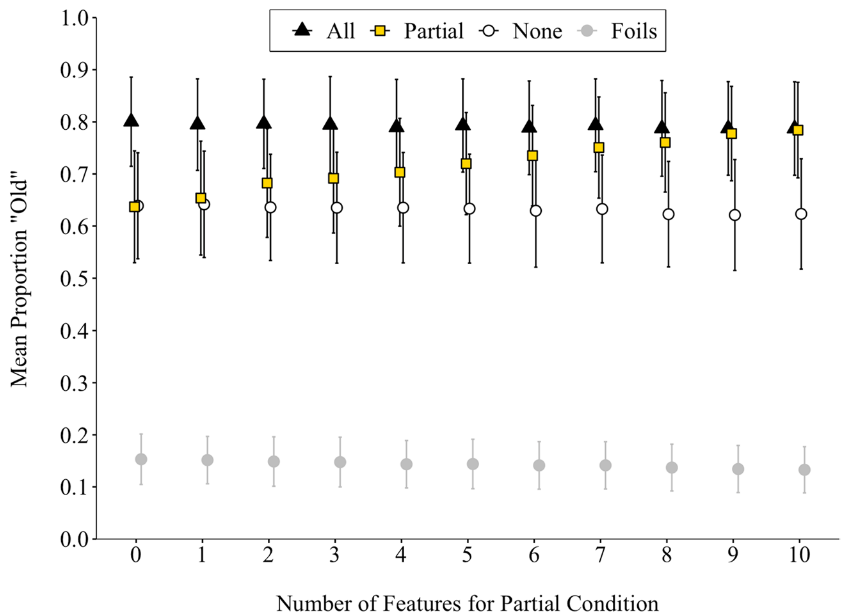


**Fig. 2** Production-associated features account in REM: Hit rate as a function of production with false alarm rates, all as a function of the number of features available in the Partial Production condition. Error bars are ± the standard deviation

features. The probe in this model contained all of the features, thus the "match" to the probe improves as the number of produced letters increases. This assumption raises an important question about how the absolute number of letters in a (typing) production modulates the production effect. We examined partial production in a context wherein the absolute target length was constant. Thus, partial production involved less production in an absolute sense (e.g., three vs. six letters) and a reduced proportion of the item (e.g., 50% vs. 100%). Is our observed modulation of the production effect by *partial* production driven by the smaller absolute number of letters produced *or* by the partial production being an incomplete representation of the *full* study item?

Early evidence from an unpublished experiment in our laboratory examining the standard typing production effect in an item set wherein items varied in number of letters suggests no clear effect of target length on the production effect when those items varied from five to eight letters.[3] While these lengths do not include the lengths used here (i.e., two vs. three vs. six letters), if there is no absolute length effect, this would suggest that the partial production effect is due to the proportion produced rather than the absolute amount produced. This idea would also provide an important challenge for models of the production effect. As noted above, the model presented here translated the number of letters produced into the number of additional features stored. This assumption would seemingly predict an influence of the absolute number of features. Thus, further investigation into the nature of the production promises deeper insight into how production influences memory.

It is also worth noting that the present results are inconsistent with a distinctiveness account wherein a constant production effect is generated if the production is distinct (ignoring how distinct the item might be) in the context of the list. As noted in the *Method* section, the items used here yielded productions that were unique in the context of the list. Although such an account would rely on a simplified conceptualization of distinctiveness, providing evidence inconsistent with it clearly demonstrates a need to establish what distinctiveness means in the context of distinctiveness accounts of the production effect.

## Conclusion

While previous work has clearly demonstrated a robust effect of producing an entire item (i.e., reading a word aloud, typing it, singing it), we report a novel form of the production effect – a *partial production* effect. This effect can be smaller than the standard production effect, but it demonstrates that complete production is not necessary for a relative memory benefit.

[3] We examined item length (ranging from five to eight letters) – thus, absolute amount produced, as a predictor of the production effect and found no relation between number of letters produced and the production effect, $b < .01$, $SE = 0.01$, $t = 0.60$, $p = .552$. Note that this work was not designed to address the effect of word length on the production effect and that these word lengths (five to eight) differed from those of the present work (two and three letters).

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02360-9.

**Open practices statement** All the data and analyses code are available via the Open Science Framework at: osf.io/rsgpe. The preregistrations of Experiments 1a, 1b, and 2 are available here, respectively: osf.io/k3h8j, osf.io/g2jpd, and osf.io/evm56.

**Author note** This work was supported by a Discovery Grant (#04091) from the Natural Sciences and Engineering Research Council of Canada (NSERC), an Early Researcher Award from the Province of Ontario (#ER14-10-258), funding from the Canada Foundation for Innovation and Ontario Research Fund (#37872) from the Canada Research Chairs (#950-232147) program to E.F.R., and an Alexander Graham Bell Canada Graduate Scholarship from NSERC to M.O.K.

## References


Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201–210.

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.

Bodner, G. E., Jamieson, R. K., Cormack, D. T., McDonald, L., & Bernstein, D. M. (2016). The production effect in recognition memory: Weakening strength can strengthen distinctiveness. *Canadian Journal of Experimental Psychology, 70*(2), 93–98.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language, 26*(3), 341–361.

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science, 27*(5), 302–308.

Fawcett, J. M. (2013). *The production effect benefits performance in between-subject designs*. A meta-analysis. *Acta Psychologica, 142*(1), 1–5.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40*(7), 1046–1055.

Jamieson, R. K., Mewhort, D. J. K., & Hockley, W. E. (2016). A computational account of the production effect: Still playing twenty questions with nature. *Canadian Journal of Experimental Psychology, 70*(2), 154–164.

Jamieson, R. K., & Spear, J. (2014). The offline production effect. *Canadian Journal of Experimental Psychology, 68*(1), 20–28.

Kelly, M. O., Ensor, T. M., Lu, X., MacLeod, C. M., & Risko, E. F. (2022). Reducing retrieval time modulates the production effect: Empirical evidence and computational accounts. *Journal of Memory and Language, 123*, 104299.

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in methods and practices*. *Psychological Science, 4*(1), 2515245920951503.

MacLeod, C. M., & Bodner, G. E. (2017). The production effect in memory. *Current Directions in Psychological Science, 26*(4), 390–395.

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*(3), 671–685.

Quinlan, C. K., & Taylor, T. L. (2013). Enhancing the production effect in memory. *Memory, 21*(8), 904–915.

Quinlan, C. K., & Taylor, T. L. (2019). Mechanisms underlying the production effect for singing. *Canadian Journal of Experimental Psychology, 73*(4), 254–264.

Roberts, B. R. T., MacLeod, C. M., & Fernandes, M. A. (2022). The enactment effect: An integrative review of behavioral, neuroimaging, and patient studies. *Psychological Bulletin, 148*(5-6), 397–434.

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language, 118*, 104219.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145–166.

Zhou, Y., & MacLeod, C. M. (2021). Production between and within: Distinctiveness and the relative magnitude of the production effect. *Memory, 29*(2), 168–179.