American Psychological Association

Journal of Experimental Psychology: Applied

1076-898X/19/$12.00

2019, Vol. 25, No. 4, 543–557
http://dx.doi.org/10.1037/xap0000225

From JOLs to JOLs+: Directing Learners' Attention in Retrieval Practice to Boost Integrative Argumentation

Sarah Shi Hui Wong and Stephen Wee Hun Lim
National University of Singapore

The effects of retrieval practice on complex, meaningful learning outcomes that require more than just basic recall are of ongoing interest in the test-enhanced learning literature. Across two experiments, we investigated the extent that retrieval practice boosts integrative argumentation—the integration of opposing viewpoints to form conclusions. Participants were tasked to form an integrative argumentation response after reading a text containing arguments for and against an issue. We found that retrieval practice alone produced superior long-term retention of text content, but not better use of integrative stratagems relative to repeated study (Experiment 1). However, when retrieval practice was augmented with judgments of higher order learning (JOLs+) that oriented learners' attention toward the critical elements of integrative argumentation (Experiment 2), it led to the use of more integrative stratagems, relative to retrieval practice supplemented with judgments of learning (JOLs) that assessed the degree of material learned or remembered, and a notetaking condition paired with JOLs+. Importantly, the improvement in learners' use of integrative stratagems persisted even after controlling for the number of idea units in their responses. These findings suggest that JOLs+ serve as a potent metacomprehension monitoring intervention when paired with retrieval practice to enhance higher order learning outcomes.

Public Significance Statement
To what extent does retrieval practice boost higher order learning performance, beyond basic memory recall? Here we show that learners' integrative argumentation performance—their ability to evaluate, weigh, and combine opposing sides of an issue—is enhanced only when they have practiced retrieving information in an argumentative text from memory and are, crucially, further guided to direct their attention to the key elements required for effective integrative argumentation.

Keywords: integrative argumentation, complex learning, retrieval practice, attention, metacomprehension

Supplemental materials: http://dx.doi.org/10.1037/xap0000225.supp

Across a thriving body of literature, retrieval practice has been robustly demonstrated to produce durable learning of educationally relevant knowledge (for reviews, see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Karpicke, 2017). In the standard retrieval-based learning paradigm, learners either repeatedly study educational materials (restudy condition) or study and then practice retrieving the materials from memory (retrieval practice condition) before taking a final test in which their learning is assessed (Roediger & Karpicke, 2006a). While retrieval practice has been consistently shown to facilitate long-term retention of facts and concepts, recent studies have begun to explore the extent to which it may also benefit complex, higher order learning outcomes such as organizational processing (e.g., Zaromb & Roediger, 2010), analogical problem-solving (e.g., Wong, Ng, Tempel, & Lim, 2017), and transfer of learning (for reviews, see Carpenter, 2012; Pan & Rickard, 2018), such as to tests that require inferences to be drawn from the previously learned material (e.g., Butler, 2010; Chan, McDermott, & Roediger, 2006). Notably, engaging in retrieval practice after studying educational texts has been observed to improve not only students' ability to answer verbatim questions but, more crucially, also enhances their performance in answering higher order questions that require integrating facts and concepts from the text or applying them in novel contexts (Smith, Blunt, Whiffen, & Karpicke, 2016; see also Jensen, McDaniel, Wood-

This article was published Online First April 25, 2019.

Sarah Shi Hui Wong and Stephen Wee Hun Lim, Department of Psychology, National University of Singapore.

This research was supported in part by two National University of Singapore Educational Research grants awarded to Sarah Shi Hui Wong and Stephen Wee Hun Lim, respectively.

We are grateful to the following individuals for their invaluable assistance: Micole Wen Shi Soh for contributing to data collection and scoring in Experiment 1, and Hou Sheng Poh for contributing to the literature review, and the design, data collection, and scoring in Experiment 2.

Correspondence concerning this article should be addressed to Sarah Shi Hui Wong or Stephen Wee Hun Lim, Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4, 9 Arts Link, Singapore 117570. E-mail: psywshs@nus.edu.sg or psylimwh@nus.edu.sg

ard, & Kummer, 2014). Thus, it appears that retrieval practice can produce long-lasting and meaningful learning—organized, coherent, and integrated mental models of knowledge (e.g., Karpicke, 2012; Mayer, 2002)—with complex educational materials (e.g., Karpicke & Aue, 2015), rather than merely transient rote learning.

In the present research, we investigated the extent that retrieval practice enhances learners' argument-counterargument integration (*integrative argumentation*)—a complex learning outcome that involves critically evaluating, weighing, and integrating arguments on both sides of an issue before forming a final conclusion, much like solving a jigsaw puzzle by assembling its many pieces while having to turn over (i.e., negate) some of them in the process (e.g., Nussbaum, 2008; Nussbaum & Schraw, 2007). Specifically, across two experiments, we tested and provide evidence for the hypothesis that retrieval practice boosts integrative argumentation, but only when it is coupled with metacomprehension monitoring using judgments of learning (JOLs) that orient learners' attention to the critical elements of the intended higher order learning outcome (i.e., judgments of higher order learning [JOLs+]).

Retrieval Practice and Metacomprehension Monitoring

Notwithstanding evidence of the positive effects of retrieval practice on students' ability to answer higher order questions such as those requiring application or inference (e.g., Karpicke & Blunt, 2011; Smith et al., 2016; Smith & Karpicke, 2014), there are studies that have not observed such benefits. For instance, McDaniel, Howard, and Einstein (2009) reported the surprising finding that retrieval practice did not produce better performance than did a notetaking strategy for inferential multiple-choice and problem-solving criterial tests that required integrating and applying one's learned knowledge. To reconcile these mixed findings, one potential explanation relates to the limited success of retrieval practice alone in facilitating learners' metacomprehension monitoring when judging how well they understand the learned material, beyond alerting learners to the basic facts that they do not know or remember (Nguyen & McDaniel, 2016). Here, an important distinction must be noted between the direct and indirect effects that retrieval practice exerts on learning—whereas direct effects relate to those stemming from the act of retrieval itself in the absence of subsequent restudy or feedback, indirect effects pertain to the influences of retrieval on learning through mediating processes, such as when retrieval guides future study efforts or enhances processing of feedback (e.g., Karpicke, 2017; Roediger & Karpicke, 2006b). In particular, metacomprehension monitoring during initial retrieval and its impact on learners' utilization of subsequent restudy opportunities leans into the indirect effects of retrieval practice on learning.

Accurate metacomprehension monitoring is crucial for effectively regulating one's learning (Thiede, Anderson, & Therriault, 2003; Thiede, Wiley, & Griffin, 2011). However, the accuracy of learners' metacomprehension monitoring is often weak for tasks such as text learning and comprehension (Dunlosky & Lipko, 2007; Thiede, Griffin, Wiley, & Redford, 2009). According to construction-integration theory (Kintsch, 1988, 2005), learners should ideally engage in both surface-level text-based processing (i.e., knowledge of specific details and facts in a text) and global-level situation model processing (i.e., a more holistic understanding of what a text is about) to optimize their learning. Yet, learners tend to allocate more effort to surface-level text-based processing during initial study (e.g., Millis, Simon, & tenBroek, 1998; Stine-Morrow, Gagne, Morrow, & DeWall, 2004), and may continue to rely on such text-based factors during retrieval practice and subsequent restudy (e.g., Nguyen & McDaniel, 2016).

The implication is that although retrieval practice may offer metacomprehension advantages at a text-based (surface) level by providing learners with feedback about their memory for information in a text (e.g., Little & McDaniel, 2015), it may not necessarily alert learners to potential gaps in their situation model of the text and prompt efforts toward constructing a more cohesive and organized global representation of the text during subsequent restudy (e.g., Nguyen & McDaniel, 2016). Consequently, learners may perform poorly when they are tested on comprehension that requires inferential knowledge demanding global-level situation model (i.e., higher order) processing beyond recalled facts alone (e.g., Wiley, Griffin, & Thiede, 2005). Of particular interest is that the complex educational outcome of integrative argumentation entails multiple sophisticated processes that transcend factual recall and inferencing, as we detail in the following section.

