

Assessing the costs and benefits of production in recognition

Glen E. Bodner · Alexander Taikh ·
Jonathan M. Fawcett

© Psychonomic Society, Inc. 2013

Abstract The production effect is a memory advantage for items studied aloud over items studied silently. Although it typically is found within subjects, here we also obtained it between subjects in a recognition task—providing new evidence that production can be an effective study strategy. Our experiment, and a set of meta-analyses, also evaluated whether the within effect reflects costs to silent items and/or benefits to aloud items. Contrary to a strong distinctiveness account, we found little evidence that aloud items show an additional within-subjects benefit. Instead, silent items suffered an additional within-subjects cost. Blocking silent and aloud items eliminated this cost, suggesting that the cost was due to mixing silent and aloud items. Our discussion focuses on implications for distinctiveness and strength accounts of the production effect and on how to implement production as an encoding strategy depending on the learner's goals.

Keywords Production effect · Costs/benefits · Distinctiveness · Memory strength · Recognition memory · Meta-analyses

Does saying items aloud make them more memorable? Hopkins and Edwards (1972) were the first to address this question. They compared recognition of words pronounced aloud at study with that of words read silently in both within- and between-subjects conditions, and they found a significant pronunciation advantage only in the within condition. MacLeod, Gopie, Hourihan,

Neary, and Ozubko (2010) replicated this pattern and also established that the within effect is not specific to pronunciation; it occurs with other encoding tasks requiring distinct, item-specific responses (see also Forrin, Ozubko, & MacLeod, 2012). MacLeod et al. dubbed this class of phenomena the *production effect*, thus linking it to the well-established generation effect (i.e., enhanced memory for generated items over read items; e.g., Slamecka & Graf, 1978).

According to MacLeod et al. (2010), the production effect is more robust within than between subjects because the aloud items “are differentiated by being processed distinctively against the backdrop of the silently read (unpronounced) words” (p. 681). The memory traces for aloud items are encoded with extra detail that can be used heuristically at test to inform memory judgments (i.e., “I can recall saying this word aloud, so it must be old”). This *distinctiveness account* is currently the best-supported account of the production effect (see Bodner & Taikh, 2012; Forrin, Jonker, & MacLeod, *in press*; Ozubko, Major, & MacLeod, *in press*). A memory strength account, in contrast, predicts a memory advantage for the more strongly encoded (i.e., aloud) words in both within and between conditions. Importantly, a recent meta-analysis by Fawcett (2013) showed that the between effect is significant across studies, thus reopening the debate over the mechanism underlying the production effect.

Regardless of its cause, production may offer a simple and highly effective study strategy. However, its utility is predicated (at least in part) on the assumption that the within production effect reflects a *benefit* for aloud items, rather than a *cost* to silent items. Benefits occur if aloud items show better memory in a mixed list than in a pure aloud list. Costs occur if silent items show worse memory in a mixed list than in a pure silent list. MacLeod et al. (2010) suggested that “the production effect seems to be more an enhancement of the aloud items” (p. 681). Oppositely, Hopkins and Edwards (1972) concluded that “the effect of pronunciation appears to lie primarily in a decrement in performance for unpronounced

G. E. Bodner (✉) · A. Taikh
Department of Psychology, University of Calgary,
2500 University Drive NW,
Calgary, AB T2N 1N4, Canada
e-mail: bodner@ucalgary.ca

J. M. Fawcett
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road,
CB2 7EF Cambridge, UK
e-mail: jonathan.fawcett@mrc-cbu.cam.ac.uk

words rather than an increment for recognition memory of pronounced words” (p. 537). Our goal was to resolve this discrepancy through an experiment and a set of meta-analyses.

Begg and colleagues studied extensively the issue of whether the within-subjects generation effect reflects a true benefit for generated items versus a cost for read items (e.g., Begg & Roe, 1988; Begg & Snider, 1987; see also Slamecka & Katsaiti, 1987). According to Begg and Snider, generation establishes a criterion of having to identify words as independent entities. This criterion can result in cursory encoding of read words in a mixed list because identifying them as independent entities requires relatively little effort. Consistent with their account, comparisons of mixed- and pure-list conditions in these studies revealed that the within generation effect largely reflected costs rather than benefits. However, use of related word pairs and categorized lists in the within condition revealed a pure benefit to generation, without a concomitant cost (Begg, Snider, Foley, & Goddard, 1989). Thus, generation can reflect either costs or benefits, depending on aspects of the stimuli and method.

