Understanding the Cognitive Processes Involved in Writing to Learn

Article *in* Journal of Experimental Psychology Applied · April 2017

DOI: 10.1037/xap0000119

Mark A. Mcdaniel
Washington University in St. Louis

Elizabeth J Marsh
Duke University

Journal of Experimental Psychology: Applied
2017, Vol. 23, No. 2, 115–127

1076-898X/17/$12.00 http://dx.doi.org/10.1037/xap0000119

Understanding the Cognitive Processes Involved in Writing to Learn

Kathleen M. Arnold
Duke University

Sharda Umanath
Claremont McKenna College

Kara Thio
Duke University

Walter B. Reilly
University of California at Davis

Mark A. McDaniel
Washington University in St. Louis

Elizabeth J. Marsh
Duke University

Writing is often used as a tool for learning. However, empirical support for the benefits of writing-to-learn is mixed, likely because the literature conflates diverse activities (e.g., summaries, term papers) under the single umbrella of writing-to-learn. Following recent trends in the writing-to-learn literature, the authors focus on the underlying cognitive processes. They draw on the largely independent writing-to-learn and cognitive psychology learning literatures to identify important cognitive processes. The current experiment examines learning from 3 writing tasks (and 1 nonwriting control), with an emphasis on whether or not the tasks engaged retrieval. Tasks that engaged retrieval (essay writing and free recall) led to better final test performance than those that did not (note taking and highlighting). Individual differences in structure building (the ability to construct mental representations of narratives; Gernsbacher, Varner, & Faust, 1990) modified this effect; skilled structure builders benefited more from essay writing and free recall than did less skilled structure builders. Further, more essay-like responses led to better performance, implicating the importance of additional cognitive processes such as reorganization and elaboration. The results highlight how both task instructions and individual differences affect the cognitive processes involved when writing-to-learn, with consequences for the effectiveness of the learning strategy.

Keywords: writing-to-learn, essays, retrieval, cognitive processes, individual differences

Supplemental materials: http://dx.doi.org/10.1037/xap0000119.supp

Writing is a critical skill, but it can also be a potential tool for learning. Since the 1970s, a writing-to-learn literature has emerged suggesting that students can learn content, such as scientific theories and historical causes and effects, through writing about it (e.g., Bangert-Drowns, Hurley, & Wilkinson, 2004; Emig, 1977). This literature tradition has advocated the merits of writing activities as diverse as journal writing (Connor-Greene, 2000), note taking (Kiewra & Benton, 1988; McIntyre, 1992), summaries (Friend, 2002; Radmacher & Latosi-Sawin, 1995), short essays (Marshall, 1987; Voss & Wiley, 1997), term papers (Nicotera, Shibley, & Milakofsky, 2001), and "mini-writing" in large classes (Gingerich et al., 2014). Less clear, however, is whether these recommendations have solid empirical support. Many of these claims depend upon case studies and observation or use small sample sizes (e.g., Newell, 1984; Newell & Winograd, 1995; see Humes, 1983, for a review of methods), and the work is often situated within specific academic disciplines. Furthermore, some researchers have reached the opposite conclusion, finding writing to be no better or, at times, even less effective than other learning strategies (Penrose, 1992; Spirgel & Delaney, 2016).

This confusing set of findings is in part due to the conflating of so many diverse activities all under the single umbrella of writing-to-learn. Writing activities can differ in many ways; they can be high or low stakes, short or long, single-shot attempts or repeatedly revised, about a single topic or more integrative—among many other differences (see Bangert-Drowns et al., 2004 and Hebert, Gillespie, & Graham, 2013, for meta-analyses exploring some factors that moderate writing-to-learn). The impact these differences may have on learning are not always taken into account. For

This article was published Online First April 27, 2017.

Kathleen M. Arnold, Department of Psychology and Neuroscience, Duke University; Sharda Umanath, Department of Psychology, Claremont McKenna College; Kara Thio, Office of Research Support, Duke University; Walter B. Reilly, Center for Neuroscience, Department of Psychology, University of California at Davis; Mark A. McDaniel, Department of Psychological & Brain Sciences, Washington University in St. Louis; Elizabeth J. Marsh, Department of Psychology and Neuroscience, Duke University.

This research was supported by Grant R305A130535 to Duke University from the Institute of Education Sciences, U.S. Department of Education. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Kathleen M. Arnold, Department of Psychology & Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27708. E-mail: kathleen.arnold@duke.edu

example, consider how cognitive psychologists sometimes conflate free recall, a common laboratory task that consists of remembering everything one can, with essays, asserting that free recall is reflective of "typical classroom assignments" such as "brief essays" (Glover, 1989, p. 398; see also Schmidt, 1983 and Roediger & Karpicke, 2006a, for similar arguments). However, although free recall is clearly an effective mnemonic device (e.g., Roediger & Butler, 2011; Roediger & Karpicke, 2006b), remembering as much of a text as is possible differs from writing an essay about that text. Imagine the educational analog: free recall would have the student writing down as much of the assigned reading as possible (without concern for coherence, deeper meaning, or the length of assigned readings). In contrast, writing an essay would involve selecting the relevant facts while excluding the irrelevant and organizing the information into a narrative structure (Hayes & Flower, 1986).

In other words, different writing tasks will engender different processing, which will in turn have different consequences for learning and retention. The writing-to-learn literature has made this point before, noting the importance of focusing on cognitive processes, rather than on tasks. For example, in an early review, Applebee (1984) concluded that different tasks afford different levels of processing, with tasks that promote deep processing, such as essay writing, more likely to enhance learning than tasks that lead to more superficial processing, such as note taking. However, although this work did push the field in the direction of processes, "deep processing" is too vague to be useful, and often leads to circular conclusions—a task involves deep processing when it leads to good learning, and it leads to good learning when it involves deep processing (Nelson, 1977).

A more recent review (Klein, 1999) suggested that the inconsistent results from the writing-to-learn literature result from asking the wrong question; researchers should not ask if writing promotes learning, but how it does so by focusing on what processes writing engages. Since this review, at least some writing-to-learn researchers have heeded this call and focused on the use of particular cognitive processes (e.g., elaboration) as strategies for improving learning from writing (e.g., Glogger, Schwonke, Holzapfel, Nückles, & Renkl, 2012; Martínez, Mateos, Martín, & Rijlaarsdam, 2015), an important applied goal.

This primary focus on applied goals sets the writing-to-learn literature apart from the related cognitive psychology literature on learning, which generally begins with a more theoretical rather than an applied approach. Although both traditions have explored theoretical and applied questions about learning, they have developed largely independently resulting in different, yet at times parallel, views on learning. Both approaches have their strengths but either in isolation can lead researchers to miss important insights from the other perspective. The present work attempts to bridge these two traditions, bringing together evidence from both literatures and identifying important gaps in our understanding of the role of cognitive processes in learning from writing.

Although the cognitive psychology and writing-to-learn traditions have largely developed independently, they have converged in identifying the importance of two cognitive processes: elaboration and organization. *Elaboration* involves connecting what one is learning to what one already knows, to one's personal experiences, and so forth (Weinstein & Mayer, 1986; Stein, Littlefield, Bransford, & Persampieri, 1984). In cognitive psychology, this may involve putting meaning on something meaningless (for example, remember RFK—1120 as Robert F. Kennedy—November 20th [his birthday]); in the writing-to-learn literature, this may involve comparing the to-be-learned content to another construct (e.g., How does mitosis compare to meiosis?). *Organization*, on the other hand, typically involves structuring the to-be-learned material often in a way that creates a new structure (i.e., reorganization). A list of randomized category exemplars is later remembered with similar category exemplars together (Bousfield, 1953; clustering), and people benefit from being told to identify headings and subheadings (Grant, 1993). The two traditions often differ in their empirical approaches; for example, one writing-to-learn approach involves engineering writing assignments that will increase the use and quality of these processes, via the particular prompts given (e.g., Schwonke, Hauser, Nückles, & Renkl, 2006). Another approach is to target the individual, training him/her to use these strategies, with consequent benefits to learning (e.g., Martínez, Mateos, Martín, & Rijlaarsdam, 2015). In contrast, cognitive psychology support for these processes generally comes from experimental work unrelated to writing (e.g., Carpenter & DeLosh, 2006; Einstein, McDaniel, Bowers, & Stevens, 1984; McDaniel & Donnelly, 1996).