Integrative Argumentation

As a valuable higher order learning outcome, argumentation has captured the interest of scholars and educators (e.g., Muller & Perret-Clermont, 2009; Schwarz & Baker, 2017; for reviews, see Asterhan & Schwarz, 2016; Manz, 2015). Broadly conceived as an interactive or introspective process in which individuals formulate and evaluate arguments (e.g., Golanics & Nussbaum, 2008; Nussbaum, 2005), argumentation can occur whenever there is a real or imagined conflict of ideas or opinions, and takes place across multiple academic disciplines such as philosophy, computing, linguistics, communications, psychology, and law (van Eemeren et al., 2014). Current dominant argumentation theories have dialectical, logical, and rhetorical dimensions, although a comprehensive discussion of these dimensions is beyond the intent and scope of this research (for a detailed perspective, see van Eemeren et al., 1996). Here, we focus on formal dialectical approaches that have commonly been used in contemporary educational psychology (e.g., Duschl, 2007; Nussbaum, 2008, 2011; van Eemeren, 2010; van Eemeren & Grootendorst, 1992; van Eemeren & Houtlosser, 2003; Walton, 1996, 2007, 2010), due to their direct relevance for teaching and learning today.

According to dialectical approaches, a dialogue underpins argumentation—through the process of pitting arguments against counterarguments and responding to counterarguments, a resolution is achieved (Walton, 2007). However, many students tend to struggle with this task and often do not generate counterarguments or address them adequately in their written argumentation (e.g., Perkins, Farady, & Bushey, 1991; Stapleton, 2001; Wolfe, Britt, & Butler, 2009). Termed the *myside bias*, Wolfe and Britt (2008) found that 50% of the participants in their study neglected to include any other-side information in their essays when arguing for or against a controversial issue. Even when students wrote their essays while conferring with an intervention text that covered both sides of the argument, 33% of the participants still exhibited the myside bias. Presumably, this bias may occur when students possess less sophisticated argumentation schemas and believe that

argumentation is simply about "lining up facts" (Wolfe & Britt, 2008).

Yet, for strong argumentation, one must not only include opposing (other-side) information in one's responses, but must also adequately address such information. In particular, Pollock (1987) has stressed the notion of defeasibility in argumentation, whereby good arguments should be able to withstand and defeat opposing arguments (also see Erduran, Simon, & Osborne, 2004; Nussbaum, 2008; Vorobej, 2006). Merely conceding or dismissing opposing claims without providing any counterarguments to rebut other-side information has limited effectiveness in improving the persuasiveness of one's argument, and might even weaken readers' impressions of the author (Wolfe et al., 2009). Thus, rather than only considering one side of the issue at hand, critically evaluating and meaningfully integrating both arguments and counterarguments to form conclusive statements in integrative argumentation (e.g., Nussbaum, 2008, 2011; Nussbaum & Schraw, 2007) is essential for increasing overall written argument quality (e.g., Nussbaum, Winsor, Aqui, & Poliquin, 2007; Voss & Van Dyke, 2001). Moreover, integrating arguments and counterarguments offers benefits for critical thinking (Nussbaum & Schraw, 2007), which has been viewed as comprising skills such as analysis when identifying and analyzing arguments, evaluation when assessing claims and arguments, and inference when querying evidence, conjecturing alternatives, and drawing conclusions (e.g., Facione, 1990). Accordingly, enhancing students' integrative argumentation may steer them toward better critical thinking.

To form an integrative argumentation response, two kinds of stratagems can be adopted: *refutational stratagems* and *integrative stratagems* (Nussbaum & Edwards, 2011). Refutational stratagems include strategies such as *refutation*, in which one explains why a counterargument is invalid, flawed, or false. While the use of refutational stratagems produces a more effective response than simply conceding or dismissing a counterargument, such stratagems are considered the least integrative because they primarily emphasize only the counterargument (Nussbaum & Edwards, 2011). Conversely, integrative stratagems include *weighing* and *design claim* strategies. The weighing strategy involves contrasting the merits of an argument with its counterargument, and is an integrative strategy because it considers both sides of an argument by showing that the benefits or strengths of an argument outweigh the benefits or strengths of a counterargument, or that the benefits of an argument outweigh the costs posited by a counterargument. Alternatively, the design claim strategy involves creating a compromise between two sides of an argument, or suggesting a new alternative solution to circumvent the problem. Design claim arguments are integrative as they attempt to mitigate the negative consequences put forth by the counterargument while preserving the benefits of the argument. In light of their particular pertinence for integrative argumentation, our research focused on integrative stratagems.

The Present Research

The key question we asked in the present research was whether retrieval practice promotes the higher order learning outcome of integrative argumentation, as assessed via learners' use of integrative stratagems. Specifically, effective integrative argumentation requires that learners critically evaluate, weigh, and combine arguments and counterarguments, before these are weaved into written discourse in support for their inferred overall conclusion (e.g., Nussbaum & Schraw, 2007). This stands in contrast to inference-based assessments in extant research that place relatively lower demands on learners to assess and integrate their learned knowledge, even though such assessments commonly involve the flexible and generative use of information when, for example, responding to multiple-choice questions that require synthesizing two or more ideas from a studied text (e.g., McDaniel et al., 2009; Nguyen & McDaniel, 2016). Accordingly, in view of the mixed evidence for the benefits of retrieval practice alone in enhancing inference performance, whether this same learning strategy may boost the relatively more complex educational outcome of integrative argumentation is, by logical extension, even more controversial.

In sum, while retrieval practice may improve retention of learned information related to an argumentative text, it may not necessarily lead to better integrative argumentation performance if it did not also orient learners' attention toward higher order processing that is crucial for a more holistic understanding of the text. We pursued this hypothesis through two experiments.

Experiment 1

In Experiment 1, learners were presented with an argumentative text that described arguments for and against an issue, and were randomly assigned to either restudy the text (repeated study condition) or practice retrieving the text's content from memory (retrieval practice condition). All learners then returned to the lab a week later, during which they were first trained on integrative argumentation. Learners then recalled, in a final integrative argumentation test, the arguments and counterarguments from the text they had studied in the previous week, and wrote an integrative conclusion using the argumentation skills they had learned during training.

In line with existing argumentation research (e.g., Nussbaum & Edwards, 2011; Nussbaum & Kardash, 2005; Nussbaum & Schraw, 2007), we examined the effects of retrieval practice on (a) learners' long-term retention of verbatim knowledge on the final test in correctly recalling idea units from the studied argumentative text and (b) their integrative argumentation performance (i.e., use of integrative stratagems when integrating arguments and counterarguments).

We made two predictions. First, practicing retrieval of the argumentative text during learning would lead to a higher number of idea units recalled during the final test, as opposed to repeatedly studying the text. Second, given our conjecture that retrieval practice may not necessarily boost integrative argumentation performance if, crucially, it did not also judiciously orient learners' attention during learning, retrieval practice alone would confer limited benefits on participants' abilities to use integrative stratagems in their written responses.

Method

Participants. Fifty-nine undergraduate students (37 were female) between 19 and 29 years of age ($M = 22.03$, $SD = 2.03$) from the National University of Singapore (NUS) participated in the experiment. All participants spoke English as their first language and received either course credit or monetary reimburse-

ment for their participation. In both reported experiments, the target sample size of 25 participants per condition was determined based on prior studies on retrieval practice (e.g., Karpicke & Blunt, 2011; McDaniel et al., 2009). We have also reported all measures, conditions, and data exclusions for both experiments. This research was conducted with the appropriate ethics-review-board approval by the NUS, and participants provided their written informed consent.