Experiment

Our experiment examined whether the within production effect in recognition is due to enhanced recognition of aloud items and/or impaired recognition of silent items and whether blocking the aloud and silent items at study modulates the cost/benefit pattern. To this end, benefits and costs in a *mixed group* were gauged relative to pure-list *silent* and *aloud groups*. To evaluate the possibility that mixing might impair memory for the silent items, we also tested two *blocked groups* (*silent–aloud*, *aloud–silent*). If a cost for silent items in the mixed group were due to lazy reading of the silent items, for example, then blocking should reduce the cost effect. The effect of blocking on costs has not been examined for either generation or production. However, Bertsch, Pesta, Wiscott, and McDaniel’s (2007) meta-analysis revealed that the within generation effect is larger when items are mixed (vs. blocked).

Method

University of Calgary undergraduates participated in one of five groups (48 per group): *mixed*, *silent–aloud blocked*, *aloud–silent blocked*, *silent*, or *aloud*. The stimuli were 100 nouns used in other production effect studies (e.g., MacLeod et al., 2010). They were assigned to four sets of 25 items. Two sets served as new items at test. In the between groups, two sets served as either aloud or silent items, and both sets were studied in either orange or green font. In the within groups, one set served as aloud items, and the other served as silent items; one set was studied in orange font, and the other in green font. Assignment of sets and the color of studied sets

were counterbalanced across participants. Items in the mixed group were randomly mixed.

Participants were tested individually via computer. They were informed that they would study words for an unspecified memory test. The mixed group was instructed to read the orange words aloud and the green words silently or vice versa. The blocked groups read an instruction screen before each study list informing them whether to read the upcoming list aloud or silently. The aloud group was instructed to read each word aloud, and the silent group was instructed to read each word silently. Items were shown one word at a time in a random order at a rate of 2 s per word, with a 0.5-s blank screen between each word, in 36-point Arial font. Participants then received a 100-item recognition test. The 50 studied words and 50 new words were presented one word at a time in a random order, in black 36-point Arial font. Participants pressed the left button (“old”) on a response box if they thought the word had been studied or the right button (“new”) if they thought it had not been studied.

Results

Table 1 lists the mean hit and false alarm rates and a signal detection measure of discrimination (d') for which floor false alarms and ceiling hit rates (a few per group) were adjusted using a $1/2N$ correction (Macmillan & Creelman, 1991). Results were significant at the .05 level unless otherwise indicated.

The mixed group showed a robust within production effect on hits favoring aloud items over silent items (.83 vs. .63), $F(1, 47) = 67.39$, $MSE = .01$. The blocked groups were analyzed using a 2 (list order: *silent–aloud* vs. *aloud–silent*) \times 2 (item type: *aloud* vs. *silent*) mixed-factorial ANOVA. There was a significant within production effect in the blocked

Table 1 Means (with *SEs*) for each group and item type

| Measure/Group | Aloud Items | Silent Items | New Items |
|-------------------------|-------------|--------------|---------------------|
| Hits and false alarms | | | |
| Mixed-list group | .83 (.02) | .63 (.03) | .17 (.02) |
| Silent–aloud group | .81 (.02) | .70 (.02) | .14 (.01) |
| Aloud–silent group | .80 (.02) | .70 (.03) | .15 (.02) |
| Between groups | .76 (.02) | .72 (.02) | .09 (.01)/.18 (.02) |
| Discrimination (d') | | | |
| Mixed-list group | 2.18 (0.12) | 1.44 (0.10) | |
| Silent–aloud group | 2.21 (0.12) | 1.84 (0.11) | |
| Aloud–silent group | 2.17 (0.12) | 1.79 (0.08) | |
| Between groups | 2.21 (0.09) | 1.72 (0.12) | |

Note. False alarm means for the between groups refer to the aloud and silent group, respectively

groups (.80 vs. .70), $F(1, 94) = 28.30$, $MSE = .02$, and no list order effect or interaction, $F_s < 1$. Thus, the production effect survived when silent and aloud items were blocked to reduce the potential for lazy reading of silent items. However, the production effect was smaller in the blocked design than in the mixed design on hits (.10 vs. .20), $F(1, 142) = 11.36$, $MSE = .02$, and in d' (.37 vs. .74), $F(1, 118) = 8.79$, $MSE = .24$; the basis of this reduction is followed up below. Finally, although the aloud and silent groups had similar hit rates (.76 vs. .72), $F(1, 94) = 1.78$, $MSE = .02$, $p = .19$, the aloud group's false alarm rate was half that of the silent group (.09 vs. .18), $F(1, 94) = 14.64$, $MSE = .01$. As a result, there was a robust between production effect in d' (2.21 vs. 1.72), $F(1, 94) = 10.79$, $MSE = .54$.