Neither tradition places much emphasis on what is known as "rehearsal" in the writing-to-learn literature or copying in the cognitive literature.[1] That is, the physical act of writing something is not necessarily beneficial to learning; the argument is that one has to transform rather than simply repeat the information (Bereiter & Scardamalia, 1987). Similarly, *copying* is a low-level skill that does not require meaning extraction. However, here is where the cognitive literature offers a perspective that is missing from the writing-to-learn literature; relatively verbatim copying can be very beneficial if and only if people are doing so from memory. Dozens, if not hundreds, of studies have shown that retrieving information from memory boosts retention (Roediger & Karpicke, 2006a) and transfer of knowledge (Butler, 2010; Carpenter, 2012), and these effects have been demonstrated in classrooms as well as in the laboratory (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013). Importantly, retrieval is more than simple rehearsal or reexposure to the material; retrieval enhances memory and learning over and above restudying, an effect that is especially robust on delayed tests (Roediger & Karpicke, 2006b).

Table 1 lays out one conceptualization of how three different writing tasks may afford these key cognitive processes (elaboration, reorganization, and retrieval). To be clear, we are not advocating that these tasks are process pure, as the actual processes involved will differ depending upon the exact instructions, the student's ability, and so on—a point we will return to later in this article. Further, the processes themselves may overlap. For example, one theory suggests that retrieval processing benefits learning at least in part by activating elaborative processing (Carpenter, 2009). Tasks provide a starting point for our discussion; we choose

[1] The term *rehearsal* also has a long history in cognitive psychology, but typically refers to the effect of repeated verbatim repetition (rote rehearsal) rather than the effect of physically copying stimuli (see Benjamin & Bjork, 2000).

Table 1
Key Cognitive Processes Likely Engaged by the Four Learning Activities Used in the Current Experiment Given the Instructions and Procedures Used

Learning activities	Writing task	Cognitive processes		
		Reorganization	Elaboration	Retrieval
Highlighting	No	No	No	No
Note taking	Yes	Unlikely	Unlikely	No
Free recall	Yes	Unlikely	Unlikely	Yes
Essay	Yes	Yes	Yes	Yes

note taking and essay writing as they are both commonly used in authentic educational contexts (Hounsell, 1997; Palmatier & Bennett, 1974), and free recall as it is commonly used as a laboratory analog for writing tasks (e.g., Glover, 1989). Our experimental design also included a highlighting condition as a commonly used nonwriting study activity (Gurung, Weidert, & Jeske, 2010; Kornell & Bjork, 2007) with which to compare the effectiveness of the writing conditions.

The first two processes listed in Table 1, reorganization and elaboration, likely often co-occur in writing tasks. By definition, these processes are not part of highlighting. Although they can be involved in note taking (Einstein, Morris, & Smith, 1985) and free recall (Zaromb & Roediger, 2010), we argue that for many students (as developed below, individual differences likely come into play here), these processes are unlikely to be engaged especially when the information is already presented in an organized fashion (McDaniel & Einstein, 1989). In contrast, essays often involve reorganization of learned material and elaborating on that material (Wiley & Voss, 1999). That is, to write an essay, learners should organize the information into a new structure (i.e., reorganize) and write about the information (i.e., elaborate) to fit it within the new context and provide support for their thesis. In this sense, a good essay is like a good story, in which a writer takes a perspective on events and structures her account depending on her goals. Of course, not all students are successful at writing good essays (e.g., Torrance, Thomas, & Robinson, 2000), and student approaches to the same task can vary (e.g., Segev-Miller, 2007). Therefore, the extent to which essay-writing engages these processes will certainly vary (see discussion on individual differences below).

The last process listed in Table 1, retrieval, is a component of writing activities that require students to write without access to other materials such as books or one's notes. Assuming that free recall and essay writing both take place in a closed-book environment in which learners are unable to refer to the source material while writing, these two writing activities engage retrieval processes. In contrast, highlighting and note taking are presumed to occur (and in the former case, must occur) while learners have access to the material. Of course, under different conditions, such as writing with access to the original materials (open-book conditions), the degree to which these tasks would engage retrieval processes would differ. This variability in situational demands again highlights the importance of focusing on processes rather than on tasks; the same writing activity can engage different processes depending on the particular instantiation of the task.

In the present experiment, participants studied two technical passages on astronomy in concert with one of the four assigned learning activities (highlighting, note taking, free recall, essay writing). Two days later, participants were given a multiple-choice test on factual information in each passage and on information that could have been inferred from the passages (but was not directly stated). They also responded to several short answer questions requiring problem solving based on the passage content. Based on the foregoing theoretical analysis, we formulated two main predictions. First, given the established benefits of retrieval for retention and transfer, we posited that the two writing activities that required retrieval (free recall and essay writing) would produce better retention of the factual information and more flexible use of the acquired content (inferencing and problem solving) than would the two activities that did not require retrieval. Second, from the perspective that elaboration and reorganization promote learning and retention, we examined whether essay writing would produce additional benefits on retention and problem solving over that of free recall, given that essays are more likely to stimulate elaboration and organization. Finally, because of the concern that tasks do not always map perfectly onto processes, we note here that the analysis section describes how we measured actual retrieval and organization in writing responses.

Individual Differences

A second major focus of this study was to consider the moderating influence of several individual differences that might influence the cognitive processes involved in the assigned writing activities (see Table 1). That is, learners vary in ability, which may affect the degree to which the learning activities stimulate the use and/or effectiveness of cognitive processes. For example, writers use different strategies (which in turn vary in effectiveness) when approaching writing assignments (e.g., Torrance, Thomas, & Robinson, 2000; Segev-Miller, 2007). This variability provides a second way to examine the role of cognitive processes in learning from writing; that is, we examined people's written products for evidence of particular cognitive processes and in turn linked that to learning. As the use of the cognitive processes differs, so should the effectiveness of the writing assignment. As explained below, we chose three individual differences that we suspected might bear on the processes involved in the writing conditions implemented in the experiment: working memory capacity, writing ability, and structure-building ability.

Working memory describes one's ability to remember and manipulate information in the face of distraction (Engle, 2002). At the most general level, working memory is known to predict many skills involved in the activities faced by participants in our experiment, including reading comprehension (Daneman & Carpenter, 1980), writing (McCutchen, 1996; Swanson & Berninger, 1996), and reasoning ability (Conway, Kane, & Engle, 2003). There are two reasons why working memory capacity might moderate the effects of writing activity. First, given that working memory is associated with greater retrieval from long-term memory (Unsworth & Engle, 2007; for reasons why, see Kane & Engle, 2000 and Unsworth, Brewer, & Spillers, 2012), working memory should matter most when the learning activities involved retrieval (essay, free recall). However, working memory may also facilitate one's ability to effectively reorganize and elaborate on material as one writes an essay. To the extent that this matters, we would expect that the advantage of essay writing relative to free recall would be most evident with increasing working memory.

We assessed *writing ability* with SAT writing and ACT English/writing combined scores, which are standardized instruments typically used to measure writing ability (Kellogg, 2001; Mattern, Camara, & Kobrin, 2007). They are validated by their predictive utility for overall first year GPA and grades in English composition courses (Norris, Oppler, Kuang, Day, & Adams, 2006). We speculated that better writing ability could be relatively important for the successful incorporation of elaboration and reorganization into essay writing. Thus, better writing ability might be associated with a more robust advantage for essay writing relative to free recall.

Structure building is an individual difference in the ability to construct coherent and organized mental representations of experienced events and texts (Callender & McDaniel, 2007; Gernsbacher et al., 1990). Structure building is related to reading comprehension (it correlates modestly with standard reading assessments; Maki, Jonas, & Kallod, 1994) but is viewed as a process that is more overarching than reading alone. Supporting this view, interventions that assist learners in building better mental representations enhance learning for low structure builders, but not for low reading comprehenders. When a story was presented with sentences in a random order and readers were required to rearrange the sentences to create a coherent story, recall for the story improved for low structure builders (relative to when the story was presented in normal order) but not for low reading comprehenders (McDaniel, Hines, & Guynn, 2002). Rearranging the sentences forced low structure builders to focus on the structure, helping them build a more coherent mental model then they would have otherwise.