Design. The single between-subjects factor of interest was learning condition, in which participants were randomly assigned either to the repeated study or retrieval practice condition. Participants' performance on a written integrative argumentation test administered one week later was assessed. The dependent variables were the total number of idea units that participants correctly recalled, and the number of integrative stratagems that they used.

Materials.

Argumentative text. The debate on the introduction of Daylight Saving Time (DST) in Japan served as the topic of the argumentative text. Adapted from Kobayashi (2010), the text contained arguments for and against DST that were organized in two columns, respectively (available in the online supplemental materials: see "Argumentative Text on Daylight Saving Time"). Both positions were matched on the number of words ("for" DST: 233 words vs. "against" DST: 237 words) and idea units ("for" DST: 26 idea units vs. "against" DST: 25 idea units).

Preexperiment questionnaires. Prior to attending the experiment, participants completed three questionnaires that measured potential covariates, including their need for cognition, epistemological understanding, and English language proficiency.

As an assessment of their need for cognition, participants completed the short form of the Need for Cognition Scale (NFC; Cacioppo, Petty, & Kao, 1984), which contains 18 items that measure the degree to which one tends to engage in and enjoy effortful cognitive activities (Cacioppo, Petty, Feinstein, & Jarvis, 1996). NFC has been found to positively predict the number of arguments that participants recall (e.g., Cacioppo, Petty, & Morris, 1983). Sample items include "I would prefer complex to simple problems" and "I like tasks that require little thought once I've learned them" (reverse-scored). Participants performed all ratings on a 5-point Likert scale (1 = *extremely uncharacteristic*; 5 = *extremely characteristic*).

The 15-item Epistemological Understanding Scale (Kuhn, Cheney, & Weinstock, 2000) was used as a measure of epistemological understanding—beliefs about knowing and knowledge. The scale assesses individual differences in levels of epistemological understanding (absolutist, multiplist, or evaluativist) across five domains: personal taste (e.g., "Robin says the stew is spicy vs. Chris says the stew is not spicy at all"), aesthetic judgments (e.g., "Robin thinks the first painting they look at is better vs. Chris thinks the second painting they look at is better"), value judgments (e.g., "Robin thinks lying is wrong vs. Chris thinks lying is permissible in certain situations"), judgments of truth about the social world (e.g., "Robin agrees with one book's explanation of how children learn language vs. Chris agrees with another book's explanation of how children learn language"), and judgments of truth about the physical world (e.g., "Robin believes one book's explanation of how the brain works vs. Chris believes another book's explanation of how the brain works"). After being presented with two contradicting statements in each scale item, participants were to decide if "only one could be right" (absolutist answers, scored as 1 point) or if "both could have some rightness." If the latter option was chosen, participants were then asked to decide if "one could not be more right than the other" (multiplist answers, scored as 2 points) or "one could be more right than the other" (evaluativist answers, scored as 3 points). The maximum possible score was 45. Epistemological understanding has been found to predict argumentation skills (e.g., Mason & Scirica, 2006) and influence students' critical interpretation of a dual-position text (e.g., Mason & Boscolo, 2004).

In addition, participants' English language proficiency was assessed through 10 questions adapted from the Verbal Reasoning section of the GRE. The maximum possible score was 10.

Prelearning questionnaire. At the start of the experiment, participants read an introductory paragraph on DST (adapted from Kobayashi, 2010; available in the online supplemental materials: see "Introductory Paragraph on Daylight Saving Time") before rating their prior attitude toward DST on four items: (a) "Daylight Saving Time should be introduced in Japan," (b) "Japanese society will greatly benefit from Daylight Saving Time," (c) "The introduction of Daylight Saving Time will bring a lot of problems to Japanese society," and (d) "Daylight Saving Time is unsuitable for Japan." All ratings were made on a 7-point Likert scale (1 = *strongly disagree*; 7 = *strongly agree*). Negative items were reverse-scored.

At the same time, participants' prior familiarity with DST was assessed via two items: (a) "How familiar are you with Daylight Saving Time?" and (b) "How familiar are you with the controversy over the introduction of Daylight Saving Time in Japan?" Participants indicated their responses on a 7-point Likert scale (1 = *not familiar at all*; 7 = *extremely familiar*). Participants also rated the personal importance of DST to them on a 7-point scale (1 = *not important at all*; 7 = *extremely important*): "How personally important is the introduction of Daylight Saving Time in Japan to you?"

Postlearning questionnaire. A four-item postlearning questionnaire was administered after the learning phase. Specifically, participants rated on a 7-point Likert scale the extent to which the text was interesting (1 = *not interesting at all*; 7 = *extremely interesting*) and understandable (1 = *not understandable at all*; 7 = *extremely understandable*). Participants also rated their prior knowledge of the text content on a 7-point scale (1 = *not very much*; 7 = *very much*): "How much information in the text did you know prior to reading it?" and "How well did you know the subject matter covered in the text prior to reading it?"

Integrative argumentation training. To ensure that all participants understood what was required of them in integrative argumentation, they were trained using printed materials adapted from Nussbaum (2008) and Nussbaum and Schraw's (2007) argument criteria training protocol (available in the online supplemental materials: see all "Integrative Argumentation Training" handouts). Specifically, arguments and counterarguments, as well as refutation, weighing, and design claim strategies, were introduced and explained to all participants using an issue that was unrelated to DST ("Should candy be banned from school?"), with an example of each strategy provided. Participants were also given a written sample of how the three strategies could be used in combination to form an integrative conclusion. Following this, participants were given another issue ("Should the university lease personal parking

spaces?") and were asked to practice writing a response using the three strategies and forming an integrative conclusion on a lined handout.

After participants had completed their practice responses, they were presented with a printed handout containing four sample integrative conclusions that ranged from low to high quality, with an accompanying explanation of each sample conclusion's strengths and/or weaknesses. Participants were also explicitly instructed on the qualities of good integrative argumentation (e.g., a clear position, adequate supporting reasons, counterargumentation, integrative final conclusion, and essay organization) using Nussbaum and Schraw's (2007) "Criteria for a Good Argument" handout.

Final integrative argumentation test. To facilitate all participants' planning of their integrative argumentation responses on the final test, they were provided with an argumentation vee diagram (AVD) that served as a graphic organizer. Adapted from Novak and Gowin (1984), the AVD is a visual aid that is shaped like a *V* and is meant to direct learners' focus to a central question while reducing cognitive load when helping them to organize arguments and counterarguments. For example, consider the question that all participants were presented with during their integrative argumentation training: "Should candy be banned from school?" (see Figure 1).

Using the AVD, arguments for and against banning candy can be organized side-by-side to facilitate comparisons between opposing sides, with prompts ("Which side is stronger, and why?" and "Is there an 'in-between' solution?") included to guide learners' consideration and adoption of integrative stratagems (e.g., Nussbaum, 2008; Nussbaum & Schraw, 2007). Providing scaffolding through the AVD has been found to enhance integrative argumentation when delivered in conjunction with direct instruction on integrating arguments and counterarguments (Nussbaum, 2008).

Procedure. The experiment comprised two phases: learning and testing. Both phases were conducted 1 week apart, and participants were run in groups of three or fewer. Before attending the experiment, participants were asked to complete the preexperiment questionnaires that were administered online.

Learning phase. Upon their arrival at the laboratory, participants were told that they would be studying a text, and that they should study the text to the best of their ability knowing that they will need the information for a later test. Participants were not informed of the specific nature of the test.