Turning to the cost/benefit analyses, discrimination of aloud items was not enhanced in the mixed group, relative to the aloud group (2.18 vs. 2.21), $F < 1$. On the other hand, discrimination of silent items was marginally worse in the mixed group than in the silent group (1.44 vs. 1.72), $F(1, 94) = 3.41$, $MSE = .55$, $p = .07$. Comparing the blocked-list and aloud group, there was no benefit for aloud items (2.19 vs. 2.21), $F < 1$, and comparing the blocked-list and silent group, there was no cost for silent items (1.82 vs. 1.72), $F < 1$. Finally, comparing the mixed- and blocked-list groups, discrimination of aloud items was not enhanced in the former group (2.18 vs. 2.19), $F < 1$, but discrimination of silent items was worse (1.44 vs. 1.82), $F(1, 94) = 10.15$, $MSE = .44$.

The between production effect makes it harder to show enhancement for aloud items in the within groups, relative to the aloud group. Therefore, we also gauged benefits by comparing aloud items in each *within* group with those in the *silent* group; we refer to these as *benefits-over-silent*. Doing so revealed superior discrimination for aloud items in the mixed group relative to the silent group (2.18 vs. 1.72), $F(1, 94) = 7.76$, $MSE = .02$, and in the blocked group relative to the silent group (2.19 vs. 1.72), $F(1, 142) = 10.86$, $MSE = .65$.

Next, we also evaluated whether the within groups showed a net increase in discrimination, relative to the silent group; we refer to these as *net benefits*. Overall d' (averaged across silent and aloud items) was nearly identical in the mixed group and silent group (1.73 vs. 1.72), $F < 1$, but was marginally higher in the blocked group than in the silent group (1.94 vs. 1.72), $F(1, 142) = 3.33$, $MSE = .47$, $p = .07$.

Finally, we also evaluated the possibility that the mixed group engaged in a processing trade-off in which they continued to process aloud items during silent item trials. This type of trade-off would work to reduce the positive correlation between aloud and silent discrimination one would typically expect to see. Contrary to this possibility, there was a significant positive correlation between d' for aloud and silent items in the mixed group, $r(46) = .62$, and critically, similar significant positive correlations were obtained

in the silent–aloud and aloud–silent groups, $r_s(46) = .52$ and $.56$, who would not be able to engage in this trade-off strategy (silent–aloud group) or would have less opportunity to engage in it (aloud–silent group).

Discussion

We obtained a robust between production effect in recognition.¹ The between-subjects production effect provides unequivocal evidence that production can enhance recognition. Although the between effect can be consistent with either a strength or a distinctiveness account (see Bodner & Taikh, 2012; Fawcett, 2013), we found no evidence that recognition of aloud items was enhanced more within subjects than between subjects, as should occur if making a subset of studied items distinctive via production renders them particularly memorable (MacLeod et al., 2010). Instead, discrimination was remarkably similar for aloud items across our mixed, blocked, and aloud groups (see Table 1). Nevertheless, discrimination was certainly higher for aloud items in our within groups than in our silent group (i.e., benefits-over-silent).

On the other hand, we found new evidence that recognition of silent items is impaired in a mixed design. The production effect was halved when items were blocked rather than mixed, and this reduction reflected poorer discrimination of silent items in the mixed group, relative to the blocked group and silent group (although the latter result was marginal). A mixed design could promote cursory reading on silent trials, or it might encourage postproduction monitoring on silent trials that follow aloud trials, either of which would inflate the overall production effect. Unfortunately, we did not collect study trial sequence information to examine such possibilities, so it remains for future research to consider. However, the positive correlation between aloud and silent discrimination was similar in our mixed and blocked groups. Had the mixed group focused on processing or rehearsing the aloud items during silent item trials, we would have expected a reduced correlation in the mixed group.