Further separating this construct from reading comprehension, structure building relates to a more general comprehension ability that cuts across modalities; whether information is presented through text, spoken words, or pictures, structure building captures one's ability to build a mental model (Gernsbacher, Varner, & Faust, 1990). Indeed, structure-building ability is related to performance in college courses, a context in which information is presented in multiple modalities through readings, lectures, videos, and so forth (Arnold, Daniel, Jensen, McDaniel, & Marsh, 2016; Maki & Maki, 2002).

Most importantly for present purposes, poor structure-building skills are assumed to result in fragmented representations of texts, which in turn reduce learners' ability to remember information from a text, including academically authentic texts (Callender & McDaniel, 2007; Callender & McDaniel, 2009, Experiment 4; Martin, Nguyen, & McDaniel, 2016). Based on this literature, we anticipated that the benefit for writing activities that required retrieval (essay writing, free recall) would be especially pronounced for learners with greater structure-building ability because those learners would be able to support the retrieval necessary for composing an essay or producing free recall. In contrast, essay writing and free recall would not be favored for learners with lower structure-building ability (relative to note taking or highlighting) because these learners would struggle to produce reasonable levels of recall or retrieve information to compose an essay.

In short, the current study involved a comparison of the four learning activities listed in Table 1: highlighting, note taking, free recall, and essay writing. We investigated how completing these different activities, which engage different cognitive processes as suggested in Table 1, affected learning, retention, and transfer of science concepts on a test two days after the initial learning phase. We further examined how, regardless of the assigned task, individual differences in the use of cognitive processes affected learning. We investigated if the use of these processes was modified by three individual differences: working memory, measured with operation span and reading span tasks; writing ability, as operationalized by SAT and ACT writing scores; and structure-building ability as determined by performance on the Multi-Media Comprehension Battery (MMCB; Gernsbacher & Verner, 1988). Overall, the goal was to shed light on writing as a learning tool, by focusing on the underlying cognitive processes and linking them to student learning.

Method

Participants and Design

One hundred Washington University undergraduates participated in exchange for either partial fulfillment of a class requirement or $25. Participants were randomly assigned to one of four between-subjects conditions: writing essays, recalling the passages, note taking, or highlighting the passages. One participant in the highlighting condition was excluded for not following directions. Three dependent variables were measured—performance on factual multiple-choice questions, inference multiple-choice questions, and problem-solving short answer questions.

Materials

Passages. Two astronomy passages were used: one on detecting extraterrestrial life and one on solar activity.[2] The passages were created using information from an undergraduate level astronomy textbook (Karttunen, Kröger, Oja, Poutanen, & Donner, 2006). The detecting life passage was 928 words long and was written at an 11.8 Flesch-Kincaid grade level. The solar activity passage was 810 words long and was written at a 9.9 Flesch-Kincaid grade level.

Test questions. For each passage, four factual multiple-choice, four inference multiple-choice, and four problem-solving short answer questions were constructed. Factual multiple-choice questions could be answered using information explicitly stated in the passage (e.g., "What is the size of the magnetic fields in sunspots?"). Inference multiple-choice questions required participants to integrate across or extrapolate from two stated facts (e.g., "Based on their visual characteristics, with regards to temperature, the relationship between that of prominences and sunspots would best be described as which of the following?").

For each set of problem-solving short answer questions, participants were asked to imagine a scenario prior to solving the problems. For the problems about solar activity, participants were instructed to answer as if they were astronomers who liked to watch solar activity from their backyards. For problems about detecting life in outer space, participants were instructed to answer as if they were researchers who strongly believed that there is life

[2] Passages are available upon request from Kathleen M. Arnold.

in outer space. Each question required participants to solve a problem by making inferences and drawing connections across several facts from the passage; for example,

> You want to show a friend solar activity in the sky, but you do not have access to a telescope at the moment. Which of these solar activities (sunspots, faculae, eruptive prominences, solar flares) would you be most likely to be able to see? Please give two reasons to explain your answer.

Individual difference measures. Working memory was measured using two computerized tasks: the automated operation span task (OSpan) and the automated reading span task (RSpan; Unsworth, Heitz, Schrock, & Engle, 2005). In both tasks, participants had to remember a series of three to seven letters, presented one at a time. In the OSpan task, in between the presentation of each letter, participants were shown a math problem—for example, $(3 \times 4)+6$—followed by a number (e.g., 19) and had to respond "true" or "false" to indicate if the number matched the correct solution to the problem. In the RSpan task, between presentations of letters, participants read sentences (e.g., "During the week of final spaghetti, I felt like I was losing my mind") and had to indicate whether or not the sentence made sense.

Writing ability was measured via SAT writing scores and/or ACT English/writing composite scores for participants who consented to release their scores ($n = 81$). Some participants had taken only the SAT ($n = 37$) or only the ACT ($n = 26$), whereas some participants took both tests ($n = 18$). To put the scores on the same metric, ACT scores were converted to SAT scores using a guide from The College Board (2009). After converting ACT scores, SAT scores ranged from 590 to 800, with an average score of 707.9.

Following previous studies, the reading portion of the MMCB (Gernsbacher & Verner, 1988) was used to measure structure-building ability (Arnold et al., 2016; Bui & McDaniel, 2015; Callender & McDaniel, 2007; Callender & McDaniel, 2009; Martin, Nguyen, & McDaniel, 2016). This measure consists of four narratives ranging from 538 to 957 words, each of which has 12 corresponding multiple-choice questions asking about key details from the story. Because of a coding error in the program, one multiple-choice question had to be excluded, resulting in a maximum possible score of 47.

Procedure

The experiment consisted of two sessions. In the first session, participants were told they would read scientific passages and complete a learning activity (essay, free recall, note taking, or highlighting, depending on condition) to help them learn the passages. They were told to learn the information as well as they could as they would be asked questions about the passages when they returned for the second session. All subjects learned about detecting life in outer space before completing the same learning activity on a passage on solar activity.

Participants received specific instructions about their learning strategy prior to reading. In the highlighting condition, participants received a highlighter and a paper copy of the text. They were told to highlight sections from the passage as they would if they were reading a textbook and trying to learn the information for a class. In the note-taking condition, participants were told to take notes on the passage like they would for a class; they received a paper copy of the text and typed their notes on the computer.

Before participants read the passage, they were told they would have to write an essay or recall the passage from memory (depending on condition) after reading it. After they finished reading, they returned the passage to the experimenter and received further instructions. In both conditions, participants were told to type their responses on the computer. Participants in the free recall condition were instructed to recall everything they could from the passage. Participants in the essay condition wrote a response to the following prompt:

> Write an essay describing the indicators of life that may be used to detect other intelligent civilizations and how we have attempted to communicate with these possible civilizations. Be as clear, detailed, and thorough as possible so that a high school student who has not read the text could understand. Your essay should have an introduction and a clear thesis, and you should make sure to back up your points with supporting details.

After reading and completing the learning activity for the detecting life passage, participants read and completed the learning activity for the solar activity passage in the same manner. That is, participants highlighted or took notes on the passage as if they were preparing for a class, recalled the passage from memory, or wrote an essay (from memory) in response to the following prompt:

> Write an essay describing the different types of solar activity, including their properties, their relationships with one another, and their effects on Earth. Be as clear, detailed, and thorough as possible so that a high school student who has not read the text could understand. Your essay should have an introduction and a clear thesis, and you should make sure to back up your points with supporting details.

In all four conditions, reading and learning activities were self-paced,[3] and no feedback was given after learning activities were completed. After finishing both passages, participants completed the OSpan task. This session lasted about 1 hour.