Participants then read an introductory paragraph on DST (adapted from Kobayashi, 2010) and completed the prelearning questionnaire. Following which, participants were presented with the argumentative text on DST. In the retrieval practice condition, participants studied the text for 7 min, practiced retrieving the material for 7 min without reference to the text, restudied the text for 7 min, and practiced retrieving again for 7 min (as in Karpicke & Blunt, 2011). During retrieval, participants were given a blank sheet of paper and were instructed to write down as much information from the text as they could recall. Participants were also told that the information could be paraphrased, recalled in any order, or written in point form. Conversely, in the repeated study

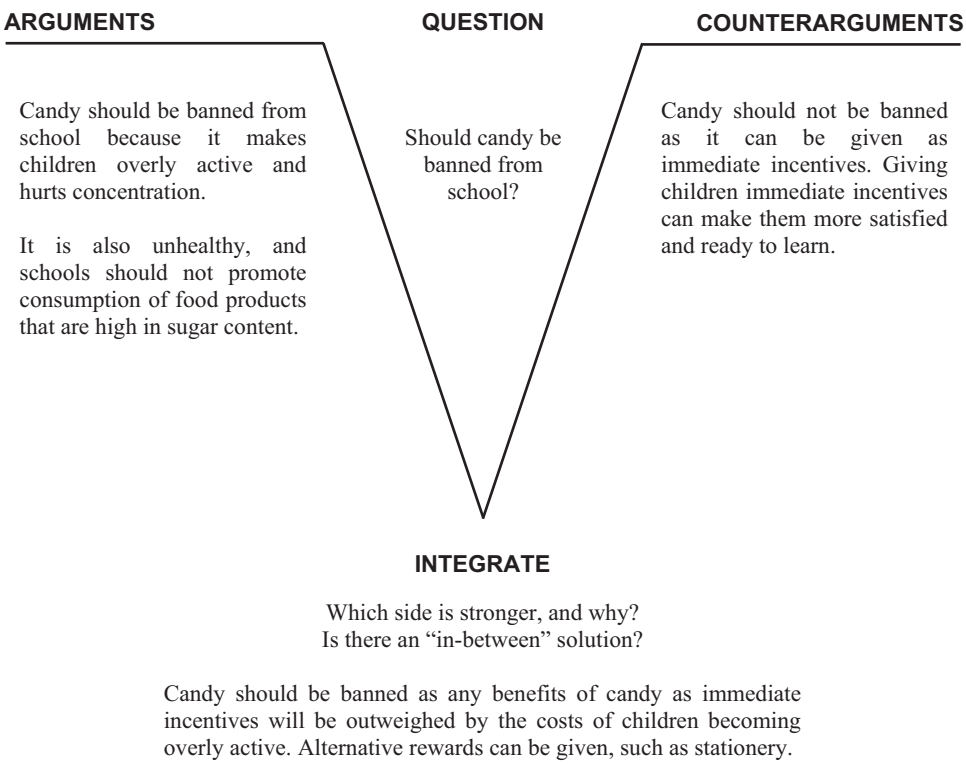

Figure 1. Sample argumentation vee diagram (AVD) adapted from Nussbaum (2008) and Nussbaum and Schraw (2007).

condition, participants studied the text in each of the four 7-min blocks. This procedure ensured that the total learning duration was exactly matched across both the repeated study and retrieval practice conditions. After completing the learning task, participants filled in the postlearning questionnaire. They were then thanked and dismissed.

Testing phase. One week later, participants returned for the testing phase of the experiment. They first underwent a training session on written integrative argumentation, in which they received a handout that introduced arguments and counterarguments using the issue of "Should candy be banned from school?" and that also explained and provided examples of how argumentation stratagems could be used to write an integrative conclusion. All training instructions were verbally reinforced by the experimenter. Participants were then given 10 min to practice using the argumentation stratagems and writing an integrative conclusion in response to the issue of "Should the university lease personal parking spaces?" Upon completion of their practice responses on a lined handout, participants were given the "Criteria for a Good Argument" handout (Nussbaum & Schraw, 2007), as well as a handout containing four sample integrative conclusions and accompanying explanations of their respective strengths and/or weaknesses. The experimenter verbally explained all instructions, and participants were asked to consider if the practice responses they had just written included all components of good integrative argumentation.

After completing the training session, participants were given 20 min to complete the final test without access to the argumentative text on DST; they were asked to use the material they had learned in the previous week to answer the question, "Should Daylight Saving Time be introduced in Japan?" Specifically, participants were instructed to recall arguments and counterarguments from the text they had studied and write them down on a blank AVD that was provided as a planning device, before writing an integrative conclusion in a blank text box using the argumentation skills they had learned during the training session (the full instructions are detailed in the online supplemental materials: see "Integrative Argumentation Final Test Instructions [Experiment 1]"). Finally, participants provided their demographic information, and were debriefed and reimbursed accordingly.

Results

Scoring. Scoring procedures were based on those developed by Nussbaum and Kardash (2005), Nussbaum and Schraw (2007), and Nussbaum and Edwards (2011), and included scoring participants' argumentation responses based on (a) the total idea units that they correctly recalled and (b) the total number of integrative stratagems that they used. Raters were trained to identify the types of stratagems. Two raters first independently scored five out of the 59 responses. Discrepancies between the two raters were then reviewed and resolved through discussion to reach 100% agreement. A further 15 responses were then independently scored by both raters. Interrater reliability was high ($\alpha = .93$). Discrepancies were further reviewed and resolved to yield 100% agreement. Given the good interrater agreement, the remaining responses were scored by one rater.

Total idea units. As a measure of recall performance, participants' final test responses were scored based on the total number of idea units included. The argumentative text was decomposed into 51 idea units, and one point was awarded for each idea unit that participants correctly recalled. Idea units had to be recalled without losing their original meaning. For example, for the idea unit "people will return from work before dark," a response that simply mentioned "people go home earlier" without specifying that it was "before dark" or "in daylight" would not earn any points.

Integrative stratagems. As a measure of participants' integrative argumentation performance, their responses were additionally scored for the number of integrative stratagems used, including those that involved weighing (i.e., contrasting the relative merits of an argument or counterargument) and design claims (i.e., proposing a new alternative solution or creating an in-between position).

Table 1
Explanations and Examples of Coding Categories

Category	Explanation	Examples
Integrative stratagem: Weighing	Contrasting the relative merits of an argument or counterargument	"It could be argued that Daylight Saving Time (DST) impacts the lifestyle of many and causes sleep disorders with the shifting time patterns. However, the human body is capable of adapting to much more adverse conditions than a change of a mere few hours to the body clock. The benefits of adopting a healthier lifestyle outweigh the possible negative impact on the body clock." "In the long run, the benefits that DST brings about, including economic growth from key sectors such as tourism, will outweigh the costs it has for transportation and financial institutions."
Integrative stratagem: Design claims	Developing an in-between position that combines the merits of both sides	"DST can be introduced in Japan, but the government must also implement laws and policies to regulate work and school culture, and ensure that people leave the office and school at a certain time."
	Suggesting alternative solutions	"A more practical solution would be to look at other sources of generating electricity, such as through solar panels, to reduce energy consumption in the summer."

Explanations and examples of the coding categories are detailed in Table 1.

Preliminary analyses. Independent-samples *t* tests were conducted to assess statistical equivalence between the repeated study and retrieval practice groups on their responses to the preexperiment, prelearning, and postlearning questionnaires. Means and standard deviations are presented in Table 2.

Both groups were comparable on their mean NFC scores, epistemological understanding scores, GRE scores, mean prior familiarity with DST, personal importance of DST, ratings of text interestingness and understandability, as well as mean prior knowledge of the text content, all *p*s $> .05$. However, participants in the retrieval practice condition ($M = 4.53$, $SD = 0.95$) reported significantly more favorable mean prior attitudes toward the introduction of DST in Japan, as compared to participants in the repeated study condition ($M = 4.03$, $SD = 0.91$), $t(57) = -2.03$, $p = .047$, $d = 0.54$.

