# Writing to Learn Increases Long-term Memory Consolidation: A Mental-chronometry and Computational-modeling Study of “Epistemic Writing”

Angélica M. Silva° and Roberto Limongi*

° Pontificia Universidad Católica de Valparaíso | Chile

* University of Western Ontario | Canada

Abstract: In this paper, we provide a mental-chronometry measurement (reaction time, RT) and a mathematical model to support the hypothesis that writing increases long-term memory (LTM) consolidation. Twenty-five subjects read short passages, wrote or spoke summaries of the texts, and performed a word-recognition episodic memory task. In the recognition task, participants responded faster in the written condition than in the spoken condition. We fit 15 drift-diffusion models to the accuracy and RT data to explore which components of the memory retrieval process reflect the learning effect of writing. Model selection methods showed that the nondecision parameter accounts for this effect, suggesting that initial stages of learning through writing are associated with fast episodic-memory retrieval. We suggest that the current approach could be used as a tool to compare different models of writing to learn. Furthermore, we show how combining mental chronometry, evidence-accumulation models of behavioral data, and dynamic causal models of functional magnetic resonance imaging could further the goal of understanding how writing affects learning. With a broader perspective, this approach provides a feasible experimental link between the field of writing to learn and the cognitive neurosciences.

Keywords: Epistemic writing; writing to learn; writing across the curriculum; writing in the disciplines; drift-diffusion model; writing research

journal of
WRITING RESEARCH

Silva, A.M., & Limongi, R. (2019). Writing to learn increases long-term memory consolidation: A mental-chronometry and computational-modeling study of “epistemic writing”. *Journal of Writing Research, 11*(1), 211-243. https://doi.org/10.17239/jowr-2019.11.01.07

Contact: Roberto Limongi, Robarts Research Institute, University of Western Ontario, 1151 Richmond Street N, London, ON N6A 5B7 | Canada – rlimongi@uwo.ca

## 1. Introduction

Many fields related to psychology, education, and linguistics accept that writing plays a prominent role in both academic achievement and professional success. In the 1970s, researchers in both scholarly and college writing began to consider written productions as an essential part of the learning process (Britton, 1980; Britton, Burgess, Martin, McLeod, & Rosen, 1975; Emig, 1977). These studies began with the Bullock report (Bullock, 1975) in Great Britain and were regarded as "a movement". This movement was first collectively referred to as "language across the curriculum" and later as "writing across the curriculum" (WAC) (Britton et al., 1975). The WAC movement expanded to the U.S and Canada (Thaiss & Porter, 2010), and later to Latin America (Navarro et al., 2016). As we detail below, all these works gave rise to more systematic inquiries about the effects of written productions on learning, suggesting that writing entails complex cognitive operations oriented by social and discursive goals (Bangert-Drowns, Hurley, & Wilkinson, 2004; Bazerman, 2018; Klein & Boscolo, 2016; Plane et al., 2017).

Since the 1980s, two perspectives about writing have been respectively studied by comparable research programs: WAC and writing in the disciplines (WID) (Bazerman et al., 2005; Carter, Ferzli, & Wiebe, 2007; Klein & Boscolo, 2016; McLeod, Miraglia, Soven, & Thaiss, 2001). Both the WAC and the WID movements have been largely influenced by cognitive (McCutchen, Teske, & Bankston, 2008) and rhetorical studies (Bazerman, 2018; Miller, 1984; Russell, 2002). Furthermore, both movements currently influence the curricula in disciplines such as psychology (Nevid, Ambrose, & Pyun, 2017), biology (Mynlieff, Manogaran, St. Maurice, & Eddinger, 2014), linguistics (Petrucci, 2002), history (Martínez, Mateos, Martín, & Rijlaarsdam, 2015), neuroscience (Prichard, 2005), pedagogy (Mateos, Martin, Villalon, & Luna, 2008), and second language acquisition (Al-Murtadha, 2013).

Writing helps students to both acquire disciplinary concepts (i.e., writing to learn) and socialize disciplinary knowledge via discipline-specific genre (i.e., learning to write) (A. Young, 2006). However, research in writing to learn differs from research in learning to write. Specifically, research in writing to learn focuses on the cognitive processes that makes writing itself a learning activity (Klein & Boscolo, 2016; McCutchen et al., 2008). Crucially, as a unique human characteristic, writing has an epistemic property because it serves learning, self-reflection, and knowledge acquisition (Brown, 1998). Interestingly, this property is closely related to the notion of epistemic action, embraced by current Bayesian frameworks of brain function such as predictive coding and active inference (Parr & Friston, 2017; Pezzulo, Cartoni, Rigoli, Pio-Lopez, & Friston, 2016). An epistemic action allows us to gain information, reduce uncertainty, and boost curiosity. Therefore, because epistemic is at the heart of writing as an action, we could refer to writing to learn as "epistemic writing".

It is uncontested that writing enhances learning. However, the questions of how it happens and how we could measure the learning effects from an experimental perspective remain open (Ackerman, 1993; Arnold et al., 2017; Klein, 1999). Furthermore, in the educational field there is an increasing interest in investigating how the knowledge about the neural basis of learning furthers the educational goal of improving teaching strategies. Naturally, knowing the neural mechanisms underlying the effects of writing on cognition would benefit the field of "neuroscience and education". However, this is an effort that demands rigorous conceptual and methodological links between education, psychology, and neuroscience (Bowers, 2016; Howard-Jones et al., 2016).

Based on the above and from the cognitive science perspective of levels of analysis proposed by Marr (1982), we have three agendas in the current work: substantiating conceptually and experimentally the thesis that the description of the effects of writing on learning should consider long-term memory (LTM) processes, advancing a formal model of the mechanism that underlies LTM consolidation during writing, and providing an experimental link between the field of writing to learn and the cognitive neurosciences. We start by describing the facts that make our case that mental-chronometry measurements (i.e., reaction time, RT) reveal the effects of writing on learning in terms of LTM consolidation. We then report the strategy we performed to evaluate the hypothesis that writing decreases the RT of LTM retrieval. We also explored, the specific component(s) of memory retrieval affected by writing. We pursued this by using evidence accumulation models of accuracy and RT. Finally, in the discussion section we establish the relationship between the proposed experimental paradigm, the current findings, and the cognitive neurosciences in terms of a link between the cognitive processes that subjects would deploy when they learn during writing activities and the underlying neural mechanisms.

## 2. Writing Facilitates LTM Consolidation

Klein (1999) proposes four hypotheses that could explain how writing affects learning: shaping at the point of utterance, genre, backward search, and forward search. The first hypothesis "shaping at the point of utterance" (Britton, 1980) states that writers produce texts by "writing down the speech" in terms of free associations of utterances. Specifically, both speakers and writers transform implicit knowledge into explicit knowledge during syntactic and semantic selection through propositional association. Put simply, this hypothesis assumes that retrieving one piece of information activates related concepts — which from a cognitive perspective is reminiscent to the spreading activation theory of semantic processing (Collins & Loftus, 1975).

The genre hypothesis states that writing enhances learning if writers follow an a-priori structure (i.e., genre) defined by specific relations among constituent parts. Consider a task in which students read a text and perform a writing activity. Following, they are asked to recall (c.f., LTM retrieval) as much information as possible about the source text (i.e., a free-association posttest). Essay-writing tasks lead to better recall of the source-text's content than non-essay writing activities such as note taking and question-answering tasks (Langer & Applebee, 1987).

The backward search hypothesis (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980, 1981) says that rhetorical constraints drive writing. Writers set goals, plan, write a content, and revise both goals and content. It is especially relevant the dual-problem space model proposed by Bereiter and Scardamalia (1987). In this model, writing leads to learning so long as the writer - in pursuing rhetorical goals - elaborates on retrieved information from LTM. Specifically, this model claims that learning, regarded as discovery, is a consequence of an interaction between a content space (the writer's beliefs stored in LTM) and a rhetorical space (the writer's representation of actual or intended text in terms of its rhetorical function). In the dual-space model novice writers tend to just retrieve information from the content space and "translate" it into text, i.e. knowledge-telling. More expert writers search and evaluate potential content in terms of rhetorical goals (represented in a rhetorical space), and knowledge transforming is a consequence of retrieving different material from LTM. At present, most of the writing-to-learn research invokes the dual-problem space model in motivating the writing-to-learn activities that they use (Klein, Haug, & Arcon, 2017) and to propose new models (Baaijen & Galbraith, 2018).

Finally, the forward search hypothesis says that writers revise texts iteratively aiming to find and resolve contradictions via inferential processes (R. Young & Sullivan, 1984). An exemplar case of this hypothesis is the proposal of discovery through writing described by the dual-process model (DPM) (Galbraith, 2009; Galbraith & Baaijen, 2015; Galbraith & Torrance, 1999). The DPM regards learning through writing as the effect of two processes on ideas activated in episodic LTM and ideas activated in semantic LTM. In one process, a dispositionally-guided process, writers produce texts by activating semantic content in LTM. It involves an initial synthesis within a distributed (i.e., connectionist) representation of content activated by the writing task specifications, followed by feedback from this initial output to the representation, leading to further syntheses. In another process, a knowledge-transforming process, writers operate over ideas stored in episodic memory. These ideas could be either those generated in the dispositionally-guided process or ideas already stored in episodic memory (e.g., ideas recently read in a text).

The analysis proposed by Klein (1999) remarks that the four hypotheses about how writing would affect learning relate to each other. For example, the forward hypothesis seems to be the natural evolution of the "shaping at the point of utterance" hypothesis. Similarly, from our point of view, we regard the dual-space model as a special case of the more general DPM.

We find the relationship between models of writing to learn not surprising for two reasons. First, as we detail below, in terms of Marr's levels of analysis (Marr, 1982) all of the models partially (the DPM) or totally fall in the algorithmic level. Second, at this level, all of the models rely directly or indirectly on a classical cognitive architecture with production systems operating on symbols retrieved from LTM. From this perspective, learning is strongly associated with LTM consolidation of novel categories or schemas (De Brigard, Brady, Ruzic, & Schacter, 2017). In what follows, we elaborate on these two points. Furthermore, within the specific case of the DPM, we reveal a gap in the field (lack of mental-chronometry measurements) and argue that filling this gap is necessary for the identification of the underlying *measurable* mechanism of the effect of writing on learning.

## 3. Models of writing to learn and Marr's levels of analysis.

Marr (1982) proposed that a cognitive system could be studied at three levels: computational (what the system's goal is –for example, creating/acquiring new knowledge through writing), algorithmic (the set and order of operations the system deploys on representations so that to achieve the computational goals — e.g., retrieving representations from LTM, transforming those representations in working memory, and creating the motor plan to be deployed during writing), and implementational (the physical realization of the operations –e.g., interaction between brain regions or the flow of information between the CPU and the hard disk of a computer). At present, Marr's levels of analysis prevail over other approaches in the cognitive-psychology literature (Peebles & Cooper, 2015).

At the algorithmic level of analysis, the cognitive-psychology field agrees that subjects achieve LTM consolidation[1] of new representations via two (non-orthogonal) operations: rehearsal and elaborative rehearsal (Tulving & Craik, 2000). Based on this assumption, we argue that current models of writing-to-learn fall within the algorithmic level because they propose cognitive processes that comprise rehearsal or elaborative rehearsal of retrieved information from LTM. For example, in the dual-space model the evaluation of information in the context of rhetorical goals demands active maintenance of both the retrieved information and the rhetorical goals in working memory, this can be achieved only via rehearsal operations. Similarly, knowledge transforming is a special case of elaborative rehearsal of the retrieved content.

