Acta Psychologica 142 (2013) 1–5

Contents lists available at SciVerse ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The production effect benefits performance in between-subject designs: A meta-analysis ☆

Jonathan M. Fawcett *

Dalhousie University, Canada

ARTICLE INFO

Article history:
Received 4 April 2012
Received in revised form 17 September 2012
Accepted 7 October 2012
Available online 9 November 2012

PsycINFO classification:
2300 Human Experimental Psychology
2340 Cognitive Processes
2343 Learning & Memory

Keywords:
Production effect
Meta-analysis
Between-subjects
Distinctiveness
Human memory

ABSTRACT

Producing (e.g., saying, mouthing) some items and silently reading others has been shown to result in a reliable advantage favoring retention of the produced compared to non-produced items at test. However, evidence has been mixed as to whether the benefits of production are limited to within- as opposed to between-subject designs. It has even been suggested that the within-subjects nature of the production effect may be one of its defining characteristics. Meta-analytic techniques were applied to evaluate this claim. Findings indicated a moderate effect of production on recognition memory when varied between-subjects ($g = 0.37$). This outcome suggests that the production effect is not defined as an exclusively within-subject occurrence.

1. Introduction

The notion that producing (compared to silently reading) a word could benefit memory has been around for at least four decades (e.g., Hopkins & Edwards, 1972). Since then, a great deal of research has supported this assertion using tasks ranging from speaking the word aloud to silently mouthing it (see MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010). Recently branded the production effect, the performance benefit in recognition memory for produced compared to non-produced words has been attributed to the concept of distinctiveness. That is to say that producing a word results in a production trace that can be reconstituted at test to discriminate (produced) study items from (non-produced) distractor items (e.g., Dodson & Schacter, 2001). Therefore, the distinctiveness account attributes the production effect to an interaction between the distinctive processes (i.e., production) applied at study and the retrieval strategies employed at test.

Curiously the production effect has been demonstrated almost exclusively using within- as opposed to between-subject designs. This has resulted in the presumption that the benefits afforded by production are evident only when tested in relation to other non-produced items (e.g., Ozubko & Macleod, 2010). Hourihan and Macleod (2008) further argued that the absence of a reliable between-subjects production effect provides compelling evidence against a single process account, such as one based purely on the strength of the study item in memory (for further discussion, see Ozubko & Macleod, 2010). If producing an item merely strengthened the associated memory trace (see, e.g., Wickelgren, 1969) performance should favor produced relative to non-produced items regardless of the study design. The finding that production benefits memory only relative to other non-produced items from the same session is instead most congruent with a distinctiveness account such as the one summarized above (Hourihan & Macleod, 2008).

The issue of whether the production effect is limited to within-subject designs was most recently addressed by MacLeod et al. (2010) in an article delineating the production effect and its boundary conditions. They summarized three published articles manipulating production between-subjects (Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972). Of those studies, only Gathercole and Conway (1988) reported a benefit of produced compared to silently read study items. MacLeod et al. (2010) then reported

☆ The author would like to thank Dr. John Christie, Dr. Vin LoLordo, Dr. Tracy Taylor, Dr. Glen Bodner, Chelsea Quinlan and Kate Thompson for their feedback, quantitative advice and encouragement. The author would also like to thank Dr. Colin MacLeod and Dr. Kathleen Hourihan for providing access to their raw data.

* Dalhousie University, Department of Psychology, Halifax, NS, Canada B3H 4J1. Tel.: +1 902 494 3001.
E-mail address: jmfawcet@dal.ca.

0001-6918/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2012.10.001

two experiments of their own in which they manipulated production between-subjects and then tested memory using either a yes–no (Experiment 2) or a two-alternative forced choice (Experiment 3B) recognition task. Neither experiment found a significant effect of production, resulting in the conclusion that the production effect is indeed a within-subjects phenomenon. They speculated that the absence of a between-subjects production effect is a defining characteristic of this paradigm (see also, Hourihan & Macleod, 2008; Ozubko & Macleod, 2010; Ozubko, Gopie, & MacLeod, 2011).

Another possibility is that production does have an impact when manipulated between-subjects — it is merely very small. This hypothesis is supported by an examination of the directionality of the null findings from the between-subject studies described above. The majority of these comparisons – despite being non-significant – were still in the predicted direction. That is to say that even though the respective p-values were often above .05, performance tended to favor the produced relative to the non-produced study items. This leaves us with a simple count of the significant and nonsignificant outcomes which disagrees with the apparent reliability of the pattern observed within those comparisons. The goal of the current article was to resolve this tension by providing a brief meta-analytic evaluation of the evidence for (or against) the production effect in between-subject designs.