We further examined the correlations between participants' responses on the preexperiment, prelearning, and postlearning questionnaires and their final test performance, including total idea units and use of integrative stratagems. Participants' GRE scores positively correlated with the number of integrative stratagems they used, $r = .33$, $p = .012$. No other questionnaire items significantly correlated with the final test performance variables, all *p*s $> .05$. Based on these observations, we controlled for participants' GRE scores and mean prior attitudes toward DST in our subsequent main analyses.

Main analyses.

Total idea units. A one-way analysis of covariance (ANCOVA) was conducted with participants' total idea units recalled as the dependent variable, learning method (repeated study vs. retrieval practice) as the independent variable, and participants' GRE scores and mean prior attitudes toward DST as covariates. To first check the homogeneity of slopes assumption, interactions between each of the two covariates with learning method were entered into the model, alongside the main effects. None of the interactions were significant, all *p*s $> .05$, indicating that the homogeneity of slopes assumption had been met. The interaction terms were then removed from the model, which was reestimated. Consistent with our prediction, participants in the retrieval practice condition ($M_{adj} = 11.02$) recalled significantly more idea units than did those in the repeated study condition ($M_{adj} = 7.12$), even after controlling for GRE scores and mean prior attitudes toward DST, $F(1, 55) = 15.21$, $MSE = 13.73$, $p < .001$, $\eta_p^2 = .22$.

Integrative stratagems. A one-way ANCOVA was performed on the number of integrative stratagems that participants used in their responses, with the same two covariates and learning method as the independent variable. After ascertaining that the homogeneity of slopes assumption had not been violated, we found that the number of integrative stratagems that participants used, indeed, did not significantly differ across the retrieval practice ($M_{adj} = 1.55$) and repeated study ($M_{adj} = 1.31$) conditions after controlling for GRE scores and mean prior attitudes toward DST, $F(1, 55) = 0.93$, $MSE = 0.84$, $p = .34$, $\eta_p^2 = .017$.

Table 2
Means and Standard Deviations of Participants' Responses on Preexperiment, Prelearning, and Postlearning Questionnaires (Experiment 1)

	Repeated study		Retrieval practice	
Variables	*M*	*SD*	*M*	*SD*
Need for cognition	3.25	.66	3.18	.60
Epistemological understanding	33.50	3.47	33.17	4.29
English proficiency (GRE)	3.57	1.59	3.38	1.57
Prior attitude toward DST	4.03	.91	4.53	.95
Prior familiarity with DST	2.17	1.15	2.36	1.16
Personal importance of DST	1.93	1.36	2.76	1.90
Text interestingness	4.90	1.09	4.90	1.40
Text understandability	6.23	.63	6.03	.91
Prior knowledge of text content	1.72	.95	2.07	1.19

Note. $N = 59$. DST = Daylight Saving Time. Mean scores were computed for need for cognition, prior attitude toward DST, prior familiarity with DST, and prior knowledge of text content.

Discussion

Consistent with a growing base of literature showing that testing promotes long-term retention of knowledge (e.g., Dunlosky et al., 2013; Karpicke, 2017; Roediger & Karpicke, 2006a, 2006b), we found that retrieval practice significantly enhanced participants' verbatim recall performance relative to repeated study, as assessed by the total number of idea units they recalled from the argumentative text. However, we found no evidence that retrieval practice led to the use of more integrative stratagems in learners' responses. Although retrieval practice as implemented in the present experiment improved learners' memory of the argumentative text content, it did not yield better integration of arguments and counterarguments. To explicate this finding, one potential account is that while learners' metacomprehension accuracy of surface-level propositional representations may have benefited from retrieval practice that increased their metacognitive awareness of what they do not know or remember (e.g., Little & McDaniel, 2015), the same may not have applied to the accuracy of their global-level situation models related to their overall understanding of the text (e.g., Nguyen & McDaniel, 2016). In particular, when engaging in retrieval practice in the present experiment, learners simply wrote down as much information as they could recall from the argumentative text. Such retrieval did not involve focusing learners' attention on the critical higher order learning outcome of integrative argumentation and orienting learners toward global-level situation model processing that is presumably crucial to achieve this desired learning outcome (e.g., Nguyen & McDaniel, 2016).

Accordingly, one potential solution to improve the usefulness of retrieval practice in enhancing integrative argumentation is to supplement it with a metacomprehension monitoring intervention that guides learners to consider the intended higher order educational outcome and whether they are adequately prepared to meet that outcome. Experiment 2 tested this possibility.

Experiment 2

Learners' metacomprehension monitoring during retrieval practice can be facilitated through making metacognitive judgments before restudy that direct them toward global-level situation model processing. For instance, Nguyen and McDaniel (2016) found that instructing learners to make judgments of inferencing (JOIs; i.e.,

judgments about their mental models and how well they can apply their learned knowledge) alerted them to gaps in their comprehension when restudying a text after a retrieval attempt. In their study, participants read a technical text passage on brakes/pumps, and were asked to make JOIs related to "how well they knew how brakes/pumps work," "how likely they were able to distinguish between the different types of brakes/pumps," and "how likely they were able to help a friend who has problems with their brakes/pumps." Notably, retrieval practice that was augmented with JOIs produced superior inference performance on questions that required integrating knowledge from different paragraphs and prior knowledge, relative to a control notetaking condition that was similarly paired with JOIs. Conversely, this advantage over notetaking was not observed for a standard retrieval practice condition that did not include JOIs. Presumably, making JOIs after retrieval stimulated participants to engage in additional global-level situation model processing during restudy that was oriented toward the intended learning outcome of inferencing, thereby enhancing their inference performance. In contrast, learners who made JOIs after notetaking spent less time studying the text during restudy—without the metacognitive benefits of retrieval practice in highlighting gaps in their knowledge, learners' higher sense of subjective fluency during their initial study may have inflated their perceived understanding of the text, such that they underutilized the restudy period and subsequently displayed poorer inference performance despite being oriented toward global-level situation model processing via JOIs (Nguyen & McDaniel, 2016).

Interestingly, Nguyen and McDaniel (2016) further found that learners' inference performance was superior after engaging in retrieval practice that had been augmented with JOIs rather than JOLs, although both conditions produced comparable free recall performance. Whereas JOIs involve asking learners to judge how well they can apply their knowledge, JOLs involve asking learners to judge how well they have learned the material or how well they have remembered the information (e.g., Koriat, 1997). While JOLs have been reported to confer learning benefits when they lead to adaptive studying (for a review, see Schwartz & Metcalfe, 2017), such judgments appear to be less effective for more complex tasks such as inferencing, given that JOLs only assess the degree of material learned or remembered, and may thus be inadequate to push learners' monitoring of learning toward global-level situation model processing (Nguyen & McDaniel, 2016).

Judgments of Higher Order Learning (JOLs+)

Taken together, extant research provides a logical basis to expect that learners' integrative argumentation performance can potentially be enhanced through retrieval practice paired with metacomprehension judgments that orient their attention to the critical elements required to achieve the desired higher order learning outcome, beyond merely judging the extent to which they remember or understand learned information (i.e., JOLs). We termed such metacomprehension judgments *JOLs+*, which refer collectively to judgments related to any intended higher order learning outcome, and can thus be broadly applied to diverse educational contexts that encompass, for instance, making inferences from learned information (i.e., JOIs), and even creating novel products. In the present research, however, we focused on JOLs+ applied to integrative argumentation as the key higher order learning outcome of interest. Specifically, we developed JOLs+ questions that guided learners toward the multiple complex processes required for effective integrative argumentation, such as addressing counterarguments, weighing arguments and counterarguments, and creating compromises or novel alternative solutions to reconcile opposing sides.