A more recent example of the "algorithmic core" of current models of writing to learn is the DPM which assumes that new ideas or discoveries (c.f., learning) emerge from transient changes in neuron-like units via "mutual constraint satisfaction". Crucially, Baaijen and Galbraith (2018) regard mutual constraint as a different (i.e., sub-symbolic) mechanism to access information in LTM. However, based on a different read of the literature, we think that this mechanism is indeed the implementation of the retrieval processes described in symbolic frameworks (O'Reilly, Munakata, Frank, Hazy, & Contributors, 2012). As we argue below, the neural substantiation of the DPM speaks to the implementation of the two main algorithmic processes (i.e., knowledge-transforming and dispositionally-guided processes).

### 3.1 Linking the algorithmic and implementational levels of analysis

The initial proposal of levels of analysis in cognitive science (Marr, 1982) gave rise to *self-contained* algorithmic models of cognition and, by extension, models of writing to learn (e.g., the dual-space model). However, there is an increasing consensus in cognitive science that models of cognition and learning should also be explained at the neural level (Peebles & Cooper, 2015). In Marrian terms, this implies linking the algorithmic and implementational levels of analysis. In the field of writing to learn, at present, only the DPM (Baaijen & Galbraith, 2018; Galbraith, 1992, 2009; Galbraith & Baaijen, 2015; Galbraith & Torrance, 1999) suggests such a link by relying on the assumptions of the complementary learning system (CLS) (O'Reilly, Bhattacharyya, Howard, & Ketz, 2014; Schapiro, Turk-Browne, Botvinick, & Norman, 2017) —which is one of the most accepted learning models in the computational cognitive neuroscience literature.

In the CLS, learning new knowledge or LTM consolidation takes place in two different yet complementary ways which in the current version of the CLS (Schapiro et al., 2017) are assumed to be at the implementational level of analysis. First, single experienced events or episodes are quickly encoded in independent representations. In this phase, contextual (i.e., episodic) information strongly modulates changes in synaptic weights (i.e., connections) between neurons of the hippocampus. These implementational operations would take place during the knowledge-transforming process. Second, overlapping representations of single episodes give rise to abstract (i.e., semantic) patterns. This second phase demands many independent episodes and is reflected in changes of synaptic weights between cortical neurons. These second set of implementational operations would underlie the dispositionally-guided text production process.

To summarize, processes described in current models of writing to learn fall within the algorithmic level of analysis. A clear exception is the DPM in which dispositionally-guided text production and problem-solving are special cases (algorithms) of elaborative rehearsal that are thought to be neurally (sub-symbolically) implemented via the CLS. In what follows, we provide the conceptual and experimental support to the thesis that "mental chronometry" measurements (i.e., RT) are necessary to identify a mechanistic explanation of the effect of writing on learning at the algorithmic level of analysis.

### 3.2 The mental chronometry of the effect of writing on LTM consolidation

As we suggested above, writing-to-learn models speak to LTM consolidation as an important effect of writing on learning. Clearly, research in writing to learn relies on response accuracy (e.g., how well subjects recall the content of a source text after performing a writing activity) as an index of LTM consolidation. However, it is surprising that whereas in cognitive psychology and cognitive neuroscience (including the CLS model) the mental operations that give rise to LTM consolidation have been studied using not only response accuracy but also RT, the writing-to-learn field has not provided this mental-chronometry measurement as evidence of the effects of writing upon learning. This is especially relevant in the case of the DPM. Achieving a CLS-grounded DPM of writing to learn requires identifying a *quantitative* relationship between behavioral and neurophysiological measurements. Whereas the neurophysiological measurements should capture changes in the connectivity strength between neurons of the hippocampus and the neocortex, the RT measurement should capture the differential effects of dispositionally-guided and problem-solving processes on LTM consolidation. More in general, none of the current research trends in the field acknowledge the importance and necessity of this measurement. This is evident not only in the writing-to-learn literature (Klein & Boscolo, 2016) but also in the general research areas of academic and non-academic writing (Bazerman, 2018; Plane et al., 2017). We think that this is an important gap in the field. Filling this gap could further the goal of both resolving how writing leads to learning (i.e., comparing competing algorithmic models of writing to learn) and, more importantly, identifying the writing-to-learn mechanism at the implementational level of analysis —the neural level.

### 3.3 Measuring LTM consolidation

One way to measure LTM consolidation is computing how fast the cognitive systems retrieves stored information. This is frequently regarded as the latency from a stimulus presentation (e.g., a word in a text) to the activation of its schematic representation (e.g., the word's meaning) in LTM. In the laboratory, this latency has been used to study the relationship between the speed with which readers recognize a word —referred to as lexical access— word-to-text integration, and the ensuing text comprehension (Perfetti, 1985, 2007; Perfetti & Lesley, 2002; Perfetti & Stafura, 2014; Stafura & Perfetti, 2014; Taylor & Perfetti, 2016). Based on these facts, if writing features learning as a hallmark then it should decrease word-recognition time as an index of LTM consolidation. Furthermore, this effect should be observed at the level of the stages comprised in the word-recognition (retrieval) process. In the remainder of this introduction, we sketch the general strategy we used to test our hypothesis and to identify the affected stage of the retrieval process.

## 4. Experimental approach

We implemented a simple experimental paradigm in which participants read short passages, wrote or spoke a summary with the main idea of the passage, and performed an episodic-memory word recognition task. We decided to use spoken productions as a baseline condition because comparing production modalities could reveal subtle writing-specific effects (Cohen-Goldberg, 2017). Specifically, we needed to control for the effect of "language production". This is because language itself mediates learning (Langacker, 2008). In concrete, we hypothesize that writing should decrease the RT in the recognition task (compared with the effect of spoken productions) as an index of LTM consolidation.

Neuropsychological, computational, and animal-learning data show that episodic cues facilitate recognition processes (Craik & Tulving, 1975) (Lepage, Habib, & Tulving, 1998; Tulving, 1972, 1983). In addition, recent findings suggest that episodic memory facilitates the reactivation of the situation model associated with a written text (Johansson, Oren, & Holmqvist, 2018), which speaks to episodic memory as an index of text comprehension and learning. In neurophysiological terms, episodic memory influences the (cortical) consolidation of semantic memory, via activity of the hippocampus (Ketz, Morkonda, & O'Reilly, 2013; O'Reilly et al., 2014; O'Reilly & Norman, 2002; Schapiro et al., 2017). Therefore, in the context of a writing-to-learn activity we suggest that an initial behavioral effect of writing on word-recognition processes takes place at the level of episodic memory, specifically in the ability to discriminate whether a word has been recently read.

To explore and identify the specific stage(s) of the retrieval process that is affected by learning through writing, we fit drift-diffusion models to data collected in the episodic-memory word recognition task. Recently, lexical access has been regarded as an evidence-accumulation process (Anders, Riès, van Maanen, & Alario, 2015) - a family of decision-making models that fall at the algorithmic level of analysis. One of the most influential evidence-accumulation models of lexical access is the Ratcliff's drift-diffusion model (Ratcliff, 1978; Ratcliff, Gomez, & McKoon, 2004) which pertains to a large family of models of fast (i.e., at a millisecond scale) two-alternative forced choice (2AFC) decision-making tasks (Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff, Smith, Brown, & McKoon, 2016).

Understanding the dynamics of the drift-diffusion model is straightforward. Consider the RT distribution collected from a subject that performed an episodic-memory task in which he or she responded as quickly and as accurately as possible whether a test word appeared in a previously read passage. In this model, the subject accumulates information until reaching a threshold. After reaching this threshold, the subject executes a response.

Formally, the model comprises four basic parameters (Figure 1) representing the accumulation threshold (a), the starting point of the accumulation process (z), the accumulation (or drift) rate (v), and the nondecision processes (t) such as stimulus encoding and motor execution (i.e., sensorimotor delay). Parameters representing the intertrial variability of the drift rate (sv), starting point (sz), and nondecision processes (st) are also included. In the field of language processing, previous works have reported the effects of experimental manipulations and populations on these parameters. For example, the word-frequency effect on lexical access is captured by the drift rate (Ratcliff, Gomez, et al., 2004) whereas the effect of age on episodic LTM retrieval is capture by the nondecision parameter (Spaniol, Madden, & Voss, 2006). In the current work, we expect that the effect of writing on LTM consolidation will be captured by one or more parameters of the drift-diffusion model.

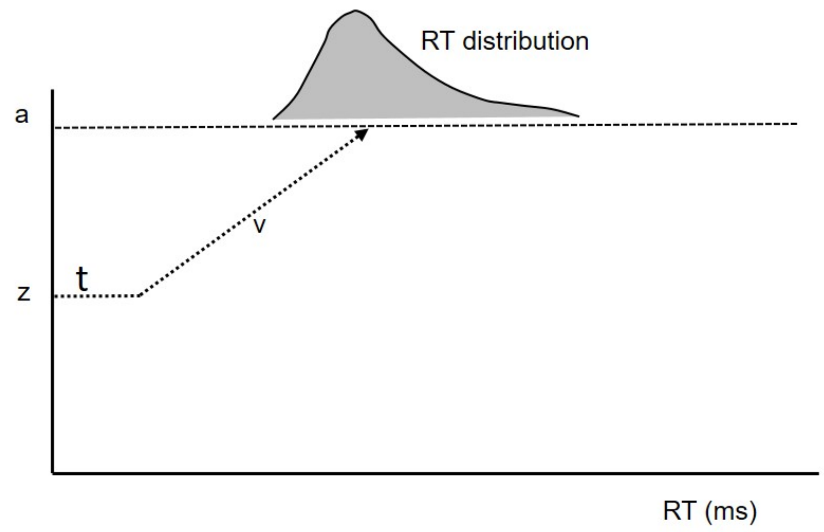


*Figure 1.* Basic Formulation of an Evidence-accumulation Model. The basic model comprises four parameters: decision threshold (a), starting point (z), drift rate (v), and nondecision time (t).

## 5. Materials and Methods

### 5.1 Participants

In this study, 25 university students (22 females, *M* age = 20.59, *SD* = 2.37) from Pontifical Catholic University of Valparaíso were recruited via campus and class advertisements. They voluntarily participated in the experiment, signed informed consent forms, and received course credit as compensation. The study was approved by the Institutional Review Board of the Pedagogical Institute of Caracas as part of a larger project (Silva, 2016). Participants spoke Spanish as L1.