Session 2 occurred 2 days after the first session. It began with participants answering the multiple-choice and problem-solving questions for the detecting life passage followed by the multiple-choice and problem-solving questions for the solar activity passage. Following these questions, participants completed the MMCB and the RSpan task. For the MMCB, participants read

[3] The tasks were self-paced to reflect how they would typically be instantiated in the real world to prevent artificial constraints on the participants. Thus, it is not surprising that the time spent learning varied across conditions, $F(3, 95) = 31.49$, $p < .05$, $\eta^2 = .13$. Participants in the essay condition ($M = 22.95$ min) spent more time on task than participants in all other conditions: recall, $M = 14.21$; note-taking, $M = 12.73$; highlighting, $M = 6.49$; smallest, $t(48) = 4.50$, $p < .001$, $d = 1.30$. From an applied perspective, equating time on task does not make sense; a quality essay would not be written in the time it takes to highlight, and forcing readers to spend more time highlighting will not improve their processing. In fact, spending more time highlighting was not associated with better performance on either the multiple-choice, $r = .18$, $p = .39$, or problem-solving questions, $r = .15$, $p = .49$. Further, in their meta-analysis on school based writing-to-learn interventions, Bangert-Drowns et al. (2004) found that spending more time writing led to smaller effects on learning. Much research has shown that it is not the quantity of time but how time is used that is key (e.g., Craik & Lockhart, 1972).

each of the four narratives at their own pace. One to two paragraphs were presented on the screen at a time, and when ready, participants pressed a button on the keyboard to move on to the next paragraph(s). Following each narrative, participants answered the corresponding multiple-choice questions. Finally, participants were given a form to fill out if they consented to releasing their SAT/ACT scores. This session lasted approximately 1.5 hours.

Results

Scoring Final Learning

For the test responses, only the problem-solving short answer questions required scoring by judges; multiple-choice answers were auto-scored. A scoring rubric was created that awarded each problem a score from 0 to a maximum possible score of 2–4 (depending on the question), and the total number of earned points was summed. Performance was measured as the proportion of total possible points earned. Two independent coders scored responses (Cohen's $\kappa = .78$), and discrepancies were resolved through discussion.

Analytic Approach

To preview, for each dependent measure we first report a one-way analysis of variance comparing all four conditions. We then conduct more focused contrasts that are structured to answer two main questions. One contrast collapses over conditions to focus on the process of retrieval. That is, we analyze performance as a function of retrieval, collapsing essay writing and free recall conditions, which both required retrieval, and comparing them to note-taking and highlighting conditions, neither of which required retrieval. A second contrast examines any added benefit from processes like reorganization and elaboration, over and above retrieval. This analysis compares the benefits of essay writing, which likely engages all of the Table 1 processes (reorganization, elaboration, and retrieval), to free recall, which as developed above is a writing activity that involves a retrieval but not necessarily an elaboration/reorganization component.

The next set of analyses focuses on the individual difference patterns, especially as they relate to the effects of the learning activities. For each of our focused questions (effect of retrieval, effect of reorganization and elaboration), we conduct hierarchical regression analyses to examine how individual differences affected performance and whether or not learning activity conditions modified these effects.

Finally, we compared the essay and free recall responses for content and structure. The goals of these final analyses were to determine how, if at all, participants responded differently to the essay and free recall prompts, and, if differences were found, how they affected final test performance. These analyses test our assumption that the different prompts encouraged the engagement of different cognitive processes.

Learning-Condition Effects

Multiple-choice questions. Comparing all four learning activities without collapsing across conditions revealed no significant effects of learning activity or of multiple-choice item type (fact, inference) and no interaction—learning activity: $F(3, 95) = 1.47$, $p = .23$, $\eta^2 = .04$; item type: $F < 1$; interaction: $F < 1$; see Table 2.

To inform our more focused theoretical questions, we next conducted two between-groups contrasts (that included type of multiple-choice item as a within-subjects variable). The first indicated that the learning activities that required retrieval of information from memory (essay and free recall) led to significantly better performance on multiple-choice questions than the two tasks that did not require retrieval (note taking and highlighting) with no difference across multiple-choice item type—$M = .64$ versus .57; learning activity: $F(1, 97) = 4.29$, $p = .04$, $\eta^2 = .04$; item type: $F < 1$; interaction: $F < 1$. The second focused contrast revealed no benefit of writing an essay over free recalling the passage across either item type, $M = .65$ versus .63; all Fs < 1.

Problem-solving questions. Performance on the problem-solving short answer questions did vary across learning activities, $F(3, 95) = 2.75$, $p = .047$, $\eta^2 = .08$ (see Figure 1). Post hoc follow-up tests showed that participants in the essay condition performed significantly better on the problem-solving questions than those in the highlighting condition, $t(47) = 2.60$, $p = .01$, $d = .76$, with the difference approaching a large effect size. No other pairwise comparisons were significant.

The planned contrasts revealed that when retrieval was required (essay and free recall), performance was significantly better on the problem-solving questions than when it was not required (note-taking and highlighting)—$M = .38$ versus .32; $t(97) = 2.17$, $p = .03$, $d = .44$. However, writing an essay afforded no significant advantage over free recall—$M = .41$ versus .36; $t(48) = 1.02$, $p = .31$, $d = .29$.

In summary, similar patterns were obtained across all final test measures: Learning activities involving retrieval led to better learning of the material, whereas there were no additional (significant) benefits of reorganization and elaboration presumably conferred by essay writing.

Individual Difference Analyses

Retrieval component. In general, test performance was better following learning activities that involved retrieval, but was this true for all participants? To answer this question, we conducted two hierarchical regression analyses, with proportion correct on the multiple-choice and problem-solving questions as the two dependent variables (see Tables 4 and 5). In Step 1, we tested for main effects of three individual difference measures: working

Table 2
Mean Proportion Correct on Final Multiple-Choice Questions (Standard Deviations in Parentheses)

Learning activity	Multiple-choice question type: Fact	Inference	Total
Essay	.64 (.14)	.65 (.23)	.65 (.15)
Free recall	.66 (.24)	.60 (.24)	.63 (.18)
Note-taking	.58 (.16)	.55 (.22)	.57 (.15)
Highlighting	.57 (.17)	.58 (.25)	.58 (.17)
Total	.61 (.18)	.60 (.23)	.60 (.16)

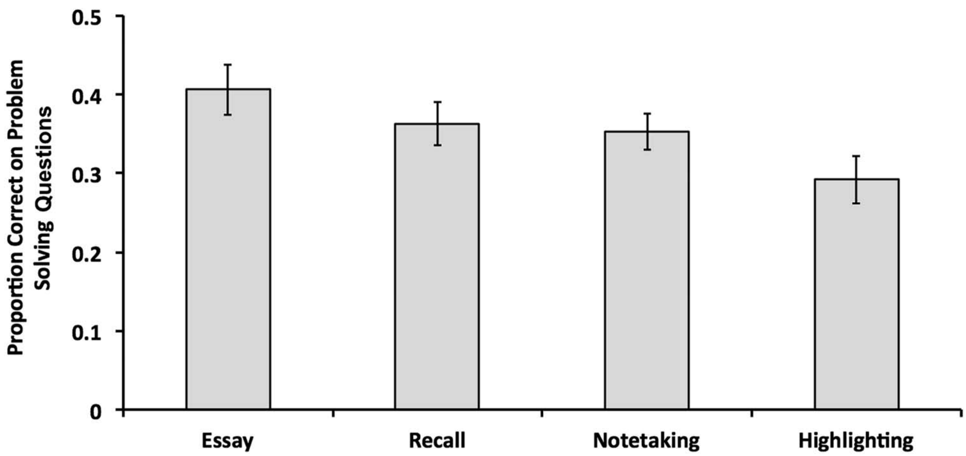

Figure 1. Proportion correct on problem-solving questions as a function of learning activity. Error bars represent standard error of the mean.

memory capacity (an average of RSpan and OSpan scores), SAT writing scores, and structure building (all mean-centered; see Table 3 for descriptive data). Here we found only a main effect of structure building; higher MMCB scores predicted better performance. In Step 2, we added the focused learning activity comparison (learning activities requiring retrieval vs. those that did not) to the model and found that it explained a significant additional proportion of the variance. When controlling for individual differences, participants in learning activity conditions with retrieval still performed better than those in learning activity conditions without retrieval.