2. Method

2.1. Literature search

A search was conducted of the online resources Google, Google Scholar, PsycINFO, PsychARTICLES, and JSTOR using numerous combinations and variations of the keywords: produce, say, speak, aloud, mouth, read, pronounce, memory, recognition, recall, and between-subjects. Only articles containing between-subject comparisons fitting the definition of the production effect provided above were considered for inclusion. This search was conducted until July 2011 but succeeded in locating only two articles (Hopkins, Boylan, & Lincoln, 1972; Ozubko & Macleod, 2010) in addition to the four referenced above (Dodson & Schacter, 2001; Gathercole & Conway, 1988; Hopkins & Edwards, 1972; MacLeod et al., 2010). Two further unpublished studies were procured through direct communication with the authors (Major & MacLeod, 2008; Ozubko & Macleod, 2009). Therefore, the sample consisted of eight studies contributing twelve independent effect sizes which are summarized as a forest plot in Fig. 1. Articles contributing one or more effect sizes are indicated in the reference section by an asterisk (*). Data were coded for measures of yes–no recognition, two-alternative forced choice and list-discrimination as the proportion correct responses for the target items. Notably, none of the between-subject studies identified throughout the search employed recall as a dependent measure. Therefore, whereas the analyses which follow are applicable to recognition performance, caution must be used when generalizing these findings to recall performance.

2.2. Effect size calculation and analysis

Effect sizes were calculated as the standardized mean difference between the production and control groups using the escalc function from the metafor package (Viechtbauer, 2010) within R version 2.12.1 (R Development Core Team, 2010). This function employs the procedure recommended by Hedges (1982) with a correction for positive bias (see Hedges & Olkin, 1985). In most cases the group variances were either calculated from the raw data (when available) or estimated from the reported statistics. In one instance (Gathercole & Conway, 1988) only the means were available. In this case the group variances were approximated by pooling the variances from all other studies that used the same dependent measure (yes–no recognition). Importantly, sensitivity analyses demonstrate that the effects reported below are robust across a range of imputed values for the group variances within this study — and that the same pattern is evident even if this study were excluded.

Effect sizes have been calculated such that a positive value represents greater performance for produced as opposed to non-produced items. Therefore, higher (positive) effect sizes represent a larger production effect. A random-effects model (using a restricted maximum-likelihood estimator) was then fitted to the aggregate data to estimate the overall impact of production on memory performance. This model was generated using the rma function from the metafor package (see Viechtbauer, 2010).

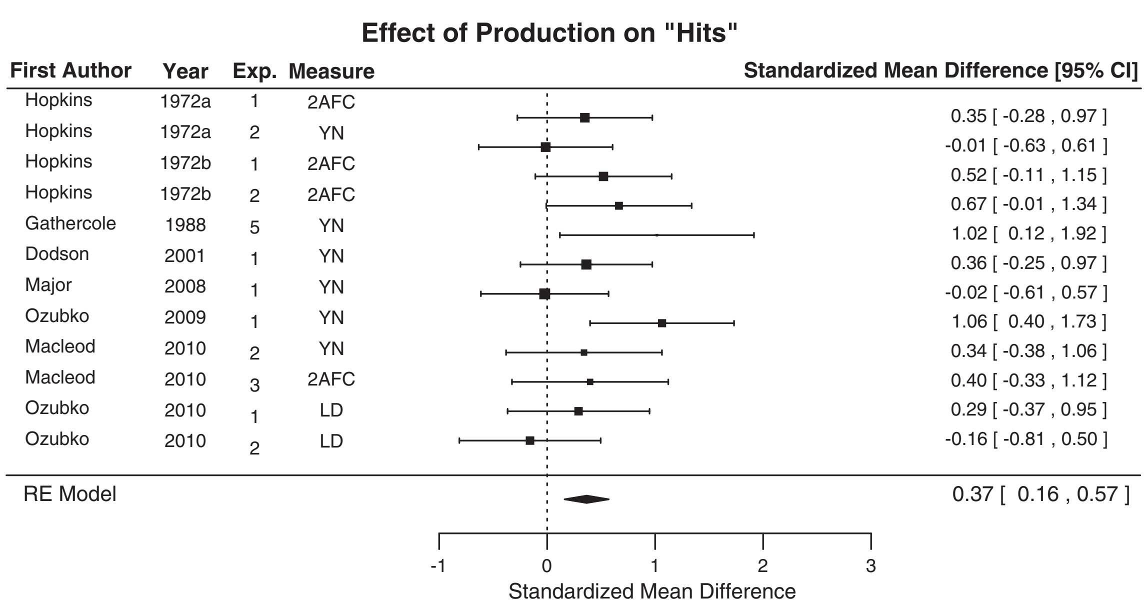

Fig. 1. Effect sizes and confidence intervals for the hits reported within each study. The polygon presented at the bottom represents the summary effect calculated using a random-effects model. Relative weight within the model is depicted by the size of the square representing the point estimate. 2AFC = two-alternative forced choice, YN = yes–no recognition, LD = list discrimination.