## 5.2 Procedure

Participants began the experiment by putting on a head set and reading the general instructions on a computer screen. They performed one familiarization block followed by eight experimental blocks with two conditions (spoken and written, Figure 2). Each condition comprised four blocks, and each block included reading, production, and memory tasks —executed in this order. In the reading task, participants read and summarized (mentally and silently) short passages ($M_{words}$ = 203, SD = 37.67; $M_{sentences}$ = 9.37, SD = 3.29). In the production task, they wrote or spoke a summary of the passage. Participants were informed about the linguistic production modality just at the beginning of the production task. In the memory task, participants performed an episodic-memory recognition task with text and no-text words. They had to decide as

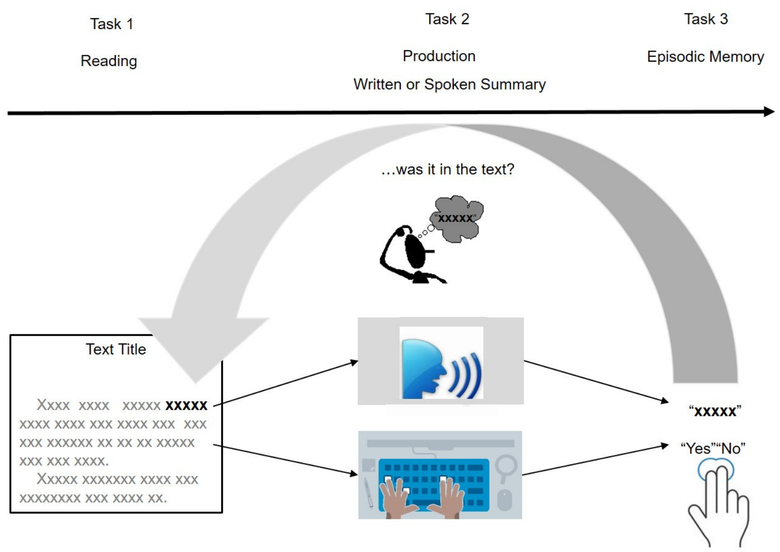


*Figure 2.* Experimental Paradigm.

quickly and as accurately as possible whether or not the test word appeared in the passage. In total, there were 60 text and 60 no-text words per condition. After finishing the memory task, a new block began. The stimulus delivery program was E-prime 2.0 (Schneider, Eschman, & Zuccolotto, 2012). Below, we describe each task in detail.

### Reading task

The reading task began with a text window depicting a passage that included a title at the top of it. The whole passage was visible on the screen during 2 min. On the top right corner of the screen, a timer showed the remaining time. Each passage was selected randomly (without replacement) from a pool of eight texts. The rhetorical structure of four texts was argumentative and with predominant hypotactic organization whereas the rhetorical structure of the other four texts was expositive and with predominant paratactic organization. We provide examples of each type of text in appendices A and B. A given passage could appear in any modality across participants. During this task, participants were not allowed to take notes. After the 2-minute reading time, the text window disappeared.

### Production task

The production task began with an instruction window. Depending upon the linguistic production condition, participants heard or read that they had to pronounce or write a summary with the main idea of the passage. In the spoken condition, participants heard an instruction asking to pronounce the summary of the passage after hearing a tone. During the duration of the instruction window, a cartoon depicting a woman wearing a head set was displayed. In the written condition, participants read an instruction on the center of the screen asking to type the summary when a textbox appeared. The instruction window lasted 11 s. When this time elapsed, the instruction (or cartoon) disappeared and a production window appeared.

The production window lasted 90 s, a timer on the top right corner of the screen showed the remaining time. In the spoken condition, a cartoon depicting a microphone appeared on the center of the screen with the following instruction below "ahora pronuncia la idea central del texto" (now say the main idea of the passage). In the written condition, a textbox appeared on the center of the screen with the following instruction below "ahora escribe la idea central del texto" (now write the main idea of the passage). For a summary to be considered valid, participants had to produce at least two sentences. Grammar and linguistic style were not considered as a correct-response criterion. We informed participants on these conditions during the general instructions and the familiarization block.

### Memory task

In the memory task, we instructed participants to respond as quickly and as accurate as possible if the test word appeared in the passage they read and summarized. On the computer keyboard, participants pressed the "k" key for "yes" responses and the "l" key for "no" responses. A word was shown in the middle of the screen and remained visible until a response was detected or until 2500 ms had elapsed. In a single block, the task comprised 15 texts and 15 no-text words. Text words were noun words extracted from the passage and were passage specific. This is, they appeared only one time in the experiment. No-text words were chosen randomly (without replacement across the experiment) from a pool of 120 words selected from Current Spanish Reference Corpus (Real Academia Española). A new test word appeared immediately after the termination of a trial. We provide examples of the source text, text and no-text words in appendices A and B.

## 5.3 Data Analysis

We performed two analyses of the memory-task data using Bayesian statistics. Bayesian statistics is now a common method in psychological sciences (Appelbaum et al., 2018) and represents a response to the call for a "new statistics" in terms of a shift from the frequentist null hypothesis significance testing to estimation of magnitudes with uncertainties (Kruschke & Liddell, 2017a, 2017b). Therefore, we performed estimations with quantified uncertainties of the parameters of interest.

In the first analysis, we computed the posterior distribution over mean RT differences of correct and incorrect responses between conditions. For completeness, we also computed mean RT differences between correct and incorrect responses within conditions. The mu ($\mu$) parameter represented the estimated mean difference. We analyzed correct and incorrect responses separately. Participants' responses were coded as "correct" or "incorrect" depending upon whether they matched the expected responses. This is, "yes" responses were expected for text words whereas "no" responses were expected for no-text words.

Prior distributions were informative. To estimate posterior distributions of parameters we used Markov chain Monte Carlo (MCMC) methods. The chain length (i.e., the MCMC sample size) was 100,000. The number of burn-in iterations was 1000, and the chains were generated with no thinning. We assessed convergence by computing the R-hat ($\hat{R}$) statistic (Gelman & Rubin, 1992). We report the estimate and the posterior 95 % highest density interval (HDI) which indicates the most probable parameter value given the data. As a decision rule, we accepted a difference in mean RT if the 95 % HDI comprising the mean of the most credible values for $\mu$ fell below 0. The analysis was implemented in R and JAGS (Kruschke, 2013a).

In the second analysis, we fit 15 Bayesian hierarchical drift-diffusion models (HDDM) to the distributions of RT and accuracy data (Wiecki, Sofer, & Frank, 2013). Hierarchical Bayesian methods allow to estimate both subjects and group parameters simultaneously. We explored all possible combinations of "a", "z", "v", and "t" as free parameters (Table 1). Furthermore, intertrial variability of z, v, and t ("sz", "sv", and "st") was estimated in every model. Prior distributions were informative. The chain length of each model was 5,000. The number of burn-in iterations was 250, and the chains were generated with no thinning.

*Table 1.* Model Space and Results of the HDDMs

| Model | Parameters | | DIC | $\hat{R}$ |
|---|---|---|---|---|
| | Free | Fixed | | |
| M1 | a | v, z, t | 5036.68 | — |
| M2 | v | a, z, t | 5037.41 | — |
| M3 | z | a, v, t | 5035.99 | — |
| M4 | t | a, v, z | 5027.61 | 1 |
| M5 | a, v | z, t | 5037.84 | — |
| M6 | a, z | v, t | 5037.20 | — |
| M7 | a, t | v, z | 5028.84 | — |
| M8 | v, z | a, t | 5038.77 | — |
| M9 | v, t | a, z | 5031.14 | — |
| M10 | z, t | a, v | 5029.41 | — |
| M11 | a, v, z | t | 5038.40 | — |
| M12 | a, v, t | z | 5030.45 | — |
| M13 | v, z, t | a | 5032.08 | — |
| M14 | a, z, t | v | 5031.76 | — |
| M15 | a, v, z, t | — | 5033.66 | — |

Note. $\hat{R}$ statistic was computed only for the winning model. DIC (deviance information criterion).

To select the winning model, we used the deviance information criterion (DIC) number as an approximation to Bayesian model evidence. At a group level, we report the parameter estimates of the winning model. For every parameter of interest, we report the percentage of the posterior estimates (i.e., the posterior proportion, PP) that differs from 0. We also report subject-wise parameter estimates. We assessed convergence by computing the $\hat{R}$ statistic. Goodness of fit was assessed via posterior predictive checks (Kruschke, 2013b).

## 6. Results

Trials with either RT > 2000 ms (1.3 %) or RT < 200 ms (1.1 %) were excluded from the analyses. Table 2 shows the relevant descriptive statistics.

### 6.1 Differences in Accuracy and RT

The first analysis revealed that response accuracy did not differ across conditions, $\mu$ = - 0.006, PP = .6 (Figure 3). However, on correct trials participants responded faster in the written condition than in the spoken condition, $\mu$ = - 0.02, PP = .951.

RTs of incorrect responses did not differ across conditions, $\mu$ = 0.008, PP = .333. However, correct responses were slower than incorrect responses in both the spoken $\mu$ = 0.119, PP = 1 and the written $\mu$ = 0.132, PP = 1 conditions (Figure 4).

*Table 2.* Summary Statistics of Variables of Interest

| Condition | Response | Dependent Variable | Mean | SD |
|---|---|---|---|---|
| Spoken | | Accuracy (%) | 81.4 | 4.16 |
| | Correct | RT (ms) | 870 | 292 |
| | Incorrect | | 971 | 333 |
| Written | | Accuracy (%) | 80.5 | 7.9 |
| | Correct | RT (ms) | 851 | 279 |
| | Incorrect | | 947 | 340 |

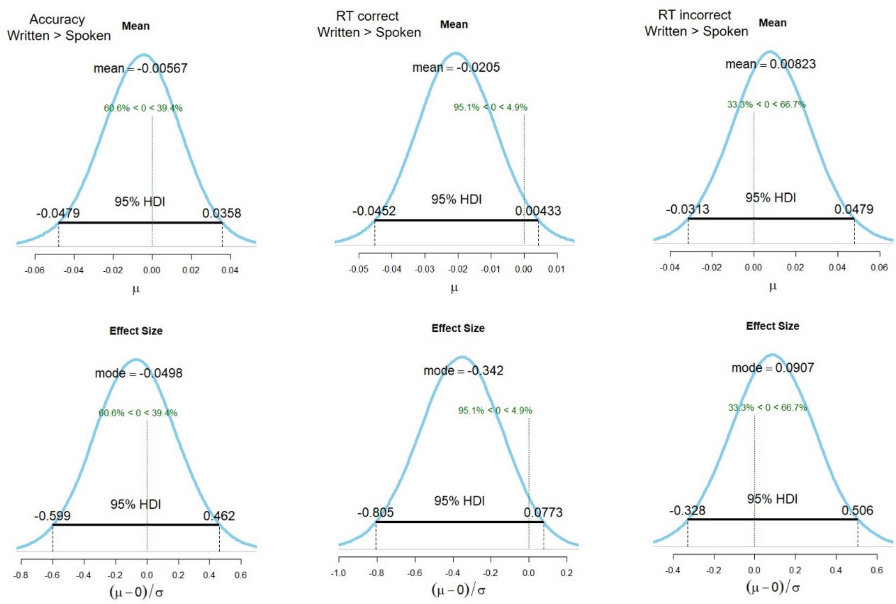


*Figure 3.* Posterior Distributions of Parameter Estimates about Differences in RT between Conditions. Correct responses were faster in the written condition whereas no difference was detected in incorrect responses. $\hat{R}$ = 1 in all estimations.