In Step 3, we asked whether the benefit of engaging in learning activities with retrieval varied as a function of scores on any of the individual difference measures. As Table 5 shows, adding these interaction terms explained a significant additional proportion of the variance for problem-solving performance. Specifically, we found that the benefits of learning activities with retrieval did vary as a function of structure building, but did not depend upon working memory capacity nor upon SAT scores. That is, the interactions were nonsignificant for SAT and working memory. In contrast, the effect of the learning activities was modified by structure-building ability (see Figure 2). Specifically, the benefits of the learning activities requiring retrieval (essay writing, free recall) relative to those not requiring retrieval (note taking, highlighting) were most prominent for learners with greater structure-building ability. In contrast, as structure-building ability declined, the advantage of the learning activities requiring retrieval relative to learning activities that did not involve retrieval was reduced or eliminated. As a consequence, engaging in retrieval was more beneficial to participants with greater structure-building ability. This pattern was numerically replicated in the multiple-choice data, although the interaction was not significant (see Table 4).[4]

Table 3
Descriptive Data of Individual Difference Measures

Individual difference measure	*M*	*SD*
Working memory	62.66	9.79
SAT writing score	707.90	52.03
MMCB	31.87	6.47

Note. Working Memory scores are the average of automated RSpan and OSpan total scores and have a maximum score of 75. Average SAT writing scores include ACT English/writing composite scores that have been converted to SAT scores and have a maximum score of 800. Multi-Media Comprehension Battery (MMCB) has a maximum of 47.

Reorganization and elaboration components. Next, we examined if individual differences modified the additional processes presumably engaged in essay writing but not in free recall (reorganization, elaboration). Although a significant essay advantage was not found in the prior analyses, examining individual differences may allow such an advantage to emerge. Specifically, we hypothesized that an essay advantage may only be present in individuals with high working memory capacity and/or with high writing ability. Having a larger working memory capacity and/or being a more skilled writer may increase a learner's ability to reorganize and/or elaborate on the material when writing an essay, thus increasing the benefit of engaging in this more complex task.

To examine these hypotheses, we again conducted two hierarchical regression analyses with the same independent variables entered into the same steps as before, with the difference that learning activities were now dummy coded as essay (1) or free recall (0). In both analyses, the Step 1 model (including only the individual difference variables) explained a significant proportion of the variance—multiple-choice: $R^2_{\text{adjusted}} = .35$, $F(3, 38) = 8.19$, $p < .001$, $\eta^2 = .39$; problem-solving: $R^2_{\text{adjusted}} = .39$, $F(3, 38) = 9.69$, $p < .001$, $\eta^2 = .43$. As before, the only individual difference that predicted performance was structure building [multiple-choice: $\beta = .57$, semipartial $r = .54$, $t(38) = 4.26$, $p < .001$; problem-solving: $\beta = .59$, semipartial $r = .56$, $t(38) = 4.56$, $p < .001$]. Learners with greater structure building ability did better on both multiple-choice and problem-solving questions.

Step 2 revealed no advantage of writing essays over free recalling the passages; adding the learning activity (essay vs. recall) to

[4] This interaction was replicated in a pilot study with 30 participants that included only the essay and highlighting conditions and experimentally controlled time on task, $\beta = .40$, semipartial $r = .32$, $t(26) = 2.16$, $p = .04$.

Table 4
Regression Results Predicting Multiple-Choice Performance Using Individual Difference Measures and Learning Activity Divided by Retrieval Processes

Predictor variables	F	Adjusted R^2	ΔR^2	β	Semipartial r
Step 1	6.68***	.18	.21***		
WM				.05	.04
SAT				.09	.09
MMCB				.43***	.42***
Step 2	7.63***	.26	.08**		
WM				.05	.04
SAT				.10	.10
MMCB				.41***	.40***
LA				.29**	.29**
Step 3	5.07***	.27	.04		
WM				−.01	−.01
SAT				.11	.07
MMCB				.27*	.20*
LA				.28**	.28**
WM × LA				.08	.06
SAT × LA				−.04	−.03
MMCB × LA				.23^	.17^

Note. $N = 78$. Learning activities (LAs) were dummy coded: 1 = activities with retrieval (essay, recall); 0 = activities without retrieval (note taking, highlighting); Individual difference measures were mean-centered. MMCB = multi-media comprehension battery; WM = working memory. ^ $p = .08$. ** $p < .01$. *** $p < .001$.

the models did not significantly increase the predictive values (multiple-choice: $\Delta R^2 = .01$, $F < 1$; problem-solving: $\Delta R^2 = .002$, $F < 1$). Further, the effect of the learning activity was not modified by any of the individual differences; adding the interactions in step 3 also failed to increase the predictive values (multiple-choice: $\Delta R^2 = .02$, $F < 1$; problem-solving: $\Delta R^2 = .005$, $F < 1$). In contrast to our predictions, higher working memory capacity learners and learners with better writing skills did not benefit more from writing essays than from free recalling the passages.

Why Was Learning Equivalent in the Essay and Recall Conditions?

Despite the presumption that essay writing should engage additional beneficial cognitive processes (see Table 1), performance did not statistically differ between this condition and the free recall condition. Further, learners with high structure building skills, high working memory capacity, and good writing skills seemed to benefit similarly from both free recall and essay writing. To better understand these patterns, we analyzed the essay and free recall responses for content and structure and then linked those properties to test performance. Our goal was to see whether essays were essentially the same as free recall (as argued by some experimental psychologists) or whether participants in these two conditions did equally well on the test for different reasons.

Content. In general, essays were longer than free recall responses—$M = 311.3$ vs. 219.1 words; $t(48) = 3.14$, $p < .003$, $d = .91$. However, essays are likely to contain words not directly related to the original passages, including introductory and concluding comments, transition phrases, and elaborations. To get a more precise estimate of the amount of original passage content included in the responses, we measured content in two ways. First, we examined the proportion of content words from the original texts included in the recall and essay responses using Linguistic Inquiry and Word Count (LIWC) software (Pennebaker, Booth, & Francis, 2007). Content words were defined as nouns, action verbs, and most adjectives from the original texts (excluding determiner adjectives [e.g., a, this, few, which] not referring to a specific quantity [e.g., 21 cm]). Although this method of scoring content is thorough in that it includes all passage content words, it fails to measure how these content words are used. That is, it does not distinguish between content words used correctly versus incorrectly, nor does it measure what idea units or facts from the passage are included in the responses. For this reason, we also used a second measure of content.

For our second measure, we identified passage facts needed to answer the multiple-choice and problem-solving questions and examined what proportion of these facts was included in the responses. Facts were limited to just those needed to answer the questions rather than all facts from the passage to target only those facts that could affect our dependent variables. Two independent coders scored all essays and free recall responses (Cohen's $\kappa = .71$), and discrepancies were resolved through discussion.

The two measures of content were highly correlated, $r = .88$, $p < .001$. For both measures of content, there was no difference between essays (content words: $M = .20$, $SD = .05$; passage facts: $M = .26$, $SD = .11$) and free recall (content words: $M = .19$, $SD = .07$; passage facts: $M = .30$, $SD = .14$) responses in the proportion of content words referenced (content words: $t < 1$; passage facts: $t(48) = 1.04$, $p = .31$, $d = .29$). Although participants in the essay condition wrote longer responses, essays and free recall responses contained similar amounts of passage information.

Where there was a difference, however, was among structure-building ability: Highly skilled structure builders included more

Table 5
Regression Results Predicting Problem-Solving Performance Using Individual Difference Measures and Learning Activity Divided by Retrieval Processes

Predictor variables	F	Adjusted R^2	ΔR^2	β	Semipartial r
Step 1	7.83***	.21	.24***		
WM				.14	.13
SAT				.11	.11
MMCB				.41***	.40***
Step 2	8.64***	.28	.08**		
WM				.13	.13
SAT				.12	.12
MMCB				.39***	.38***
LA				.28**	.28**
Step 3	6.73***	.34	.08*		
WM				.12	.09
SAT				.05	.03
MMCB				.17	.13
LA				.27**	.27**
WM × LA				−.02	−.01
SAT × LA				.09	.05
MMCB × LA				.36**	.27**

Note. $N = 78$; Learning activities (LAs) were dummy coded: 1 = activities with retrieval (essay, recall); 0 = activities without retrieval (note taking, highlighting); Individual difference measures were mean-centered. MMCB = multi-media comprehension battery; WM = working memory. * $p < .05$. ** $p < .01$. *** $p < .001$.