In Experiment 2, we adapted Nguyen and McDaniel's (2016) paradigm to investigate whether the use of retrieval practice augmented with JOLs+ that involved assessing one's knowledge related to integrative argumentation would drive an increase in learners' use of integrative stratagems, as compared to retrieval practice paired with the traditional judgments of learning (i.e., JOLs), as well as a control condition in which learners engaged in notetaking instead of retrieval, but were also tasked to make JOLs+. As in Experiment 1, we further assessed the total number of idea units that participants recalled in their responses.

At the same time, we introduced an improvement to the integrative argumentation training that all participants received. Specifically, whereas learners in Experiment 1 received a lined handout when practicing integrative argumentation during training and a blank AVD as a planning device during the final test, participants in Experiment 2 received a blank AVD with five "critical questions" added to provide better scaffolding during both training and their planning of responses on the final test. Indeed, Nussbaum and Edwards (2011) have found that guiding learners to answer critical questions during classroom verbal discussions directs students toward adopting more integrative stratagems. Rooted in Walton's (1996) dialogue theory, critical questions involve creating a burden of proof that encourages learners to determine the strength of an argument. An example of the use of a critical question is:

> *Claim:* Drinking diet soda will cause cancer.
>
> *Supporting Reason:* Dr. Tan says so.
>
> *Critical Question:* Is Dr. Tan an expert in nutrition or other health-related disciplines?
>
> *Answer:* No, he specializes in computer engineering.
>
> *Refutation:* Then your argument is flawed because Dr. Tan does not have expertise in the domain that he is making claims about.

As illustrated in this example, critical questions such as "Is the argument unlikely?" or "Is there another likely explanation?" facilitate learners' use of balanced reasoning.

We made two predictions. First, both retrieval practice conditions would yield a basic recall advantage in terms of the total number of idea units in learners' argumentative responses, as compared to notetaking (with JOLs+); recall performance would not differ across both retrieval practice (with JOLs vs. JOLs+) conditions. Second, and more important, retrieval practice coupled with JOLs+ would lead to greater use of integrative stratagems, as compared to retrieval practice with JOLs or to notetaking with JOLs+; the latter two conditions would not differ in this aspect (i.e., in enhancing integrative argumentation performance, retrieval practice and JOLs+ must be adopted in tandem; see the introduction).

Method

Participants. Seventy-eight students (60 were female) aged between 19 and 30 (M = 21.80, SD = 1.79) from the NUS participated in the study for either course credit or cash payment.

Design. The study used a between-subjects design with learning strategy (retrieval-JOLs vs. retrieval-JOLs+ vs. notetaking-JOLs+) as the independent variable of interest. Participants were randomly assigned to one of the three learning conditions. The dependent variable was performance on an integrative argumentation test, as assessed via the total number of idea units that participants correctly recalled and the number of integrative stratagems that they used.

Materials.

Argumentative text. The same argumentative text from Experiment 1 on the introduction of DST in Japan was used in this study.

Metacomprehension monitoring interventions. To facilitate metacomprehension monitoring during participants' learning of the argumentative text, both JOLs and JOLs+ questions were designed. JOLs questions (as adapted from Nguyen & McDaniel, 2016) asked learners to make a general estimate of how well they had learned the text on a scale from 0 (*definitely will not*) to 100 (*definitely will*), including (a) "How well do you think you understood the text?" (b) "How likely are you to remember this text on a future test?" and (c) "How much do you think you have learned about the case for/against Daylight Saving Time in Japan?"

In contrast, JOLs+ questions asked learners for a general estimate of their knowledge related to the higher order learning outcome of integrative argumentation. Specifically, learners were asked to respond to the following questions: (a) "How difficult do you think it is to argue the case for/against Daylight Saving Time in Japan?" (rated from 0 = *definitely not difficult* to 100 = *definitely difficult*), (b) "How confident do you think you are in arguing the case for/against Daylight Saving Time in Japan?" (rated from 0 = *definitely not confident* to 100 = *definitely confident*), (c) "If you are asked to argue for/against Daylight Saving Time in Japan, how well do you think you can (i) address the counter claims to your claims, (ii) argue that one claim is weaker than another claim, (iii) comment on the benefits and costs of a claim, or the benefits and costs of one claim compared to another, (iv) create novel solutions to resolve the arguments on both sides of the issue?" (each item rated 0 = *definitely not well* to 100 = *definitely well*).

Preexperiment questionnaires. As in Experiment 1, prior to attending the experiment, participants completed online questionnaires that assessed their need for cognition, epistemological understanding, and English language proficiency through the short-form of the NFC scale (Cacioppo et al., 1984), Epistemological Understanding Scale (Kuhn et al., 2000), and 10 questions adapted from the Verbal Reasoning section of the GRE, respectively.

Prelearning questionnaire. At the start of the experiment, participants' prior attitude toward DST and prior familiarity with DST, as well as the personal importance of DST to participants, were measured using the same questionnaire as in Experiment 1.

Postlearning questionnaire. A four-item questionnaire identical to that in Experiment 1 was administered after the learning phase to assess how interesting and understandable participants found the argumentative text to be, as well as their prior knowledge of the text content.

Integrative argumentation training. Before embarking on the final test, participants were trained on written integrative argumentation. The training procedure was identical to that in Experiment 1, with a few refinements. Specifically, when practicing integrative argumentation in response to the issue of "Should the university lease personal parking spaces?" all participants were provided with an AVD as a graphic organizer to form their arguments, counter-arguments, and integrative conclusion. Five critical questions (adapted from Nussbaum & Edwards, 2011) were also added to the AVD as prompts to guide all participants' written responses: (a) "Are any of the arguments not as important as others?" (b) "Are any of the arguments unlikely?" (c) "Is there a creative solution to any problem raised?" (d) "Is the creative solution practical (consider costs)?" and (e) "For any argument, can you think of any examples to the contrary, or other likely explanations?" Such scaffolding was intended to better guide participants' learning of integrative argumentation (e.g., Nussbaum & Edwards, 2011), in contrast to the simple lined handout that had been used in Experiment 1. After writing their practice responses, participants were then provided with a printed handout containing three sample integrative conclusions that outlined weak to strong responses, alongside explanations of each conclusion's strengths and/or weaknesses (available in the online supplemental materials: see all "Integrative Argumentation Training" handouts).

Final integrative argumentation test. As a planning device, all participants were provided with a blank AVD similar to that in Experiment 1 but with the five critical questions added as prompts, before they typed out their responses on the final integrative argumentation test.

Procedure. As in Experiment 1, participants were instructed to complete the preexperiment questionnaires before attending the study. Upon arrival at the laboratory, all participants were first tasked to complete the prelearning questionnaire after reading the same introductory paragraph on DST (adapted from Kobayashi, 2010) that had been used in Experiment 1. Participants then proceeded to complete the study's two phases: learning and testing.

Learning phase. Participants were randomly assigned to study the argumentative text on DST in Japan using either the retrieval-JOLs, retrieval-JOLs+, or notetaking-JOLs+ method. In the retrieval-JOLs condition, participants were asked to study the text for 7 min, practice written free recall without reference to the text for 7 min, answer the JOLs questions, and then restudy the text for 7 min (e.g., Nguyen & McDaniel, 2016). The retrieval-JOLs+ condition was identical to the retrieval-JOLs condition, except that participants responded to the JOLs+ questions (instead of the JOLs questions) after the free-recall task. In the notetaking-JOLs+ condition, participants studied the text for 7 min, restudied the text for another 7 min, answered the JOLs+ questions and, finally, restudied the text for 7 min; participants in this condition were also allowed to take notes (e.g., by highlighting or writing on the text) during their study. After the learning task, all participants completed the postlearning questionnaire, before immediately proceeding to the testing phase of the experiment (as in Nguyen & McDaniel, 2016).