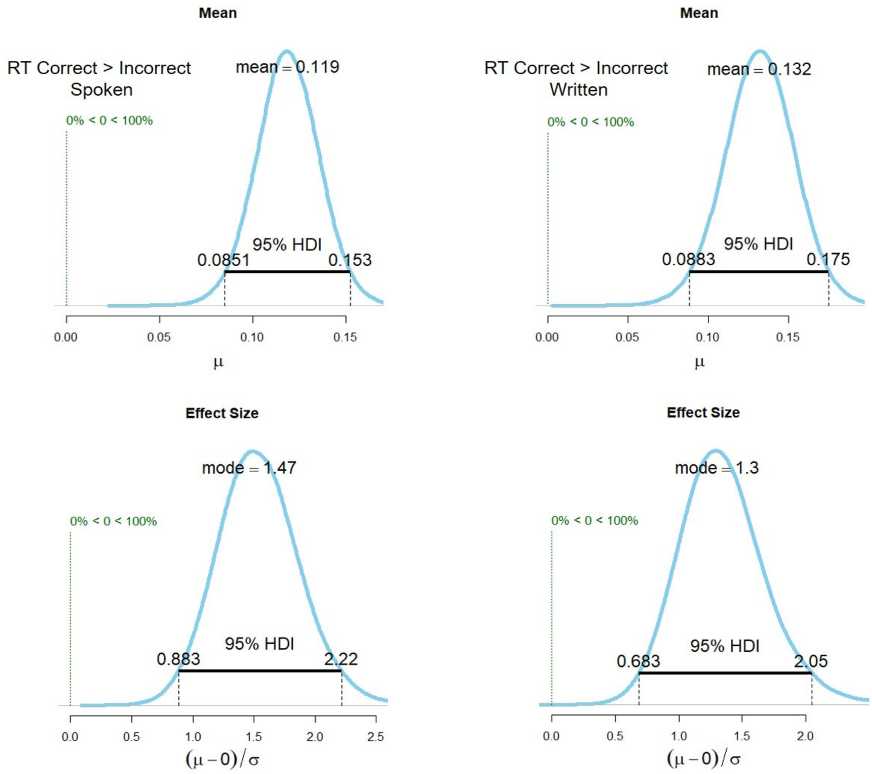


Figure 4. Posterior Distributions of Parameter Estimates about Differences in RT within Conditions. Incorrect responses were slower than correct responses in both conditions. $\hat{R} = 1$ in all estimations.

*Table 3.* Parameter Estimates of the Winning HDDM (M4)

| Parameter | a | | z | | v | | t(spoken) | | t(written) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | mean | std | mean | std | mean | std | mean | std | mean | std |
| Group | 1.489 | 0.075 | 0.495 | 0.014 | 1.532 | 0.105 | 0.511 | 0.022 | 0.496 | 0.004 |
| Subject 1 | 1.586 | 0.066 | 0.497 | 0.027 | 1.172 | 0.157 | 0.518 | 0.011 | 0.502 | 0.015 |
| Subject 2 | 1.559 | 0.066 | 0.360 | 0.027 | 1.700 | 0.170 | 0.383 | 0.012 | 0.368 | 0.016 |
| Subject 3 | 1.407 | 0.070 | 0.523 | 0.027 | 1.589 | 0.184 | 0.534 | 0.017 | 0.519 | 0.021 |
| Subject 4 | 1.214 | 0.055 | 0.527 | 0.027 | 1.148 | 0.186 | 0.465 | 0.014 | 0.450 | 0.018 |
| Subject 5 | 1.613 | 0.079 | 0.493 | 0.027 | 1.571 | 0.165 | 0.543 | 0.016 | 0.528 | 0.020 |
| Subject 6 | 1.700 | 0.085 | 0.522 | 0.027 | 1.655 | 0.180 | 0.591 | 0.016 | 0.575 | 0.020 |
| Subject 7 | 1.645 | 0.074 | 0.495 | 0.027 | 1.450 | 0.163 | 0.499 | 0.011 | 0.484 | 0.015 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Subject 8 | 1.474 | 0.080 | 0.570 | 0.027 | 1.589 | 0.193 | 0.663 | 0.019 | 0.648 | 0.023 |
| Subject 9 | 1.792 | 0.093 | 0.491 | 0.027 | 1.721 | 0.176 | 0.475 | 0.016 | 0.460 | 0.020 |
| Subject 10 | 1.011 | 0.051 | 0.554 | 0.027 | 1.684 | 0.218 | 0.647 | 0.013 | 0.631 | 0.017 |
| Subject 11 | 1.491 | 0.072 | 0.500 | 0.027 | 1.443 | 0.170 | 0.580 | 0.017 | 0.564 | 0.021 |
| Subject 12 | 0.956 | 0.047 | 0.527 | 0.027 | 1.643 | 0.210 | 0.525 | 0.012 | 0.509 | 0.016 |
| Subject 13 | 1.502 | 0.064 | 0.465 | 0.027 | 1.119 | 0.162 | 0.397 | 0.014 | 0.381 | 0.018 |
| Subject 14 | 1.528 | 0.073 | 0.513 | 0.027 | 1.275 | 0.168 | 0.564 | 0.019 | 0.548 | 0.023 |
| Subject 15 | 1.334 | 0.064 | 0.498 | 0.027 | 1.882 | 0.191 | 0.526 | 0.012 | 0.511 | 0.017 |
| Subject 16 | 1.234 | 0.060 | 0.539 | 0.027 | 1.510 | 0.186 | 0.561 | 0.015 | 0.546 | 0.019 |
| Subject 17 | 1.109 | 0.054 | 0.499 | 0.027 | 1.691 | 0.193 | 0.511 | 0.013 | 0.496 | 0.017 |
| Subject 18 | 1.350 | 0.065 | 0.565 | 0.027 | 1.589 | 0.187 | 0.653 | 0.014 | 0.637 | 0.019 |
| Subject 19 | 2.202 | 0.110 | 0.422 | 0.027 | 1.710 | 0.161 | 0.319 | 0.012 | 0.303 | 0.016 |
| Subject 20 | 1.785 | 0.079 | 0.413 | 0.027 | 1.818 | 0.177 | 0.404 | 0.009 | 0.389 | 0.013 |
| Subject 21 | 1.372 | 0.067 | 0.583 | 0.027 | 1.440 | 0.189 | 0.567 | 0.017 | 0.552 | 0.021 |
| Subject 22 | 1.365 | 0.062 | 0.524 | 0.027 | 1.143 | 0.177 | 0.556 | 0.016 | 0.540 | 0.020 |
| Subject 23 | 2.239 | 0.112 | 0.348 | 0.027 | 1.965 | 0.183 | 0.316 | 0.011 | 0.301 | 0.015 |
| Subject 24 | 1.447 | 0.062 | 0.427 | 0.027 | 1.215 | 0.161 | 0.444 | 0.016 | 0.428 | 0.020 |
| Subject 25 | 1.141 | 0.054 | 0.532 | 0.027 | 1.554 | 0.195 | 0.501 | 0.012 | 0.486 | 0.016 |

*Note*. Intertrial variability of z (sz), v (sv), and t (st) were estimated only at a group level: sz = 0.120, std =.06; sv = 1.161, std = 0.102; st = 0.244, std =.009. In every subject, the t parameter was smaller in the written condition than in the spoken condition. This suggests fixed effect at the level of the model structure and random effect at the parameter level.

## 6.2 Evidence-accumulation Models

Bayesian model comparison shows that the model comprising the nondecision time as free parameter performed better than the other models (Table 1). Table 3 shows the parameter estimates at both group and subject levels. The nondecision time parameter was larger in the spoken condition than in the written condition (PPs = 1, Figure 5). Interestingly, we observed this effect not only at a group level but also in all subjects. Posterior predictive checks show that the model fairly reproduced the observed data (Table 4).

*Table 4.* Posterior Predictive Checks

| Statistic | Observed | Model generated | SD | SEM | MSE | Credible |
|---|---|---|---|---|---|---|
| Accuracy | 0.81 | 0.81 | 0.05 | 0.0003 | 0.003 | True |
| RT (Upper bound) | 860 ms | 882 ms | 0.10 | 0.004 | 0.01 | True |
| RT (Lower bound) | -959 ms | -982 ms | 0.17 | 0.0006 | 0.03 | True |

*Note.* Observed and model-generated values are collapsed across conditions. Upper bound = correct responses, Lower bound = incorrect responses, SD = standard deviation, SEM = standard error of the mean, MSE = mean squared error, Credible (True) = in the 95% credible interval.

### 6.3 Post-hoc Analysis

As per suggestion of our reviewers and from a language production perspective (Margolin, 1984; Olive, 2014; Torrance et al., 2018), we performed a post-hoc analysis to evaluate two alternative explanations of why the written condition yielded fast RTs compared with the spoken condition. First, if the written summaries were longer or more detailed (e.g., with more propositional content) than the spoken summaries, then the inclusion of the target words in the summaries was perhaps more likely and could decrease the latency of the probe word in the recognition task. Second, it could be the case that the amount of time engaged in the summary task correlates with subsequent recognition accuracy and RT. If the effect persists after accounting for time on task, then it could indicate that writing enhances the consolidation of the words in LTM because of the knowledge-transforming process elicited through writing. If it does not persist, then the results may instead imply that writing merely engages the learner for a longer amount of time in thinking about the text without altering the qualitative nature of the learning process. To resolve these uncertainties, we first compared the quality of the summaries and the specific amount of time allocated to summary production within the 90 s window (time on task).

From the perspective of the communicability theory (Parodi, 2011), we assumed that summary quality indexes propositional content and production length in terms of (0 – 4) rating scale (Silva, 2016) with the following criteria: (0) no production, (1) retrieval of just the main idea, (2) retrieval of the main-idea along with non-specific details, (3) retrieval of the main idea along with specific supporting details, and (4) retrieval of the main idea along with elaborative rehearsal (e.g., paraphrasis). Based on this scale, a production graded with "3" was longer than a production graded with "3" was longer than a production graded with "1".

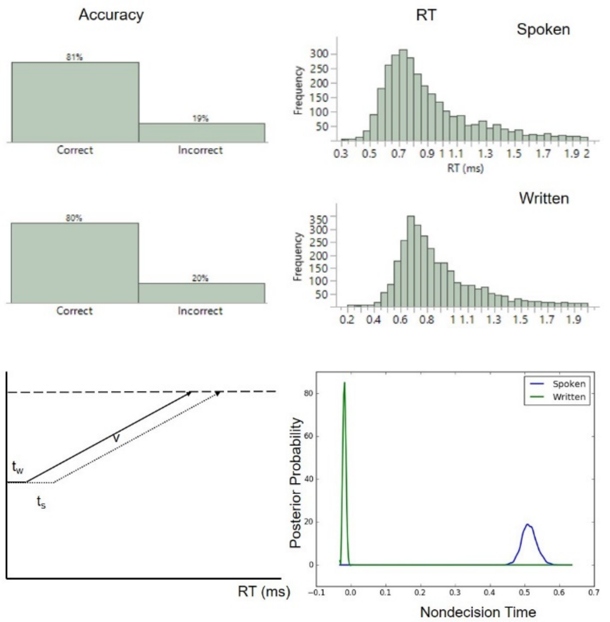


*Figure 5.* Accuracy and RT Histograms of Observed Responses, Visual Representation of the Winning Model, and Difference in Posterior Probabilities of "t" Estimates. The winning model was fitted to the distributions of RT and accuracy data of both conditions. The model reproduced these distributions (see also Table 4). Nondecision time increased in the spoken condition. The posterior distributions of "t" estimates do not overlap, meaning that they fairly differ. The 100% of posterior proportions (PP) differ from 0. The actual parameter value of the written condition is relative to the spoken condition (0.51 - 0.015).