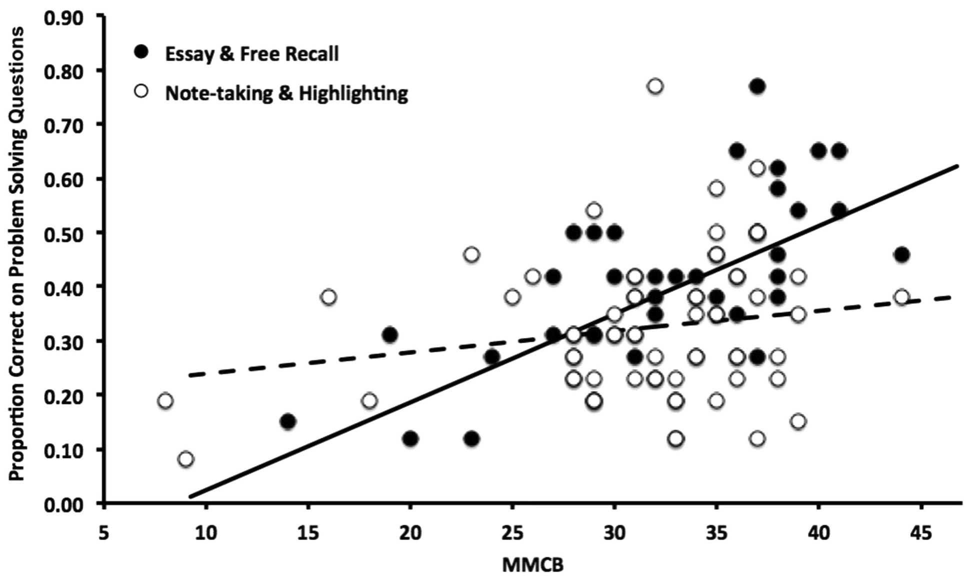

Figure 2. Proportion correct on problem-solving problems as a function of learning activity and scores on the Multi-Media Comprehension Battery. The lines are modeled from a regression equation with the solid line representing the essay and free recall conditions and the dotted line representing the highlighting and note-taking conditions.

content words in their writing than did less skilled structure builders. This pattern was confirmed with two regression analyses, one with each measure of content, which both revealed main effects of structure building—content words: $\beta = .54$, semipartial $r = .51$, $t(37) = 3.89$, $p < .001$; passage facts: $\beta = .53$, semipartial $r = .50$, $t(37) = 4.00$, $p < .001$—but no effect of SAT scores or working memory. The two analyses differed, however, in the effect of learning activity (essay vs. recall); learning activity did not predict the proportion of content words included but did predict the proportion of facts included with free recall predicting more included facts than the essay condition ($\beta = -.35$, semipartial $r = -.33$, $t(37) = -2.66$, $p = .01$). Finally, none of the individual differences interacted with the learning activity in either analysis.

This pattern could at least in part explain why those with greater structure-building ability benefited more from both essay writing and free recall; those with better structure-building skills retrieved more passage content on both types of responses, and including more passage content was associated with better test performance. Indeed, responses (both essay and free recall) that included more content words and passage facts were associated with better performance on both the multiple-choice—content words: $\beta = .51$, semipartial $r = .51$, $t(47) = 4.07$, $p < .001$; passage facts: $\beta = .46$, semipartial $r = .46$, $t(47) = 3.50$, $p = .001$—and problem-solving—content words: $\beta = .74$, semipartial $r = .74$, $t(47) = 7.63$, $p < .001$; passage facts: $\beta = .73$, semipartial $r = .72$, $t(47) = 7.27$, $p < .001$—questions. This pattern was similar in both the essay and free recall conditions, as the effect of content did not interact with learning activity in either analysis.

Structure. Good essays are generally structured in an organized manner with transitions that connect ideas with each other. Such a response likely requires the writer to reorganize and elaborate upon what is stored in memory, to make an argument complete with transitions and connections between ideas.

To measure the structure of the essay and free recall responses, we turned to crowdsourcing. Three hundred and nine “workers,” recruited using Amazon’s Mechanical Turk, each rated a random sample of 10 or 20 essay and free recall responses from participants, using a 1 (*very list-like* [defined as unconnected sentences or phrases that read like a list of facts]) to 5 (*very cohesive* [defined as connected and well organized sentences that read like an essay]). On average, each essay and free recall response was evaluated by 33.14 different people. Using a random sample of 28 ratings per response (the minimum number of ratings for any given response), a high degree of reliability was found. With a one-way random effects model, the average measure intraclass correlation (ICC) was .92 with a 95% confidence interval from .90 to .94, $F(99, 2700) = 13.00$, $p < .001$. For each written response, an average cohesion score was calculated, creating one cohesion score per passage per participant, which were then averaged together to create an overall cohesion score for each participant.

Overall, essays were rated as more cohesive ($M = 3.98$, $SD = .34$) than were free recall responses ($M = 3.07$, $SD = .69$; $t(48) = 5.91$, $p < .001$, $d = 1.71$); however, as indicated by the standard deviations and clearly depicted in Figure 3, there was more variability in the free recall condition. Some free recall responses were rated as cohesive as many of the essays.

A regression analysis revealed that learners with better writing skills wrote more cohesive responses, but this relationship was especially pronounced in the free recall condition. Overall, higher SAT writing scores predicted higher cohesion ratings—$\beta = .51$, semipartial $r = .25$, $t(35) = 2.33$, $p = .03$—but this effect differed depending on learning activity condition—$\beta = -.42$, semipartial $r = -.22$, $t(35) = -2.07$, $p = .046$. In the free recall condition, higher SAT scores were correlated with greater cohesion ratings, $r = .51$, $p = .03$, whereas in the essay condition, there was no relationship between these variables, $r = .04$, $p = .85$. These

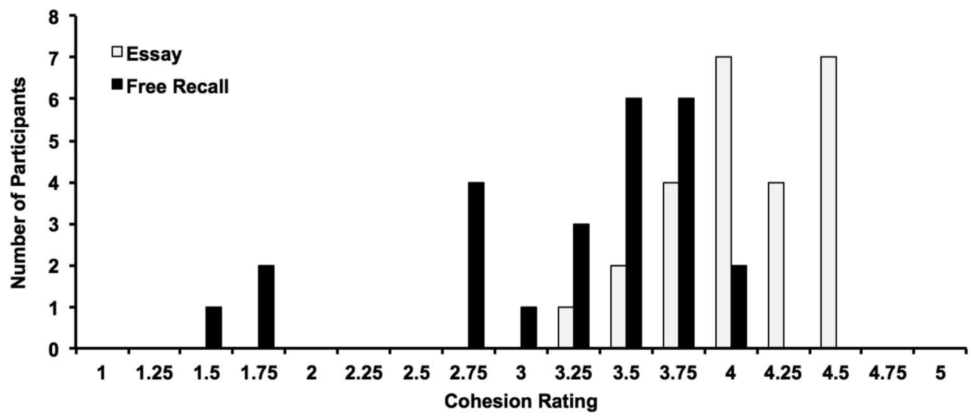

Figure 3. Histogram showing the number of participants with a given average cohesion rating in both the essay and free recall conditions. Each listed cohesion rating represents a range of ratings (e.g., 4.76–5), with the listed number indicating the highest cohesion rating within that range.

results suggest that, in the free recall condition, more skilled writers were more likely to choose to write an essay-like response, whereas in the essay condition, all learners attempted to write an essay-like response.

Given this possible interpretation, we asked whether response cohesion predicted later test performance regardless of what learning activity was assigned. The hierarchical regression analysis indicated that response cohesion predicted both problem-solving—β = .47, semipartial r = .47, $t(48) = 3.69$, $p = .001$—and multiple-choice performance—β = .31, semipartial r = .31, $t(48) = 2.28$, p = .03—and that adding the assigned learning activity condition to the equations did not significantly increase the predictive value of the models—problem-solving: $\Delta R^2 = .04$, $F(1, 47) = 2.81$, $p = .10$; multiple-choice: $\Delta R^2 = .04$, $F(1, 47) = 2.29$, $p = .14$. Further, adding an interaction term to the models also failed to increase the predictive value—problem-solving: $\Delta R^2 = .001$, $F < 1$; multiple-choice: ΔR^2 = .005, $F < 1$. These results suggest that later test performance was driven at least in part by the structure of participants' responses rather than the particular instructions they were given; more essay-like responses were associated with better performance on the later test, regardless of writing instruction.

Discussion

Our results clearly highlight the need to focus on cognitive processes when evaluating writing as a learning tool, in addition to considering different writing tasks. That is, our experiment allowed a direct comparison of two writing tasks, essays and free recall, and yet that distinction was much less important than the qualities of the writing produced. A comparison of learning in these two different instructional conditions masked important information because some participants, particularly those with better writing skills, wrote essay-like responses even when simply instructed to recall the passages. MTurk workers rated some free recall responses as just as "cohesive" and "essay-like" as those written in the essay condition, and this rating was more predictive of final test performance than which writing condition a participant had been assigned to. Writing an essay-like response likely required reorganization and elaboration, both processes known to improve memory, regardless of the particular instruction received.