Averaged across aloud and silent items, discrimination was equivalent in the mixed group and silent group and was only marginally better in the blocked group than in the silent group. Thus, if increasing overall memory accuracy is one's goal, our data suggest that the best approach is to encode *all* items aloud (given the between production effect),

¹ We replicated the between production effect when items were presented on two separate lists at study as for the blocked groups ($n = 16$ per group). Aloud and silent group hit rates were .75 and .72, $F < 1$; false alarm rates were .11 and .21, $F(1,30) = 10.53$, $MSE = .01$; and d 's were 2.18 and 1.57, $F(1,30) = 4.92$, $MSE = 1.22$. We have since obtained several more replications that are part of a current project investigating the basis of the between production effect.

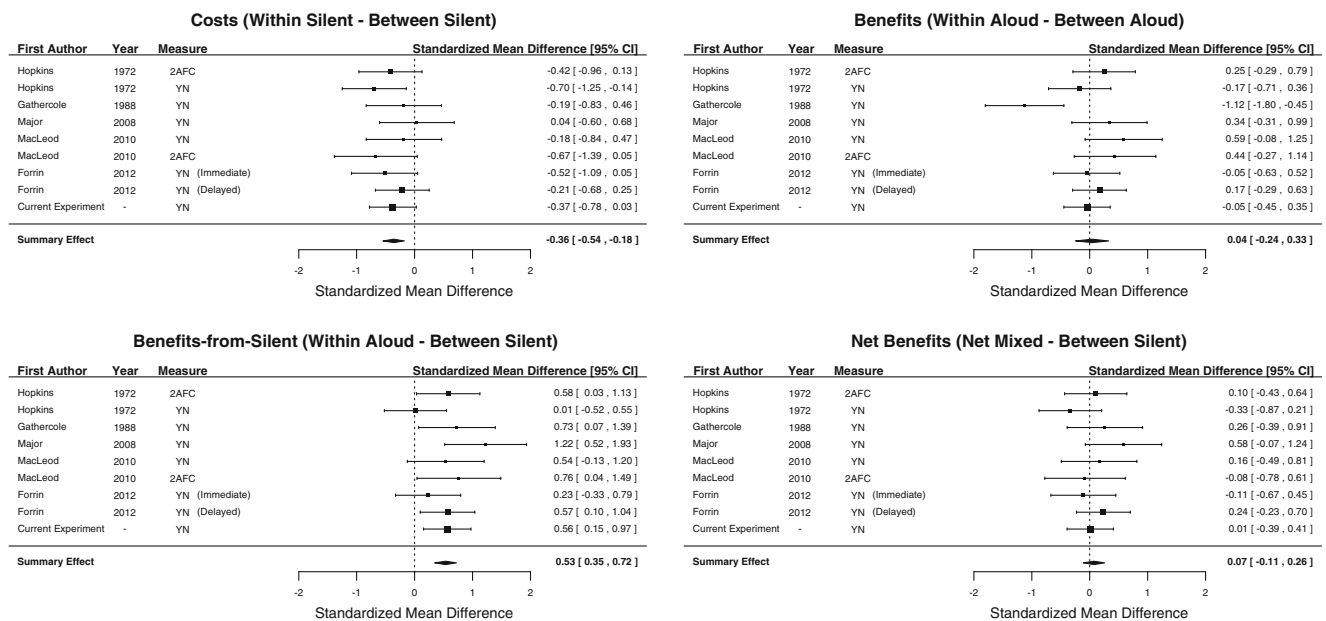


Fig. 1 Meta analyses of the costs and benefits of production in recognition. Effect sizes and confidence intervals are based on standardized mean differences in discrimination (d'). The polygon at the bottom of each panel represents the summary effect for each analysis calculated

rather than only a subset of them (given the lack of net increase in the within groups relative to the silent group). This new claim regarding the utility of production as a study strategy warrants further study.