Testing phase. All participants were told that they would be expected to type out an integrative answer to an argumentative

question, and that they would be guided on the basics of written argumentation and the use of argumentation stratagems. Participants then went through training on written integrative argumentation.

At the end of the training, all participants were given 20 min to type out an answer to "Should Daylight Saving Time be introduced in Japan?" They were reminded to include as much information as possible, involve both arguments and counterarguments, and to incorporate all strategies taught for strong argumentation (the full instructions are detailed in the online supplemental materials: see "Integrative Argumentation Final Test Instructions [Experiment 2]"). A blank AVD with critical questions added as prompts was provided for all participants to plan their answer. Upon completion of the argumentative writing task, all participants were thanked and reimbursed accordingly for their participation.

Results

Scoring. As in Experiment 1, participants' responses were scored based on the total number of idea units that were correctly recalled and the number of integrative (i.e., weighing and design claims) stratagems used. Two raters first independently scored 18 out of the 78 responses. There was high interrater reliability (α = .95). Discrepancies between the two raters were then reviewed and resolved through discussion to reach 100% agreement, and the remaining scripts were scored by one rater.

Preliminary analyses. One-way analyses of variance were conducted to assess statistical equivalence across all three learning conditions on their responses to the preexperiment, prelearning, and postlearning questionnaires. Means and standard deviations are presented in Table 3. There were no significant differences across the three groups on all questionnaire items, including: mean NFC scores, epistemological understanding scores, GRE scores, mean prior attitudes toward DST, mean prior familiarity with DST, personal importance of DST, ratings of text interestingness and understandability, and mean prior knowledge of the text content, all *p*s > .05.

We further examined correlations between participants' responses on the preexperiment, prelearning, and postlearning questionnaires and their final test performance, including total idea units and the number of integrative stratagems used. There was a significant positive correlation between participants' GRE scores and the number of integrative stratagems in their responses, r = .24, p = .038, indicating that better English language proficiency is associated with greater use of integrative stratagems. In addition, participants' ratings of text understandability positively correlated with the number of integrative stratagems they used, r = .24, p = .036. No other correlations with the final test performance variables were significant, all *p*s > .05. Based on these observations, we controlled for participants' GRE scores and ratings of text understandability in our subsequent main analyses.

Table 3
Means and Standard Deviations of Participants' Responses on Preexperiment, Prelearning, and Postlearning Questionnaires (Experiment 2)

Variables	Retrieval-JOLs		Retrieval-JOLs+		Notetaking-JOLs+	
	M	*SD*	*M*	*SD*	*M*	*SD*
Need for cognition	3.24	.53	3.09	.73	3.02	.70
Epistemological understanding	34.27	3.08	34.88	3.56	34.92	3.79
English proficiency (GRE)	2.85	1.80	3.50	1.92	2.65	1.65
Prior attitude toward DST	4.50	.81	4.03	1.06	4.13	.93
Prior familiarity with DST	2.48	1.23	2.58	1.04	2.73	1.17
Personal importance of DST	2.23	1.21	2.35	1.57	2.46	1.53
Text interestingness	5.00	1.33	4.69	1.38	4.85	1.38
Text understandability	6.19	1.02	6.62	.57	6.08	.89
Prior knowledge of text content	2.35	1.42	2.12	1.09	2.73	1.28

Note. N = 78. DST = Daylight Saving Time. Mean scores were computed for need for cognition, prior attitude toward DST, prior familiarity with DST, and prior knowledge of text content.

Main analyses.

Total idea units. A one-way ANCOVA was performed with participants' total idea units as the dependent variable, learning condition as the independent variable, and participants' GRE scores and ratings of text understandability as covariates. After ascertaining that the homogeneity of slopes assumption had been met, we found a significant effect of learning condition, $F(2, 73)$ = 13.22, MSE = 18.42, $p < .001$, η_p^2 = .27. Pairwise comparisons revealed that participants in the retrieval-JOLs (M_{adj} = 13.56) and retrieval-JOLs+ (M_{adj} = 10.68) conditions recalled significantly more idea units than did participants in the notetaking-JOLs+ condition (M_{adj} = 7.42), $p < .001$, d = 1.43, and p = .011, d = 0.76, respectively, even after controlling for GRE scores and ratings of text understandability. In addition, participants in the retrieval-JOLs condition displayed significantly better recall of the idea units than did their counterparts in the retrieval-JOLs+ condition, p = .022, d = 0.67.

Integrative stratagems. The number of integrative stratagems that participants used was analyzed through a one-way ANCOVA with learning condition as the independent variable, and participants' GRE scores and ratings of text understandability as covariates. After checking that the homogeneity of slopes assumption had been met, we found a significant effect of learning condition after controlling for both covariates, $F(2, 73)$ = 11.52, MSE = 1.94, $p < .001$, η_p^2 = .24. Pairwise comparisons revealed that participants in the retrieval-JOLs+ condition (M_{adj} = 3.91) used significantly more integrative stratagems in their argumentation responses, as compared to participants in the retrieval-JOLs (M_{adj} = 2.69) and notetaking-JOLs+ (M_{adj} = 1.97) conditions, p = .003, d = 0.88, and $p < .001$, d = 1.39, respectively. There was no significant difference between the number of integrative stratagems used by participants in the retrieval-JOLs and notetaking-JOLs+ conditions, p = .066, d = 0.52.

Importantly, these trends persisted even after additionally controlling for the total number of idea units in participants' argumentation responses as a third covariate through a one-way ANCOVA, $F(2, 72)$ = 10.33, MSE = 1.85, $p < .001$, η_p^2 = .22, suggesting that the present effects were not attributable to factual recall, per se. Pairwise comparisons indicated that participants in the retrieval-JOLs+ condition (M_{adj} = 3.90) used significantly more integrative stratagems compared to those in the retrieval-JOLs (M_{adj} = 2.45) and notetaking-JOLs+ (M_{adj} = 2.22) conditions, p = .001, d = 1.06, and $p < .001$, d = 1.24, respectively. Participants in the retrieval-JOLs and notetaking-JOLs+ conditions did not significantly differ in the number of integrative stratagems they used, p = .60, d = 0.17.

Discussion

Replicating our findings from Experiment 1 and many other demonstrations of the benefits of testing for memory retention and meaningful learning (e.g., Karpicke, 2012; Roediger & Karpicke, 2006a, 2006b), we found that retrieval practice produced better recall of the content of an argumentative text, as opposed to a notetaking condition that did not include retrieval during studying. More crucially, the key finding from Experiment 2 was that retrieval practice paired with JOLs+ led to superior integrative argumentation performance as assessed by learners' greater use of integrative stratagems, relative to retrieval practice paired with JOLs or notetaking with JOLs+. Importantly, this improvement in integrative argumentation cannot be attributed to the sheer quantity of information that learners remembered from the argumentative text, given that the improvement persisted even after controlling for the number of idea units that participants correctly recalled. In fact, verbatim recall of idea units from the argumentative text was better when retrieval practice had been supplemented with JOLs than with JOLs+. Simply put, despite recalling relatively fewer idea units, retrieval-JOLs+ participants effectively used more integrative stratagems in their argumentative responses than did retrieval-JOLs participants. This suggests that the benefits of making JOLs+, as opposed to JOLs, during retrieval practice were specific to the targeted higher order learning outcome—retrieval-JOLs+ participants outperformed those in the retrieval-JOLs condition on a measure of integrative argumentation.