Our conclusion that less proficient writers (in our sample, as measured by SAT writing scores) needed to be explicitly instructed to write essays was based on a relatively small sample size for individual difference analyses (n = 25 each in the free recall and essay conditions). To ensure that this finding was not a false positive, we collected additional data sampling participants at both Washington University and Duke University (free recall, n = 29; essay, n = 39; see the online supplementary materials for the full methods and results). This new data set confirmed our previous findings; lower-skilled writers again were less likely to write a cohesive free recall response, whereas writing skill did not affect cohesiveness in the essay condition (Learning Activity × SAT interaction: $\beta = -.32$, semipartial $r = -.20$, $t(64) = -2.2$, p = .03; free recall: $r = .42$, $p = .03$; essay: $r = .002$, $p = .99$).

Our work also highlights the mnemonic benefits of retrieval, consistent with past work from cognitive psychology (e.g., Carrier & Pashler, 1992; Glover, 1989; Roediger & Karpicke, 2006b). Learning was greater following retrieval of learned information, as occurred when students wrote essays or recalled the passages, but not when students highlighted or took notes. From a practical perspective, it is telling that this advantage for learning activities that required retrieval was quite broad, as it emerged for multiple-choice fact items, multiple-choice inference items, and short-answer (problem solving) items. Many writing tasks could easily be modified to include a retrieval component—even writing tasks typically completed open-book (e.g., summarization, see Spirgel & Delaney, 2016). Focusing on a cognitive process rather than on a task is less constraining to instructors, as they can keep their familiar tasks and yet modify them to include the desired processes. For example, consider the popular task of concept mapping, where students represent the relationship between concepts graphically by drawing links, or words or phrases, between different nodes, or concepts. This task typically is completed open-book with students having access to their notes and reading material. However, Blunt and Karpicke (2014) recently showed that students learned more from making concepts maps when the task was modified to require retrieval; students who made concept maps without access to the original text did better on a test one week later than those who were able to refer back to the text while making their maps.

There is one caveat to our recommendation to add retrieval to learning activities: in our study, participants with greater structure-building ability benefited more from retrieving information than did those with less capacity to build narrative structures. Compared to high-structure builders, low-structure builders used fewer critical content words when recalling or writing closed-book essays, which in turn reduced the benefits of these activities. Because low-structure builders do not seem to be as able to benefit from retrieval processes when writing essays, future research should explore whether open-book essays may benefit these learners more.

One challenge for future research is to understand when (and in what form) people may need cognitive supports to engage in processes known to benefit learning. Our samples came from highly selective private universities, meaning that even our relatively less-skilled participants were likely more skilled than the general college-bound population, highlighting the importance of considering individual differences. For example, our "less proficient" writers needed to be told to write essays—but the minimum writing SAT score in the sample (590) still ranked in the 80th percentile of college-bound seniors (The College Board, 2015). Thus, it would likely be necessary to explicitly instruct students to write essays in a less-selective cohort, as such students would be less likely to produce essays given free-recall instructions. A different cohort might also yield different conclusions about working memory, which had surprisingly little effect on performance here.

More generally, our work sets up many future questions, as our instantiations of essays and free recall represent just two of the many possible writing assignments; slight changes in conditions and/or materials would potentially change which cognitive processes were engaged. For example, if writers had access to the source material while writing their essays, they might be better able to organize and elaborate upon the material, which should increase learning. However, at the same time, this change would reduce retrieval processing which should decrease learning. Similarly, different essay prompts will likely afford different cognitive processes; an essay comparing-and-contrasting a new concept to a known one may encourage more elaboration. Writing an argument may require more reorganization as learners fit the source material within the constraints of their argument (for preliminary evidence for this claim, see Wiley & Voss, 1999). Future research could also more stringently test the effects of particular processes by holding the learning activity constant while manipulating only the use of a particular cognitive process. For example, essays could be made to use more or less retrieval processing by manipulating whether or not participants had access to the source material while writing.

In short, this work brings together two different traditions, writing-to-learn and cognitive psychology to identify important cognitive processes that affect learning from writing. Our results link writing-to-learn effects to specific cognitive processes, an approach we believe is likely to be more fruitful than a focus on the infinite possible variations of writing tasks. Theoretically, it makes sense to link writing-to-learn to well-understood cognitive processes drawing from both cognitive psychology and educational psychology, as opposed to building completely separate theory. This approach has more promise for both theoretical understanding and for research to be able to realistically guide education. Further, exploring how individual differences affect writing-to-learn can provide valuable insight into how, why, and when writing activities can enhance learning. These insights could be invaluable to educators who must use the complex, and often contradictory, literature to decide how best to help their students learn.

References

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research, 54,* 577–596. http://dx.doi.org/10.3102/00346543054004577

Arnold, K. M., Daniel, D. B., Jensen, J., McDaniel, M., & Marsh, E. J. (2016). Structure building predicts grades in college psychology and biology. *Applied Cognitive Psychology*, *30,* 454–459. http://dx.doi.org/10.1002/acp.3226

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74,* 29–58. http://dx.doi.org/10.3102/00346543074001029

Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61,* 213–238. http://dx.doi.org/10.3102/00346543061002213

Benjamin, A. S., & Bjork, R. A. (2000). On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 638–648. http://dx.doi.org/10.1037/0278-7393.26.3.638

Bereiter, C., & Scardamalia, M. (1987). An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. *Curriculum Inquiry, 17,* 9–30. http://dx.doi.org/10.1080/03626784.1987.11075275

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology, 106,* 849–858. http://dx.doi.org/10.1037/a0035934

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49,* 229–240. http://dx.doi.org/10.1080/00221309.1953.9710088

Bui, D. C., & McDaniel, M. A. (2015). Enhancing learning during lecture note-taking using outlines and illustrative diagrams. *Journal of Applied Research in Memory & Cognition, 4,* 129–135. http://dx.doi.org/10.1016/j.jarmac.2015.03.002

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1118–1133. http://dx.doi.org/10.1037/a0019902

Callender, A. A., & McDaniel, M. A. (2007). The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology, 99,* 339–348. http://dx.doi.org/10.1037/0022-0663.99.2.339

Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology, 34,* 30–41. http://dx.doi.org/10.1016/j.cedpsych.2008.07.001

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1563–1569. http://dx.doi.org/10.1037/a0017021

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21,* 279–283. http://dx.doi.org/10.1177/0963721412452728

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268–276. http://dx.doi.org/10.3758/BF03193405

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633–642. http://dx.doi.org/10.3758/BF03202713

Connor-Greene, P. A. (2000). Making connections: Evaluating the effectiveness of journal writing in enhancing student learning. *Teaching of Psychology, 27,* 44–46. http://dx.doi.org/10.1207/S15328023TOP2701_10

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7,* 547–552. http://dx.doi.org/10.1016/j.tics.2003.10.005

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671–684. http://dx.doi.org/10.1016/S0022-5371(72)80001-X

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19,* 450–466. http://dx.doi.org/10.1016/S0022-5371(80)90312-6

Einstein, G. O., McDaniel, M. A., Bowers, C. A., & Stevens, D. T. (1984). Memory for prose: The influence of relational and proposition-specific processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 133–143. http://dx.doi.org/10.1037/0278-7393.10.1.133

Einstein, G. O., McDaniel, M. A., Owen, P. D., & Coté, N. C. (1990). Encoding and recall of texts: The importance of material appropriate processing. *Journal of Memory and Language, 29,* 566–581. http://dx.doi.org/10.1016/0749-596X(90)90052-2

Einstein, G. O., Morris, J., & Smith, S. S. (1985). Note-taking, individual differences, and memory for lecture information. *Journal of Educational Psychology, 77,* 522–532. http://dx.doi.org/10.1037/0022-0663.77.5.522

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28,* 122–128. http://dx.doi.org/10.2307/356095

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23. http://dx.doi.org/10.1111/1467-8721.00160

Friend, R. (2002). Summing it up. *Science Teacher, 69,* 40–43.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 430–445. http://dx.doi.org/10.1037/0278-7393.16.3.430

Gernsbacher, M. A., & Verner, K. R. (1988). *The multi-media comprehension battery*. Eugene, OR: University of Oregon, Institute of Cognitive and Decision Sciences.