Likewise, a production graded with "4" had more propositional content than a production graded with "3"[2]. Time on task was logged automatically by the stimulus delivery program.

Following, we fit four mixed-effects linear models to the RT data of correct trials. Model 1 comprised the main effect of production modality. Model 2 comprised the main effect of summary quality, main effect of linguistic modality, and the Summary Quality × Production Modality interaction. Model 3 included the main effects of production modality, time on task, and the Production Modality × Time on Task interaction. We also fit a full model (model 4) comprising the main effects of production modality, summary quality, time on task, and all possible interactions. In all models, subjects were included as random effects. We relied on model comparison procedures to evaluate the post-hoc hypotheses, using the lowest AIC (Akaike information criterion) number as a decision rule for model selection. If any of the post-hoc hypotheses accounts for the differential effect on RT, the lowest AIC number would be associated to either model 2, 3, or 4.

The results of this post hoc analysis showed that the quality of the summaries was better in the written condition (*M* = 2.60, *SD* = 0.50) than in the spoken condition (*M* = 2.38, *SD* = 0.63); $\mu$ = 0.23, PP = .97 (Figure 6). Furthermore, participants took more time to produce a summary in the written condition (*M* = 78.35 s, *SD* = 11.35 s than in the spoken condition (*M* = 34.15 s, *SD* = 11.9 s); $\mu$ = 44.2, PP = .999 (Figure 6). The results of the mixed-effects models showed that the simple model 1 outperformed all the other more complex models ($AIC_{model-1}$ = -93.84, $AIC_{model-2}$ = -81.80, $AIC_{model-3}$ = -64.71, $AIC_{model-4}$ = -25.74). RT was longer in the spoken condition than in the written condition, β = .01, (SE = .007).

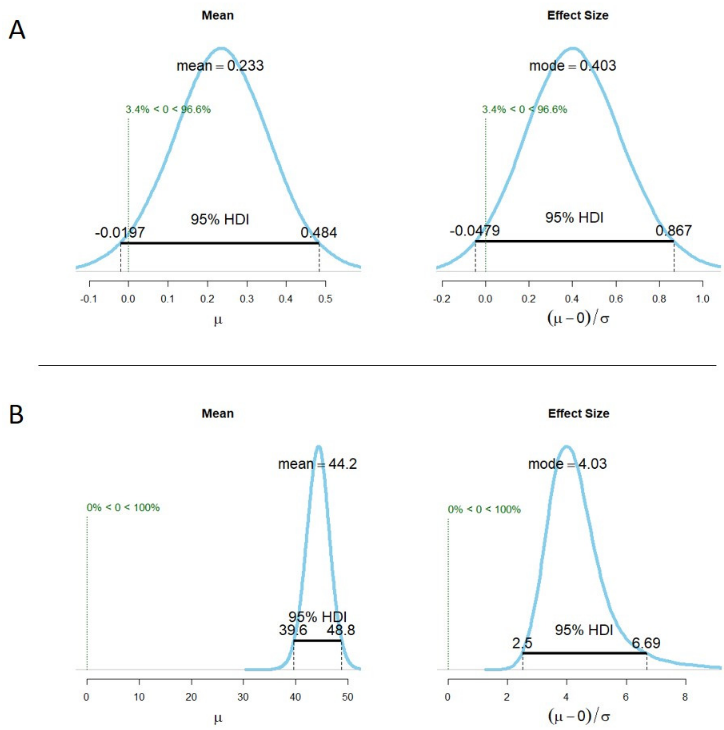


*Figure 6.* Results of the Post-hoc Analysis. Posterior Distributions of Parameter Estimates about Differences in summary quality (A) and time on task (B) across conditions.

## 7. Discussion

We introduced a paradigm to measure the effects of writing on LTM at a millisecond scale. Participants read short passages, wrote or voiced a summary, and performed an episodic-memory word recognition task. After producing the summary, participants read words and decided as quickly as possible whether each word appeared in the passage. With this paradigm, we evaluated the hypothesis that writing increases LTM consolidation (indexed by decreased word-recognition time) and explored how a drift-diffusion model captured this effect. Bayesian estimation methods showed that written productions decreased the RT of correct responses, compared with spoken productions. This differential effect mapped onto the model's nondecision parameter. Our results provide the first mental-chronometry evidence of the effect of writing on learning.

The fact that incorrect responses were faster than correct responses constitutes a proof of concept of the use of speeded memory recognition task in the context of writing to learn. This is because in 2AFC decision making tasks incorrect responses are faster than correct responses (Ratcliff, 1985). Another proof of concept of the method is the fact that writing did not increase accuracy (compared with speaking) in the memory task. In practice, RT and accuracy are negatively correlated in 2AFC decision making tasks (Heitz, 2014). This causes the subjects to engage in a speed accuracy trade off. This is relevant for future works in writing to learn because the experimenter could manipulate this trade-off effect depending upon the research objective. For example, if one wants to observe the effect of production modality on accuracy rather than on RT we could increase experimentally the accuracy in the written condition. One could do this via condition-specific instructions. In the written condition, we could ask the subjects to pay more attention to the accuracy of the responses and "to take their time to respond". In the spoken condition, we could ask them to respond both as accurately and as quickly as possible. Based on the expected speed-accuracy trade off we should expect higher accuracy, but longer RT, in the written condition than in the spoken condition.

The HDDM results suggest that writing facilitates initial stages of learning by decreasing –via episodic cues– either the encoding phase of memory retrieval or the motor response. Heuristically, this explains why "when we write the summary of a text we tend to retrieve words that we believe were present in the text". Note that adding the decision threshold (M7) or starting point (M10) as free parameters could have increased the goodness of fit of the model. However, DIC values suggest that this gain in fit would have been accompanied by a loss in generalizability, meaning that these more complex models would fit noise.

A qualitative reading of these results is that to make a correct response about whether a test word was present or not in a passage, participants accumulated the same amount of evidence in both conditions, began accumulating evidence at the same starting point, and at the same rate. However, participants needed less time to perform nondecision processes (i.e., stimulus encoding and/or motor response) after writing a summary than after pronouncing the summary.

## 7.1 Language production and memory retrieval: a specific writing effect

In our paradigm, we found a specific differential effect of writing on episodic-memory recognition using a spoken production as a control condition. The most credible value of the effect size was Cohen's d = - 0.39 –note that the negative sign only indicates the direction of the comparison. In a previous work (Arnold et al., 2017), participants performed problem-solving tasks two days after performing learning activities that required retrieval (free recall and essay writing) and non-retrieval (note taking and highlighting) processes. They found that (compared with highlighting) essay writing improved problem-solving performance –with an effect size of Cohen's d = 0.44. Interestingly, Arnold et al. (2017) used a no-language production baseline. In the context of our results, we suggest that not accounting for the main effect of language production could cause a larger effect of writing because spoken productions increase learning more than no-language activities (MacLeod & Bodner, 2017).

From a language production perspective (Margolin, 1984; Olive, 2014; Torrance et al., 2018), the post-hoc analysis showed that peripheral processes such as the physical act of forming the letters or encoding the letter sequence during spelling (i.e., the articulatory stage of language production) do not account for our findings. This speaks to central processes such as the planning or conceptualization phases as the loci of the differential effect of writing. One central process that could be responsible for this effect is memory retrieval. Trying to retrieve material from LTM following a period of initial study leads to better LTM consolidation than a further period of study for equivalent time (Karpicke & Roediger, 2008; Roediger & Butler, 2011). This suggests that "mere retrieval" - comprised, for example, in knowledge telling or dispositionally-guided text production - is sufficient to lead to learning.

Interestingly, we find congruent the above fact with a previous work in which writing activities that demanded memory retrieval led to better results on learning than non-retrieval writing activities (e.g., note taking) (Arnold et al., 2017). However, our control (spoken) condition also demanded memory retrieval. If we assume the Donder's subtraction method (Donders, 1969) - still prevalent in the mental-chronometry literature - the writing effect detected here should be associated with additional elaborative processes (i.e., elaborative rehearsal) per–formed over the retrieved information. These processes might not take place in the spoken condition. It is worth noting however, that whereas the subtraction methods reveal processes that are unique to writing it does so by subtracting the effect of processes that are shared by both modalities.

We speculate that reconstructive processes and text revision could be two writing-specific operations that facilitate LTM consolidation. When the writer attempts to summarize the text, he or she no longer has access to the source text but instead faces the uncertainty about how well the summary captures the meaning extracted from the text. At this point, there are two sources of uncertainty. The first source refers to how well the extracted meaning corresponds to the text whereas the second source refers to how well the summary text under construction captures the meaning of the text. With reconstructive processes, the writer adds in plausible inferences about what the text must have said whereas with text revision the writer formulates new text to bring the emerging summary into line with their conceptual representation. Reading (i.e., revising) the externalized text then reduces the uncertainty. These processes sound very like the knowledge-constituting process (Galbraith & Torrance, 1999). An initial attempt to summarize conceptual content mismatches the underlying content so a new synthesis is generated to express the missing content, which in turn fails to fully capture the content, so a further synthesis takes place. The result is an explicit proposition or series of propositions in episodic memory.

As revealed by the winning HDDM, the observable consequence of a specific effect of writing manifests itself at either the encoding phase or the motor response phase during the retrieval process of the word-recognition task. The nondecision parameter does not differentiate between these two phases. We found this very interesting because in emerging theories of brain function such as predictive coding and active inference (Parr & Friston, 2017; Pezzulo et al., 2016) encoding processes and motor responses are two different ways through which the organism establishes an equilibrium with the environment. In particular, from this perspective we have found that premature motor responses (i.e., actions) are strongly associated with "prior beliefs" (Limongi, Bohaterewicz, Nowicka, Plewka, & Friston, 2018) or consolidated LTM which in the context of the dual-process model is equivalent to the dispositionally-guided process. This conceptual relationship could result in a nice link between the field of writing to learn and emergent theories of brain function, which would regard writing to learn as a special case of epistemic behavior –epistemic writing. Therefore, a future study could adjudicate between these processes by manipulating experimentally, for example, the orientation of the probe words. This is because word orientation affects specifically the encoding phase (Gomez & Perea, 2014).

## 7.2 Alternative explanation

Although we controlled for the effect of language production by using a spoken condition, one could think of orthographic priming as an alternative explanation for the "writing effect" on RT. The rationale underlying this possible explanation is that spoken and written productions might comprise different cognitive (and neurocognitive) paths to access lexical knowledge (Cohen-Goldberg, 2017), lexical-semantics differences (Biber, 1988), and differences between written and spoken registers (Louwerse, McCarthy, McNamara, & Graesser, 2004).

Unlike the spoken summary, the written summary involved orthographic representations. Given that the source text and the text/no-text test words were also orthographically represented, we could naturally think that general priming of orthographic representations in the written condition could have decreased the RT. However, a previous work on priming in "masked" and "unmasked" conditions (Gomez, Perea, & Ratcliff, 2013) allows us to rule out this alternative explanation. Below, we do this by detailing the relationship between the differential effect of masked and un masked priming on HDDM parameters.