Meta-analyses

MacLeod et al. (2010) suggested that their within production effect was largely due to benefits, whereas Hopkins and Edwards (1972) reported that theirs was primarily due to costs. Rather than relying on our experiment alone to adjudicate between these possibilities, particularly given our between production effect (which neither of these studies obtained), we next report a set of meta-analyses based on nine recognition studies in which production was manipulated both within subjects (in a mixed list) and between subjects under otherwise identical conditions: (1) Hopkins and Edwards (1972, Experiment 1), (2) Hopkins and Edwards (1972, Experiment 2), (3) Gathercole and Conway (1988, Experiment 5), (4) Major, Ozubko, and MacLeod (2008, unpublished), (5) MacLeod et al. (2010, Experiment 1 vs. 2), (6) MacLeod et al. (2010, Experiment 3), (7) Forrin and MacLeod (2012, immediate test condition, unpublished), (8) Forrin and MacLeod (2012, delayed test condition, unpublished), and (9) our experiment.

Discriminability (d') was calculated from raw participant means, except for Gathercole and Conway (1988) and Hopkins and Edwards (1972), where we used the published

using a random-effects model. The square marker size indicates weight within the model. 2AFC, two-alternative forced choice; YN, yes/no recognition

group means, and standard deviations imputed from the weighted average of the other studies. Effect sizes were separately calculated as standardized mean differences using the *escalc* function from the *metafor* package (Viechtbauer, 2010) in R (R Development Core Team, 2010). This function produces a metric corrected for inherent positive bias (see Hedges, 1982; Hedges & Olkin, 1985). Separate random-effects models were fit to each analysis using the *rma* function from the *metafor* package, producing aggregate effect sizes measured on the same scale (see Fig. 1).²

These meta-analyses revealed nonsignificant benefits ($g = 0.04$, $CI_{95\%} = -0.24, 0.33$) and net benefits ($g = 0.07$, $CI_{95\%} = -0.11, 0.26$) but significant costs ($g = -0.36$, $CI_{95\%} = -0.54, -0.18$) and benefits-over-silent ($g = 0.53$, $CI_{95\%} = 0.35, 0.72$). Heterogeneity was minimal for net benefits ($I^2 < 0.01\%$; $Q_{Error} = 6.06$, $p = .64$), costs ($I^2 < 0.01\%$; $Q_{Error} = 4.91$, $p = .76$), and benefits-over-silent ($I^2 < 0.01\%$; $Q_{Error} = 9.17$, $p = .32$), but was moderately high for benefits ($I^2 = 56.41\%$; $Q_{Error} = 17.73$, $p = .02$), due to Gathercole and Conway (1988) (see Fig. 1, top-right panel). Excluding this study eliminated this heterogeneity ($I^2 < 0.01\%$; $Q_{Error} = 5.53$, $p = .57$), and although it did not result in a significant benefit effect, it was certainly closer to significance ($g = 0.13$, $CI_{95\%} = -0.06, 0.32$).

² An additional random-effects model was fit, estimating the magnitude of the between production effect in these studies. In line with Fawcett (2013), there was a moderate effect ($g = 0.49$, $CI_{95\%} = 0.29, 0.69$) with minimal heterogeneity ($I^2 = 3.22\%$; $Q_{Error} = 12.74$, $p = .12$).

On the basis of our present evidence and measures, we conclude that the production effect in mixed lists yields costs, but whether it yields benefits depends on how benefits are conceived. Production clearly yields improved discrimination for aloud items in mixed groups, relative to silent groups (i.e., benefits-over-silent), but there is not yet good evidence of a benefit in a mixed-list design, relative to pure aloud groups.

General discussion

Our experiment and meta-analyses make several new contributions to our understanding of the production effect in recognition. First and foremost, we obtained a significant between-subjects production effect, an effect previously reported only once (Gathercole & Conway, 1988). This effect accords well with Fawcett's (2013) meta-analysis, but it adds new wrinkles to accounts of the production effect and to our cost/benefit analyses. The between production effect does not rule out a distinctiveness account, given that pure aloud groups can choose to attempt to recollect saying studied items aloud to inform their recognition decisions (for a discussion, see Bodner & Taikh, 2012; Fawcett, 2013; MacLeod et al., 2010). However, it does remove an important source of support for this account over a strength account, although there are other sources (see Ozubko, Gopie, & MacLeod, 2012; Ozubko et al., *in press*).

We did not find improved accuracy for aloud items in our within groups relative to our aloud group, which MacLeod et al.'s (2010) distinctiveness account predicts. Instead, the benefits of production (relative to a silent group) were equivalent in our within and between designs. In addition, although discrimination was better in the aloud group than in the silent group, discrimination was not better in the within groups averaged across aloud and silent items (i.e., net benefits) than in the silent group. Thus, the best strategy for maximizing discrimination may be to encode all items aloud, rather than encoding half aloud and half silently.