Gingerich, K. J., Bugg, J. M., Doe, S. R., Rowland, C. A., Richards, T. L., Tompkins, S. A., & McDaniel, M. A. (2014). Active processing via write-to-learn assignments: Learning and retention benefits in introductory psychology. *Teaching of Psychology, 41,* 303–308. http://dx.doi.org/10.1177/0098628314549701

Glogger, I., Schwonke, R., Holzapfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology, 104,* 452–468. http://dx.doi.org/10.1037/a0026683

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81,* 392–399. http://dx.doi.org/10.1037/0022-0663.81.3.392

Grant, R. (1993). Strategic training for using text headings to improve students' processing of content. *Journal of Reading, 36,* 482–488.

Gurung, R. A. R., Weidert, J., & Jeske, A. (2010). Focusing on how students study. *Journal of the Scholarship of Teaching and Learning, 10,* 28–35.

Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist, 41,* 1106–1113. http://dx.doi.org/10.1037/0003-066X.41.10.1106

Hebert, M., Gillespie, A., & Graham, S. (2013). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing, 27,* 1043–1072.

Hounsell, D. (1997). Contrasting conceptions of essay-writing. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning* (106–125). Edinburgh, Scotland: Scottish Academic Press.

Humes, A. (1983). Research on the composing process. *Review of Educational Research, 53,* 201–216. http://dx.doi.org/10.3102/00346543053002201

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358. http://dx.doi.org/10.1037/0278-7393.26.2.336

Karttunen, H., Kröger, P., Oja, H., Poutanen, M., & Donner, K. J. (2006). *Fundamental astronomy*. Berlin, Germany: Springer.

Kellogg, R. T. (2001). Long-term working memory in text production. *Memory & Cognition, 29,* 43–52. http://dx.doi.org/10.3758/BF03195739

Kiewra, K. A., & Benton, S. L. (1988). The relationship between information-processing ability and notetaking. *Contemporary Educational Psychology, 13,* 33–44. http://dx.doi.org/10.1016/0361-476X(88)90004-5

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review, 11,* 203–270. http://dx.doi.org/10.1023/A:1021913217147

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14,* 219–224. http://dx.doi.org/10.3758/BF03194055

Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review, 1,* 126–129. http://dx.doi.org/10.3758/BF03200769

Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of web-based and lecture courses. *Journal of Experimental Psychology: Applied, 8,* 85–98. http://dx.doi.org/10.1037/1076-898X.8.2.85

Marshall, J. D. (1987). The effects of writing on students' understanding of literary texts. *Research in the Teaching of English, 21,* 30–63.

Martin, M., Nguyen, K., & McDaniel, M. A. (2016). Structure building differences influence learning from educational text: Effects on encoding, retention, and metacognitive control. *Contemporary Educational Psychology, 46,* 52–60. http://dx.doi.org/10.1016/j.cedpsych.2016.03.005

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,* 275–302. http://dx.doi.org/10.17239/jowr-2015.07.02.03

Mattern, K., Camara, W., & Kobrin, J. L. (2007, January). SAT writing: An overview of research and psychometrics to date. *College Board Research Notes* (p. 32). New York, NY: College Board. Retrieved from https://eric.ed.gov/?id=ED562868

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review, 8,* 299–325. http://dx.doi.org/10.1007/BF01464076

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103,* 399–414. http://dx.doi.org/10.1037/a0021782

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology, 88,* 508–519. http://dx.doi.org/10.1037/0022-0663.88.3.508

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review, 1,* 113–145. http://dx.doi.org/10.1007/BF01326639

McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language, 46,* 544–561. http://dx.doi.org/10.1006/jmla.2001.2819

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27,* 360–372. http://dx.doi.org/10.1002/acp.2914

McIntyre, S. (1992). Lecture notetaking, information-processing, and academic achievement. *Journal of College Reading and Learning, 25,* 7–17. http://dx.doi.org/10.1080/10790195.1992.10849993

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 151–171. http://dx.doi.org/10.1016/S0022-5371(77)80044-3

Newell, G. E. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English, 18,* 265–287.

Newell, G. E., & Winograd, P. (1995). Writing about and learning from history texts: The effects of task and academic ability. *Research in the Teaching of English, 29,* 133–163.

Nicotera, C. L., Shibley, I. A., Jr., & Milakofsky, L. K. (2001). Incorporating a substantial writing assignment into orgranic chemistry: Library research, peer review, and assessment. *Journal of Chemical Education, 78,* 50–53. http://dx.doi.org/10.1021/ed078p50

Norris, D., Oppler, S., Kuang, D., Day, R., & Adams, K. (2006). *The College Board SAT writing validation study: An assessment of predictive and incremental validity* (Report No. 2006–2). New York, NY: College Board.

Palmatier, R. A., & Bennett, J. M. (1974). Notetaking habits of college students. *Journal of Reading, 18,* 215–218.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *LIWC2007: Linguistic inquiry and word count*. Austin, TX: LIWC.

Penrose, A. M. (1992). To write or not to write: Effects of task and task interpretation on learning through writing. *Written Communication, 9,* 465–500. http://dx.doi.org/10.1177/0741088392009004002

Radmacher, S. A., & Latosi-Sawin, E. (1995). Summary writing: A tool to improve student comprehension and writing in psychology. *Teaching of Psychology, 22,* 113–115. http://dx.doi.org/10.1207/s15328023top2202_4

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

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Schmidt, S. R. (1983). The effects of recall and recognition test expectancies on the retention of prose. *Memory & Cognition, 11,* 172–180. http://dx.doi.org/10.3758/BF03213472

Schwonke, R., Hauser, S., Nückles, M., & Renkl, A. (2006). Enhancing computer-supported writing of learning protocols by adaptive prompts. *Computers in Human Behavior, 22,* 77–92. http://dx.doi.org/10.1016/j.chb.2005.01.002

Segev-Miller, R. (2007). Cognitive processes in discourse synthesis: The case of intertextual processing strategies. In M. Torrance, D. Galbraith, & L. Van Waes (Eds.), *Writing and cognition: Research and application* (pp. 231–250). Amsterdam, The Netherlands: Elsevier. http://dx.doi.org/10.1108/S1572-6304(2007)0000020016

Spirgel, A. S., & Delaney, P. F. (2016). Does writing summaries improve memory for text? *Educational Psychology Review, 28,* 171–196. http://dx.doi.org/10.1007/s10648-014-9290-2

Stein, B. S., Littlefield, J., Bransford, J. D., & Persampieri, M. (1984). Elaboration and knowledge acquisition. *Memory & Cognition, 12,* 522–529. http://dx.doi.org/10.3758/BF03198315

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63,* 358–385. http://dx.doi.org/10.1006/jecp.1996.0054

The College Board. (2009). *ACT and SAT concordance tables* (p. 40). New York, NY: College Board Research Notes. Retrieved from https://research.collegeboard.org/publications/content/2012/05/act-and-sat-concordance-tables

The College Board. (2015). *SAT percentile ranks*. New York, NY: College Board. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/sat/sat-percentile-ranks-crit-reading-math-writing-2014.pdf

Torrance, M., Thomas, G. V., & Robinson, E. J. (2000). Individual differences in undergraduate essay-writing strategies: A longitudinal study. *Higher Education, 39,* 181–200. http://dx.doi.org/10.1023/A:1003990432398

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition, 41,* 242–254. http://dx.doi.org/10.3758/s13421-012-0261-x

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. http://dx.doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505. http://dx.doi.org/10.3758/BF03192720

Voss, J. F., & Wiley, J. (1997). Developing understanding while writing essays in history. *International Journal of Educational Research, 27,* 255–265. http://dx.doi.org/10.1016/S0883-0355(97)89733-9

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 315–327). New York, NY: Macmillan.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91,* 301–311. http://dx.doi.org/10.1037/0022-0663.91.2.301

Zaromb, F. M., & Roediger, H. L., III. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38,* 995–1008. http://dx.doi.org/10.3758/MC.38.8.995

Received February 18, 2016
Revision received November 27, 2016
Accepted December 7, 2016 ■