Consider the priming effect of the word "book" on the recognition time of the word "library" in masked and unmasked conditions. A masked trial would comprise, for example, the following sequence: (1) a 500-ms mask (e.g., "********") followed by (2) a 60-ms prime word (e.g., "book"), and then by (3) the probe word (e.g., "library"). In the unmasked trial, there is no mask, and the prime word would last 560 ms. In the masked condition, the subjects cannot detect the briefly presented (50 ms) prime word. However, the priming effect (faster recognition of the probe word) occurs. This reduction in RT is captured by the nondecision parameter whereas the priming effect in the unmasked condition is captured by the drift rate parameter. In the current task, if there had been a priming effect of orthographic representations on RT this effect would have mapped onto the drift-rate parameter because orthographic cues would play the role of unmasked primes.

Despite our modeling argument, a future experiment could evaluate empirically the orthographic-priming explanation by including extra control conditions in which both the initial text and the recognition-memory stimuli are presented in auditory form. If the orthographic explanation for the writing superiority effect is valid, then writing should show no advantage over speech when input and output are auditory in form. By contrast, if the effect is a genuine writing superiority effect, the advantage should still be present when text and recognition cues are presented in auditory form[3].

### 7.3 Limitations

The current experiment unveils a cause-effect relationship, which is what we could expect from a laboratory study. However, it is important to interpret our results in the context of real situations. In this task, subjects did not have access to the source text when they were composing the summary. This allowed us to capitalize on the effect of retrieval on learning. However, as highlighted by one of our reviewers, this is "probably atypical of writing to learn tasks" in the classroom. Writing is generally intermixed with reading, which makes writing to learn "a discourse synthesis task". In the same line of arguments, the writing to learn literature has often included measures that are not merely recalled but require learners to draw new inferences or apply their knowledge to solve novel problems. Arguably, inference tasks are good measures of the higher level of learning that educators pursue with writing to learn activities. Interestingly, the outcome of inference tasks can also be studied using mental chronometry and computational models, being a tremendous tool for measurement purposes in education.

Another limitation of the current study is worth noticing. A major goal of our work was to make initial progress on the formal (i.e., mathematical) understanding of the effect of writing on learning. To achieve this, we opted to identify first - as we indeed did - the main effect of linguistic modality (i.e., writing vs. speaking). With this in mind, we fully randomized the stimuli across participants, conditions, and trials to control for a series of linguistic variables (e.g., word length and word frequency) that could affect the diffusion-model parameters (Aschenbrenner, Balota, Gordon, Ratcliff, & Morris, 2016; Ratcliff, Thapar, & McKoon, 2004). However, we assumed that such variables affected both conditions alike. Future works should experimentally evaluate this assumption.

### 7.4 Towards a cognitive neuroscience of writing to learn: combining neuroimaging, computational, and mental-chronometry

Our primary focus in what follows is detailing how the current paradigm could serve as a link between the description of writing-to-learn operations, the quantitative measurements of their ensuing effects, and their neural correlates. We think that a precise description of this link could further the goal of the most important claim of educational neuroscience, "…that new insights about the brain can improve classroom teaching" (Blakemore & Frith, 2005). As an exemplar case, we sketch how our approach and results could take the initiative comprised in the DPM one step further. Specifically, we could achieve a CLS-grounded DPM of writing to learn by combining the current methodology with dynamic causal models (DCM) of functional magnetic resonance imaging (fMRI) (Forstmann et al., 2016; Friston, 2007). This combined strategy is a gold standard in cognitive neuroscience and is currently applied in studies ranging from basic perception and learning processes (Stephan & Friston, 2010) to neuropsychiatric disorders (Limongi et al., 2018). Therefore, it could be used straightforwardly to test the implementational predictions of the DPM. Naturally, this approach would inform current proposals of the relationship between neuroscience and education.

Recall that the DPM assumes that writer's disposition (i.e., knowledge constituting) leads to learning via recursive activation of semantic memory (in the cortex). Conversely, writing would lead to learning via problem-solving (knowledge-transforming) if it focuses on stable activations of episodic memories (in the hippocampus)[4]. The DPM makes predictions of these two cases during the first stages of LTM consolidation (during the first milliseconds after completing the writing activity) at both the algorithmic (drift-diffusion models) and implementational (dynamic causal models) levels of analysis.

At the algorithmic level of analysis, the first case should impact the drift rate of the drift-diffusion model, because semantic activation during word recognition is associated with the quality of the encoded information (Ratcliff, Gomez, et al., 2004). The second case, as this work reveals, implies changes in the nondecision parameter. Furthermore, the DPM proposes that *optimal* learning takes place via conjoint deployment of both processes, which should map onto both the drift rate and the nondecision parameters. Interestingly, previous works have shown that low and high self-monitors tend to learn by deploying the dispositionally-guided and problem-solving processes respectively (Baaijen & Galbraith, 2018).

Note that knowledge constituting alone could affect the drift rate at later stages of LTM consolidation, but not during the first milliseconds after the production phase (i.e., the current experiment). For example, if subjects worked on the same summary over and over, we would expect a stronger cortical participation, as an index of stronger LTM consolidation. This consolidation would map onto a change in the drift rate. We speculate that this could occur when a subject works on the same draft for several days. This might nicely explain why "putting away" a draft a few days before proofreading it allows the writer to detect errors. The draft has not changed but the writer's knowledge system certainly has!

At the implementational level, the DPM predicts transient changes in connectivity strengths between the cortex and the hippocampus (writer's disposition) and within the hippocampus (problem-solving). These predictions could be tested via DCM of fMRI data. In DCM, two types of inputs give rise to the effect that one brain region (e.g., the hippocampus) exerts over another region (e.g., the entorhinal cortex): driving and modulatory inputs (Figure 7).

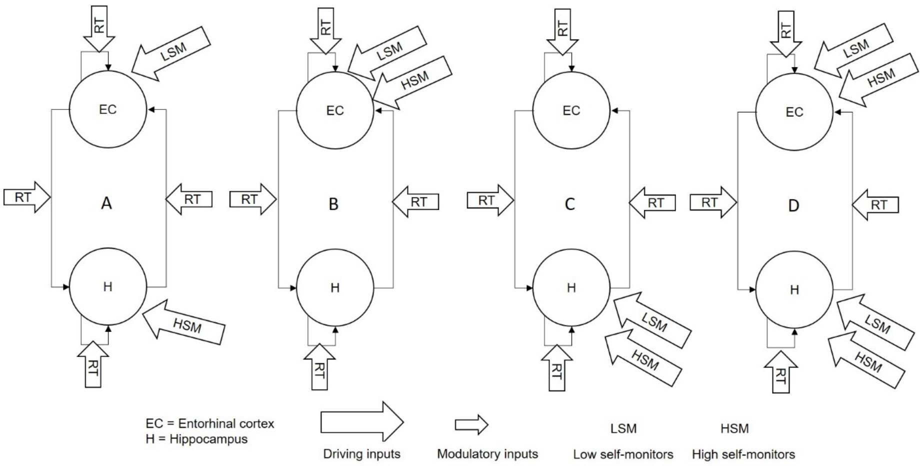


*Figure 7.* Dual-process Model Predictions at the Implementational Level of Analysis. Three dynamic causal models represent predictions informed by the current results (A) and two competing models (B - D).

Whereas driving inputs directly "perturb" a region, modulatory inputs affects the connectivity strength between regions. Furthermore, both types of inputs can modulate within-region (intrinsic) connections. We can map the DPM predictions on DCM inputs by setting trial-wise RT measurements as modulatory inputs on connections and groups (high vs. low self-monitors) as direct inputs on the hippocampus and the cortex respectively. This example provides a feasible (i.e., biologically realistic) approach for the study of the neural basis of writing to learn congruent with current algorithmic models.

## 8. Concluding remarks

On a general note, the acute reader can see that our main findings neither contradict nor confirm the assumptions of any writing-to-learn model (or any theory about the epistemic properties of writing). This is because we did not aim at resolving the question of which (algorithmic) set of operations fulfils the (computational) learning goal of writing. Rather, with the current paradigm we pursued a more modest theoretical ambition and a more general methodological tool that can be used by researchers in the writing-to-learn field. We tried to substantiate the idea that research on writing to learn should include mental chronometry and mathematical models. There are other models of writing to learn whose predictions could be tested using this approach –e.g., Klein et al. (2017), even when such models are self-contained at the algorithmic level of analysis. This is because the common denominator among these models is that they predict specific writing effects on LTM consolidation. More interestingly and with a broader perspective, some reductionist neuroscientific approaches suggest that computational goals can be explained directly from an implementational perspective (Bickle, 2015) –i.e., bypassing the algorithmic level. Regarding this point of view, the approach on offer here opens a research strategy for testing, for example, innovative hypotheses on epistemic writing relying on predictive coding and active inference (Silva & Limongi, 2017, 2018), which links directly the computational and implementational levels.

**Notes**

1. The information processing theory regards "encoding" as the process or set of processes that the system executes to store information in LTM. In this work, we use LTM consolidation. In the Experimental Approach section, we regard "encoding" as a nondecision parameter of the drift-diffusion model.
2. We acknowledge that the definitions of production length and propositional content vary across linguistic theories.
3. We acknowledge and thank one anonymous reviewer form stating this alternative explanation and the related experiment.
4. Current developments of the CLS suggest that the hippocampus could also mediate the generation of abstract patterns (Schapiro et al., 2017)

**Acknowledgements**

We thank the "Instituto de Literatura y Ciencias del Lenguaje" of "Pontificia Universidad Católica de Valparaíso" and "Instituto Venezolano de Investigaciones Lingüísticas y Literarias Andrés Bello" of "Instituto Pedagógico de Caracas" for facilitating the execution of this study.

**References**

Ackerman, J. M. (1993). The Promise of Writing to Learn. *Written Communication, 10*(3), 334-370. https://doi.org/10.1177/0741088393010003002

Al-Murtadha, M. (2013). Investigating the impact of sentence writing on second language vocabulary acquisition. *Vocabulary Education & Research Bulletin, 2*(2), 2-3.

Anders, R., Riès, S., van Maanen, L., & Alario, F. X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology, 82*(Supplement C), 57-73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3-25. https://doi.org/10.1037/amp0000191

Arnold, K. M., Umanath, S., Thio, K., Reilly, W. B., McDaniel, M. A., & Marsh, E. J. (2017). Understanding the Cognitive Processes Involved in Writing to Learn. *Journal of Experimental Psychology-Applied, 23*(2), 115-127. https://doi.org/10.1037/xap0000119

Aschenbrenner, A. J., Balota, D. A., Gordon, B. A., Ratcliff, R., & Morris, J. C. (2016). A diffusion model analysis of episodic recognition in preclinical individuals with a family history for Alzheimer's disease: The adult children study. *Neuropsychology, 30*(2), 225-238. https://doi.org/10.1037/neu0000222

Baaijen, V. M., & Galbraith, D. (2018). Discovery Through Writing: Relationships with Writing Processes and Text Quality. *Cognition and Instruction*, 1-25. https://doi.org/10.1080/07370008.2018.1456431

Bangert-Drowns, R., Hurley, M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: a metaanalysis. *Review of educational research, 74*, 29-58. https://doi.org/10.3102/00346543074001029

Bazerman, C. (2018). What Does a Model Model? And for Whom? *Educational Psychologist*, 1-18.
https://doi:10.1080/00461520.2018.1496022

Bazerman, C., Little, J., Bethel, L., Chavkin, T., Fouquette, D., & Garufis, J. (2005). *Reference guide to writing across the curriculum*. West Lafayette: Parlor Press.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ.: Lawrence Erlbaum.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Bickle, J. (2015). Marr and Reductionism. *Topics in Cognitive Science, 7*(2), 299-311. https://doi.org/10.1111/tops.12134

Blakemore, S. J., & Frith, U. (2005). *The learning brain*. Oxford, UK: Blackwell.