The production effect in a mixed design in recognition occurs in part because it includes a cost to silent items, as Hopkins and Edwards (1972) originally reported (cf. MacLeod et al., 2010). Production can impair memory for nonproduced items in a mixed list. This cost may reflect cursory processing of silent items. Participants might use the silent trials as an opportunity to prepare for the next aloud trial, rather than focusing on encoding the silent items. Consistent with this possibility, blocking the silent and aloud items eliminated the cost effect and yielded a production effect half as large as that obtained in the mixed group (.10 vs. .20). These findings lead us to advocate use of a blocked rather than a mixed design for measuring the benefits of production. Of course, the 10% production effect in the

blocked design is substantial and, indeed, comparable to the generation effect and other "distinctiveness effects" (Ozubko & MacLeod, 2010). If one's goal is to promote memory for one set of items and to reduce memory for (and interference from) another set of items, producing the former subset of items and silently reading the latter subset of items might provide a useful strategy. Alternatively, if one wishes to strengthen the produced items without incurring a cost for nonproduced items, blocking might be a useful means of achieving that goal. In other words, the best way to employ a production strategy (pure aloud vs. mixed list vs. blocked) depends on the learner's goals.

Although mixing clearly impairs recognition of silent items, there is no direct evidence that this cost reflects lazy reading (as was true of other studies; e.g., Begg & Roe, 1988). Blocking eliminated the cost to silent items, but how it did so remains unclear. On this issue, Forrin et al. (*in press*) found that requiring participants to generate or imagine items in a mixed list was additive rather than underadditive with the production effect, contrary to the lazy reading hypothesis. However, the influence of elaborative encoding on the costs versus benefits of production could not be gauged in this study because pure-list groups were not tested.

Future research should examine whether the pattern of cost/benefits for production is shared by other distinctive encoding strategies, such as generation, imagery, and levels-of-processing (McDaniel & Bugg, 2008). It will also be important to test whether the production effect in free recall (e.g., MacLeod, 2011) reflects both costs and benefits, as Slamecka and Katsaiti (1987) found for generation. Finally, other means of distinguishing between distinctiveness and strength accounts must be sought, given that the production effect can occur both within and between subjects.

Author Note The Natural Sciences and Engineering Research Council of Canada supported this research through a Discovery Grant to Glen Bodner. We thank Noah Forrin, Colin MacLeod, Jason Ozubko, and an anonymous reviewer for their feedback. Send correspondence to Glen Bodner, Department of Psychology, University of Calgary, 2500 University Drive NW, Calgary AB T2N 1N4, Canada (e-mail: bodner@ucalgary.ca).

References

(* indicates a study included in the meta-analysis)

- Begg, I., & Roe, H. (1988). On the inhibition of reading by generation. *Canadian Journal of Psychology*, 42, 325–336.
- Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 553–563.

- Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 977–989.
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 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.
- Fawcett, J. M. (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, 142, 1–5.
- Forrin, N. D., Jonker, T. R., & MacLeod, C. M. (in press). Production improves memory equivalently following elaborative vs. non-elaborative processing. *Memory*.
- *Forrin, N. D., & MacLeod, C. M. (2012). *The production effect in memory: Costs and benefits*. Manuscript in preparation.
- Forrin, N. D., Ozubko, J. D., & MacLeod, C. M. (2012). Widening the boundaries of the production effect. *Memory & Cognition*, 40, 1046–1055.
- Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory & Cognition*, 16, 110–119.
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92, 490–499.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York, NY: Academic Press.
- *Hopkins, R. H., & Edwards, R. E. (1972). Pronunciation effects in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 11, 534–537.
- MacLeod, C. M. (2011). I said, you said: The production effect gets personal. *Psychonomic Bulletin & Review*, 18, 1197–1202.
- *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.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- *Major, J. C., Ozubko, J. D., & MacLeod, C. M. (2008). [Reconsidering the status of the production effect in pure list designs]. *Unpublished raw data*.
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, 15, 237–255.
- Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition*, 40, 326–338.
- 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.
- Ozubko, J. D., Major, J., & MacLeod, C. M. (in press). Remembered study mode: Support for the distinctiveness account of the production effect. *Memory*.
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna: R Development Core Team.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592–604.
- Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory and Language*, 26, 589–607.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.