Bowers, J. S. (2016). The practical and principled problems with educational neuroscience. *Psychological Review, 123*(5), 600-612. https://doi.org/10.1037/rev0000025

Britton, J. (1980). Shaping at the point of utterance. In A. Freedman & I. Pringle (Eds.), *Reinventing the rhetorical tradition* (pp. 61-66). Conway, AR: L & S. Books for the Canadian Council of Teachers of English.

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities*. London: Macmillan.

Brown, W. (1998). Power of self-reflection through epistemic writing. *College Teaching, 46*(4), 135-138. https://doi.org/10.1080/87567559809596258

Carter, M., Ferzli, M., & Wiebe, E. (2007). Writing to Learn by Learning to Write in the Disciplines. *Journal of Business and Technical Communication, 21*(3), 278-302. https://doi.org/10.1177/ 1050651907300466

Cohen-Goldberg, A. M. (2017). Informative Differences: An Argument for a Comparative Approach to Written, Spoken, and Signed Language Research. In S. Plane, C. Bazerman, F. Rondelli, C. Donahue, A. Applebee, C. Boré, P. Carlino, M. Larruy, P. Rogers, & D. Russell (Eds.), *Research on writing: multiple perspectives* (pp. 457-476). Fort Collins: The WAC Clearinghouse.

Collins, A. M., & Loftus, E. (1975). *A Spreading Activation Theory of Semantic Processing* (Vol. 82).

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294. https://doi.org/10.1037//0096-3445.104.3.268

De Brigard, F., Brady, T. F., Ruzic, L., & Schacter, D. L. (2017). Tracking the emergence of memories: A category-learning paradigm to explore schema-driven recognition. *Memory & Cognition, 45*(1), 105-120. https://doi.org/10.3758/s13421-016-0643-6

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30*, 412-431. https://doi.org/10.1016/0001-6918(69)90065-1

Emig, J. (1977). Writing as a Mode of Learning. *College Composition and Communication, 28*(2), 122-128.

Flower, L., & Hayes, J. (1980). *The dynamics of composing: making plans and juggling constraints. Cognitive Processes in Writing*. Hillsdale New Jersey: Lawrence Earlbaum.

Flower, L., & Hayes, J. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication, 32*(4), 365-387. https://doi.org/10.2307/356600

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annu Rev Psychol, 67*, 641-666. https://doi.org/10.1146/annurev-psych-122414-033645

Friston, K. J. (2007). Dynamic Causal Models for fMRI. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (pp. 541-560). London: Elsevier. https://doi.org/10.1016/b978-012372560-8/50041-3

Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science, 21*(1-3), 45-71. https://doi:10.1007/BF00119655

Galbraith, D. (2009). Writing as discovery. *British Journal of Educational Psychology Monograph Series II 6 - Teaching and Learning Writing*, 5-26. https://doi.org/10.1348/978185409X421129

Galbraith, D., & Baaijen, V. M. (2015). Conflict in writing: Actions and objects. In G. Cislaru (Ed.), *Writing(s) at the crossroads: the process/product interface* (pp. 253-276). Amsterdam: John Benjamins Publishing. https://doi.org/10.1075/z.194.13gal

Galbraith, D., & Torrance, M. (1999). Writing as a knowledge-constituting process. In *Knowing What to Write* (pp. 137-157). Amsterdam, NL: Amsterdam University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science, 7*(4), 457-472. https://doi.org/10.1214/ss/1177011136

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *The Quarterly Journal of Experimental Psychology, 67*(12), 2455-2466. https://doi.org/10.1080/17470218.2014.937447

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance, 39*(6), 1731-1740. https://doi.org/10.1037/a0032333

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*(150). https://doi.org/10.3389/fnins.2014.00150

Howard-Jones, P. A., Varma, S., Ansari, D., Butterworth, B., De Smedt, B., Goswami, U., . . . Thomas, M. S. C. (2016). The principles and practices of educational neuroscience: Comment on Bowers (2016). *Psychological Review, 123*(5), 620-627. https://doi.org/10.1037/rev0000036

Johansson, R., Oren, F., & Holmqvist, K. (2018). Gaze patterns reveal how situation models and text representations contribute to episodic text memory. *Cognition, 175*, 53-68. https://doi.org/10.1016/j.cognition.2018.02.016

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science, 319*(5865), 966-968. https://doi.org/10.1126/science.1152408 Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta Coordinated Error-Driven Learning in the Hippocampus. *Plos Computational Biology, 9*(6), e1003067. https://doi.org/10.1371/journal.pcbi.1003067

Klein, P. D. (1999). Reopening Inquiry into Cognitive Processes in Writing-To-Learn. *Educational Psychology Review, 11*(3), 203-270. https://doi:10.1023/A:1021913217147

Klein, P. D., & Boscolo, P. (2016). Trends in Research on Writing as a Learning Activity. *Journal of writing research, 7*(3), 311-350. https://doi.org/10.17239/jowr-2016.07.03.01

Klein, P. D., Haug, K. N., & Arcon, N. (2017). The Effects of Rhetorical and Content Subgoals on Writing and Learning. *The Journal of Experimental Education, 85*(2), 291-308. https://doi.org/10.1080/00220973.2016.1143795

Kruschke, J. K. (2013a). Bayesian estimation supersedes the t test. *J Exp Psychol Gen, 142*(2), 573-603. https://doi:10.1037/a0029146

Kruschke, J. K. (2013b). Posterior predictive checks can and should be Bayesian: Comment on Gelman and Shalizi, 'Philosophy and the practice of Bayesian statistics'. *British Journal of Mathematical and Statistical Psychology, 66*(1), 45-56. https://doi:10.1111/j.2044-8317.2012.02063.x

Kruschke, J. K., & Liddell, T. M. (2017a). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*. https://doi:10.3758/s13423-017-1272-1

Kruschke, J. K., & Liddell, T. M. (2017b). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*. https://doi:10.3758/s13423-016-1221-4

Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.

Langer, J., & Applebee, A. (1987). *How writing shapes thinking: a study of teaching and learning*. Retrieved from Urbana, Illinois:

Lepage, M., Habib, R., & Tulving, E. (1998). Hippocampal PET activations of memory encoding and retrieval: the HIPER model. *Hippocampus, 8*(4), 313-322. https://doi.org/10.1002/(sici)1098-1063(1998)8:4<313::aid-hipo1>3.0.co;2-i

Limongi, R., Bohaterewicz, B., Nowicka, M., Plewka, A., & Friston, K. J. (2018). Knowing when to stop: Aberrant precision and evidence accumulation in schizophrenia. In *Schizophrenia Research* (Vol. 197, pp. 386-391). https://doi.org/10.1016/j.schres.2017.12.018

Louwerse, M., McCarthy, P., McNamara, D., & Graesser, A. (2004). *Variation in language and cohesion across written and spoken registers.* Paper presented at the 26th Annual Meeting of the Cognitive Science Society, Mahwah, NJ.

MacLeod, C. M., & Bodner, G. E. (2017). The Production Effect in Memory. *Current Directions in Psychological Science, 26*(4), 390-395. https://doi:10.1177/0963721417691356

Margolin, D. I. (1984). The neuropsychology of writing and spelling: Semantic, phonological, motor, and perceptual processes. *The Quarterly Journal of Experimental Psychology Section A, 36*(3), 459-489. https://doi.org/10.1080/14640748408402172

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.

Martínez, I., Mateos, M., Martín, E., & Rijlaarsdam, G. (2015). Learning history by composing synthesis texts: Effects of an instructional programme on learning, reading and writing processes, and text quality. *Journal of Writing Research, 7*(2), 275-302. https://doi.org/10.17239/jowr-2015.07.02.03

Mateos, M., Martin, E., Villalon, R., & Luna, M. (2008). Reading and writing to learn in secondary education: online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing, 21*(7), 675-697. https://doi.org/10.1007/s11145-007-9086-6

McCutchen, D., Teske, P., & Bankston, C. (2008). Writing and cognition: implications of the cognitive architecture for learning to write and writing to learn. In C. Bazerman (Ed.), *Handbook of research on writing* (pp. 451-470). New York: L. Erlbaum Associates.

McLeod, S., Miraglia, E., Soven, M., & Thaiss, C. (2001). *WAC for the millennium*. Chicago: Urbana Illinois University.

Miller, C. (1984). Genre as social action. *Quartely Journal of Speech, 70*, 151-167.

Mynlieff, M., Manogaran, A. L., St. Maurice, M., & Eddinger, T. J. (2014). Writing Assignments with a Metacognitive Component Enhance Learning in a Large Introductory Biology Course. *CBE-Life Sciences Education, 13*(2), 311-321. https://doi.org/10.1187/cbe.13-05-0097

Navarro, F., Ávila Reyes, N., Tapia-Ladino, M., Cristovão, V. L. L., Moritz, M. E. W., Narváez Cardona, E., & Bazerman, C. (2016). Panorama histórico y contrastivo de los estudios sobre lectura y escritura en educación superior publicados en América Latina. *Revista signos, 49*, 78-99. https://doi.org/10.4067/s0718-09342016000400006

Nevid, J. S., Ambrose, M. A., & Pyun, Y. S. (2017). Effects of Higher and Lower Level Writing-to-Learn Assignments on Higher and Lower Level Examination Questions. *Teaching of Psychology, 44*(4), 324-329. https://doi:10.1177/0098628317727645

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary Learning Systems. *Cognitive Science, 38*(6), 1229-1248. https://doi:10.1111/j.1551-6709.2011.01214.x

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2012). Learning and Memory. In R. C. O'Reilly, Y. Munakata, M. J. Frank, T. E. Hazy, & Contributors (Eds.), Computational Cognitive Neuroscience. . Colorado: Wiki Book. Retrieved from http://ccnbook.colorado.edu.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the learning complementary systems framework. *Trends in Cognitive Sciences, 6*(12), 505-510. https://doi.org/10.1016/s1364-6613(02)02005-3

Olive, T. (2014). Toward a parallel and cascading model of the writing system: A review of research on writing processes coordination. *Journal of writing research, 6*(2), 173-194. https://doi.org/10.17239/jowr-2014.06.02.4

Parodi, G. (2011). La Teoría de la Comunicabilidad: Notas para una concepción integral de la comprensión de textos escritos. *Revista signos, 44*, 145-167. https://doi.org/10.4067/s0718-09342011000200004

Parr, T., & Friston, K. J. (2017). Working memory, attention, and salience in active inference. *Scientific Reports, 7*(1), 14678. https://doi.org/10.1038/s41598-017-15249-0

Peebles, D., & Cooper, R. P. (2015). Thirty Years After Marr's Vision: Levels of Analysis in Cognitive Science. *Topics in Cognitive Science, 7*(2), 187-190. https://doi.org/10.1111/tops.12137

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., & Lesley, H. (2002). The lexical quality hypothesis. In L. Vehoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/swll.11.14per

Perfetti, C. A., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading, 18*(1), 22-37. https://doi.org/10.1080/10888438.2013.827687 Petrucci, P. R. (2002). A Writing-to-Learn Approach to Writing in the Discipline in the Introductory Linguistics Classroom. *The WAC Journal, 13*, 133-143.

Pezzulo, G., Cartoni, E., Rigoli, F., Pio-Lopez, L., & Friston, K. J. (2016). Active Inference, epistemic value, and vicarious trial and error. *Learning & Memory, 23*(7), 322-338. https://doi.org/10.1101/lm.041780.116 Plane, S., Bazerman, C., Rondelli, F., Donahue, C., Applebee, A., Boré, C., . . . Russell, D. (2017). *Research on writing: multiple perspectives*. Fort Collins, Co: The WAC Clearinghouse.

Prichard, J. (2005). Writing to Learn: An Evaluation of the Calibrated Peer Review™ Program in Two Neuroscience Courses. *The Journal of Undergraduate Neuroscience Education, 4*(1), A34-A39.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108. https://doi.org/10.1037/0033-295x.85.2.59 Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review, 92*(2), 212-225. https://doi.org/10.1037//0033-295x.92.2.212

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review, 111*(1), 159-182. https://doi.org/10.1037/0033-295x.111.1.159

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences, 20*(4), 260-281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*(4), 408-424. https://doi.org/10.1016/j.jml.2003.11.002

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27. https://doi.org/10.1016/j.tics.2010.09.003

Russell, D. (2002). *Writing in the academic disciplines*. Carbondale, IL.: Southern Illinois University.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B-Biological Sciences, 372*(1711). https://doi.org/10.1098/rstb.2016.0049

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.

Silva, A. M. (2016). *Efectos de la modalidad de producción lingüística en la memoria de largo plazo: una evidencia experimental acerca del potencial epistémico-afectivo de la escritura.* (Ph.D in Linguistics), Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile. Retrieved from http://www.giovanniparodi.cl/giovanniparodi/site/artic/20170417/asocfile/20170417121641/tesis_escrituraepistemica_angelicasilva_2016.pdfhttps://doi.org/10.4067/s0718-68512014000100015

Silva, A. M., & Limongi, R. (2017). La Escritura Epistémica en Contextos Académico Profesionales: Desafíos de Investigación Educativa, Cognitiva y Neurocientífica. *Revista Educación Superior y Sociedad (ESS)*(1), 41-58.

Silva, A. M., & Limongi, R. (2018). *Evidence Accumulation Models Confirm the Epistemic Property of Writing: A Predictive Coding and Active Inference Study of Writing to Learn*. Paper presented at the Neuroscience 2018, San Diego, CA.

Spaniol, J., Madden, D. J., & Voss, A. (2006). A Diffusion Model Analysis of Adult Age Differences in Episodic and Semantic Long-Term Memory Retrieval. *Journal Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 101-117. https://doi.org/10.1037/0278-7393.32.1.101

Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: Message level and lexical level influences in ERPs. *Neuropsychologia, 64*, 41-53. https://doi.org/10.1016/j.neuropsychologia. 2014.09.012

Stephan, K. E., & Friston, K. J. (2010). Analyzing effective connectivity with functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews-Cognitive Science, 1*(3), 446-459. https://doi.org/10.1002/wcs.58

Taylor, J. N., & Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing, 29*(6), 1069-1103. https://doi.org/10.1007/s11145-015-9616-6

Thaiss, C., & Porter, T. (2010). The State of WAC/WID in 2010: Methods and Results of the U.S. Survey of the International WAC/WID Mapping Project. *College Composition and Communication, 61*(3), 524-570.

Torrance, M., Nottbusch, G., Alves, R. A., Arfe, B., Chanquoy, L., Chukharev-Hudilainen, E., . . . Wengelin, A. (2018). Timed written picture naming in 14 European languages. *Behaviour Research Methods, 50*(2), 744-758. https://doi.org/10.3758/s13428-017-0902-x

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E., & Craik, F. (2000). *The Oxford Handbook of Memory*. New York: Oxford Univerity Press.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Front Neuroinform, 7*, 14. https://doi.org/10.3389/fninf.2013.00014

Young, A. (2006). *Teaching writing across the curriculum* (4th ed.). New Jersey: Pearson Prentice Hall.

Young, R., & Sullivan, P. (1984). Why write? A reconsideration. In R. Conners, L. Ede, & A. Lunsnford (Eds.), *Essays on classical rhetoric and modern discourse* (pp. 215-225). Carbondale, IL: Southern Illinois University Press.

## Appendix A

*Sample Passage (and Test Words) with Expositive Rhetorical Structure and with Predominant Paratactic Organization*

| Text | Test Word | |
|---|---|---|
| | Text | No Text |
| Los murciélagos tienen características fascinantes relacionadas con su genética, estilo de vida y habilidades de supervivencia. En relación a la genética, todas las especies de murciélagos existentes pertenecen al reino animal de los mamíferos con la característica esencial de que pueden volar. Acerca de su estilo de vida puede decirse que son animales que duermen suspendidos -cabeza abajo- de las ramas de los árboles o también pueden encontrarse en las salientes de las piedras en alguna cueva. Entre la variedad bien documentada de especies de murciélagos sorprenden los llamados murciélagos come-insectos. Los murciélagos comen insectos se alimentan de parásitos que dañan las cosechas y enferman a los animales de las granjas. En cuanto a las habilidades de supervivencia de los murciélagos es asombrosa la habilidad de escuchar el caminar de insectos o el aletear de los que tienen alas. Pero la mayor de las habilidades es el sistema de eco-localización, es decir, un sistema de traslado y referencia de cuerpos que impide que una onda de sonido se propague. Así, pues, los murciélagos en pleno vuelo emiten sonidos de muy baja frecuencia que chocan con los cuerpos y se devuelve hasta ellos para demarcar la ubicación de objetos como la presa. La eco-localización en los murciélagos es, por tanto, como una especie de radar que les impide chocar contra árboles y rocas en la naturaleza porque los murciélagos, a pesar de tener ojos como todo mamífero, no pueden ver. | murciélagos<br>reino<br>cueva<br>piedras<br>mamíferos<br>ojos<br>sonidos<br>alas<br>sistema<br>destrezas<br>genética<br>especie<br>rocas<br>árboles<br>naturaleza | león<br>montaña<br>cielo<br>mar<br>volcán<br>avión<br>presidente<br>ciudad<br>lápiz<br>tren<br>motor<br>pared<br>pizza<br>almohada<br>cuaderno |
| Bats have fascinating characteristics related to their genetics, lifestyle and survival skills. In relation to genetics, all existing bat species belong to the animal kingdom of mammals with the essential characteristic that they can fly. About their lifestyle it can be said that they are sleeping animals suspended-head down-from the branches of the trees or they can also be found in the ledges of the stones in some cave. Among the well-documented variety of bat species, the so-called insect-eating bats are surprising. Insect-eating bats feed on pests that damage crops and make farm animals sick. As for the survival skills of bats It is amazing the ability to hear the walking of insects or the flap of those with wings. But the greatest of the skills is the eco-localization system, that is to say, a system of transfer and reference of bodies that prevents a sound wave from spreading. Thus, bats in mid-flight emit very low-frequency sounds that collide with the bodies and are returned to them to demarcate the location of objects such as prey. The eco-localization in bats is, therefore, a kind of radar that prevents them from colliding with trees and rocks in nature because bats, despite having eyes like all mammals, cannot see. | bats<br>kingdom<br>cave<br>stones<br>mammals<br>eyes<br>sound<br>wings<br>system<br>skills<br>genetics<br>species<br>rocks<br>trees<br>nature | lions<br>mountains<br>heaven<br>sea<br>volcano<br>aircraft<br>president<br>city<br>pencils<br>train<br>motor<br>wall<br>pizza<br>pillow<br>notebook |

Appendix B.

*Sample Passage (and Test Words) with Argumentative Rhetorical Structure and with Predominant Hypotactic Organization*

| Text | Test Word | |
|---|---|---|
| | Text | No Text |
| Actualmente, la cirugía cosmética se presenta como un procedimiento quirúrgico fácil, indoloro y sin riesgo alguno que lamentar para la salud de hombres y mujeres. Así, pues, resulta muy común asociar la cirugía cosmética con procedimientos básicos tales como: el refrescamiento facial, los masajes modeladores de la figura, o el tatuaje de cejas, ojos y labios. Sin embargo, los riesgos que conlleva una cirugía cosmética incluyen reacciones adversas a la anestesia, los sangrados excesivos y las infecciones posquirúrgicas. Músculos y nervios podrían dañarse durante la cirugía; por tanto, el/la paciente estaría en riesgo de sufrir parálisis. También existe la posibilidad de que el propio cuerpo rechace el implante (por ejemplo, una prótesis mamaria o de glúteos) que ha sido colocado en la cirugía. Aún si el paciente se recupera satisfactoriamente de la cirugía, existe el riesgo de obtener resultados poco gratificantes. En dichos casos, es muy probable que cirugías adicionales sean requeridas para corregir algún cambio corporal no deseado. Finalmente, un resultado razonablemente exitoso de la cirugía cosmética podría, inclusive, incitar a la persona a realizarse más operaciones cosméticas por el refuerzo positivo recibido de verse bien. Pero lo cierto es que una cirugía cosmética de nariz más perfilada de lo normal o el agrandamiento de mamas u otro miembro corporal no asegura la felicidad, los buenos amigos, ni el amor verdadero | cirugía<br>procedimiento<br>salud<br>hombres<br>mujeres<br>senos<br>pacientes<br>amigos<br>persona<br>resultado<br>músculo<br>figura<br>implante<br>prótesis<br>labios | auto<br>casa<br>perro<br>bicicleta<br>lentes<br>disco<br>cinta<br>manzana<br>cable<br>taza<br>sopa<br>carne<br>pan<br>pantalla<br>zapato |
| Currently, cosmetic surgery is presented as an easy, painless and no-risk surgical procedure to lament for the health of men and women. Therefore, it is very common to associate cosmetic surgery with basic procedures such as: facial refreshing, shaping massages of the figure, or the tattoo of eyebrows, eyes and lips. However, the risks involved in cosmetic surgery include adverse reactions to anesthesia, excessive bleeding and post-surgical infections. muscles and nerves may be damaged during surgery; Therefore, the patient would be at risk for paralysis. There is also the possibility that the body itself will reject the implant (for example, a breast or buttocks prosthesis) that has been placed in the surgery. Even if the patient recovers satisfactorily from surgery, there is a risk of getting unrewarding results. In such cases, additional surgeries are likely to be required to correct any unwanted body change. Finally, a reasonably successful result of cosmetic surgery could even incite the person to perform more cosmetic operations by the positive reinforcement received from being good. But the truth is that a more profiled nose cosmetic surgery than normal or breast enlargement or another body member does not ensure happiness, good friends, or true love. | surgery<br>procedure<br>health<br>men<br>women<br>breast<br>patient<br>friends<br>person<br>results<br>muscles<br>figure<br>implant<br>prosthesis<br>lips | car<br>house<br>dog<br>bicycle<br>glasses<br>disc<br>ribbon<br>apple<br>cable<br>cup<br>soup<br>meat<br>bread<br>screen<br>shoe |