General Discussion

In two experiments, we investigated the extent to which retrieval practice boosts the complex, meaningful learning outcome of integrative argumentation. We predicted and subsequently showed that retrieval practice on its own was insufficient to improve learners' integration of arguments and counterarguments relative to repeated study, although it enhanced verbatim recall of information from a studied argumentative text (Experiment 1). Crucially, however, the benefits of retrieval practice for integrative argumentation emerged when it was supplemented with a metacomprehension monitoring intervention (JOLs+) that oriented learners' attention to the critical elements of the intended higher order learning outcome of integrative argumentation (Experiment 2). Specifically, learners who had practiced retrieving an argumentative text and made JOLs+ before they restudied the text subsequently achieved greater success in using more integrative stratagems, as compared to learners who had similarly practiced retrieval but made JOLs that merely involved assessing the degree of material learned or remembered, or to learners who had engaged in notetaking (a relatively more active learning strategy than repeated studying) with JOLs+. Moreover, it is noteworthy that this advantage persisted even after controlling for the number of idea units in learners' responses, suggesting that this effect was not simply due to learners remembering more verbatim information from the text (i.e., purely recalling more content from the text did not enable learners to use more integrative stratagems with the additional content).

Our findings are in line with some recent studies that have reported the limited benefits of retrieval practice alone on higher order criterial tests that involve inferencing, beyond memory-based criterial tests (e.g., McDaniel et al., 2009; Nguyen & McDaniel, 2016). In particular, for assessments requiring inferential knowledge, retrieval practice by itself was reportedly inadequate if learners focused primarily on surface-level text-based information. Likewise, in the present research, retrieval practice alone fell short in boosting integrative argumentation, which involves relatively more complex processes such as critically evaluating and weighing arguments and counterarguments, before inferring an integrative conclusion. Rather, supplementing retrieval practice with a metacomprehension monitoring intervention (JOLs+) that directs learners' attention to the intended higher order learning outcome may alert them to internally assess their global-level situation models of their overall understanding of the to-be-learned material (e.g., Nguyen & McDaniel, 2016). This may, in turn, guide learners to more effectively direct their attentional resources toward additional global-level situation model processing during restudy, which then promotes better use of integrative stratagems when constructing their final argumentative responses. Conversely, learners in our study who practiced retrieval and completed JOLs may have continued to focus on additional surface-level text-based processing during their restudy opportunity, such that they displayed only better verbatim recall but not integrative argumentation performance. In other words, these learners may have continued to err in their learning strategy relative to the crucial higher order educational goal, thereby failing it. While committing and correcting errors has been shown to be beneficial in a variety of learning contexts (see Metcalfe, 2017, for a review), the implication is that learners should further be guided to recognize and rectify their erroneous learning strategy in a way that serves the specific educational goal directly, in order for this correction to be useful in learning (see Wong & Lim, 2019, for a discussion of guiding errors).

We demonstrated that JOLs+ can enhance integrative argumentation—a complex measure of meaningful learning that requires not only inferencing information from an argumentative text, but also evaluating and integrating arguments and counterarguments into cohesive and persuasive statements. Moreover, it is worth noting that learners in our study displayed significant improvements in their integrative argumentation performance within a single session. Hence, retrieval practice with JOLs+ is not only effective but, in this regard, also efficient in enhancing students' integrative argumentation. Accordingly, it may be worthwhile to consider how this learning strategy can be practically applied in educational settings.

Educational Implications

To boost students' integrative argumentation, our results suggest that educators can take advantage of retrieval practice, which has been widely established as an effective learning strategy across various domains and learning outcomes (e.g., Dunlosky et al., 2013; Karpicke, 2017; Roediger & Butler, 2011), with demonstrated benefits in real-world school settings (e.g., Carpenter, Pashler, & Cepeda, 2009; McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; Roediger, Agarwal, McDaniel, & McDermott, 2011). By, specifically, combining retrieval practice with JOLs+ as a metacomprehension monitoring intervention that guides students' attention toward assessing their holistic understanding of their study materials beyond simply knowing surface-level content, educators may be better able to improve their students'

integrative argumentation performance in flexibly using knowledge to create coherent conclusions that consider arguments both for and against the position they endorse. For instance, after direct instruction on argumentation topics and skills, short quizzes can be administered for students to engage in retrieval practice (e.g., Roediger & Pyc, 2012), followed by having students respond to JOLs+ before reviewing their study materials. Notably, extended training does not appear to be strictly necessary, given that the benefits of retrieval practice with JOLs+ for integrative argumentation observed in the present research emerged after only a single intervention session. At the same time, students can be encouraged to implement retrieval practice with JOLs+ during their self-regulated study. Because most written assignments in college demand argumentation (e.g., Wolfe, 2011), students can engage in retrieval practice of their learned knowledge on a particular topic when preparing for essay-related assessments, while questioning themselves if they have achieved a competent, holistic understanding of the content in terms of being able to form integrative argumentation responses.

Retrieval practice with JOLs+ can also be potentially implemented in online educational contexts, in view of the growing popularity of online learning and distance education (for reviews, see Bernard et al., 2004; U.S. Department of Education, Office of Planning, Evaluation, & Policy Development, 2010; see also Loh, Tan, & Lim, 2016; Yong & Lim, 2016). In particular, technological aids such as intelligent tutoring or feedback systems have been created to aid students' argumentative discourse in online environments. For instance, Wolfe, Britt, Petrovic, Albrecht, and Kopp (2009) have developed a web-based counterargument tutor that utilizes interactive exercises to guide university undergraduates in identifying counterarguments and responses to counterarguments in didactic texts. In argumentation-based computer-supported collaborative learning, a diverse variety of knowledge representation tools (e.g., diagrams), prompts (e.g., sentence openers), and computer-supported scripts (e.g., argumentative scripts that guide students to formulate and structure well-elaborated arguments) have also been used to orchestrate argumentation (for reviews, see Noroozi, Weinberger, Biemans, Mulder, & Chizari, 2012; Scheuer, Loll, Pinkwart, & McLaren, 2010). To provide further support when guiding students' integrative argumentation in online learning contexts, making JOLs+ in concert with retrieval practice can be incorporated as a complementary tool. For example, review questions and JOLs+ prompts can be provided at regular intervals, whereby students can be periodically asked to not only recall knowledge that they had previously been taught but to also judge their level of preparedness to critically position, use, and evaluate such knowledge in working toward their targeted learning outcomes.

Future Directions

While our research has focused on dialectical theory in argumentation due to its educational relevance in fostering balanced reasoning and critical thinking (e.g., Nussbaum, 2008, 2011), it must be noted that there are diverse types of argumentation rooted in more than one argumentation theory. For instance, informal logic approaches such as Johnson and Blair's (2006) criteria focus on the extent to which arguments are relevant, acceptable, and sufficient (RAS), while rhetorical approaches (e.g., Perelman & Olbrechts-Tyteca, 1969) emphasize argument schemes based on the principles of association (connecting elements previously considered as separate) and dissociation (splitting something previously considered whole into separate elements; for a discussion, see van Eemeren et al., 2014). To provide more robust support for the efficacy of retrieval practice and JOLs+ in enhancing argumentation, future studies may investigate the use of this learning strategy in facilitating performance across both informal logic and rhetorical approaches as well.

At the same time, given the efficacy of JOLs+ in enhancing the complex educational outcome of integrative argumentation, it may be useful to examine how JOLs+ can similarly be applied in conjunction with retrieval practice to promote various other higher order learning outcomes. Indeed, the JOLs+ questions we have designed in the present research can be readily adapted depending on the desired learning objective at hand. For instance, higher order thinking has been conceived to include creating novel products (e.g., Bloom, 1956). This may encompass generating creative research questions and hypotheses, such as through integrating the (potentially conflicting) outcomes of multiple past studies (for a review, see McGuire, 1997). Accordingly, after having studied the literature in a particular field, learners can be asked to practice recalling the advances made in that area from memory, responding to JOLs+ that specifically target the critical elements of hypothesis generation (e.g., addressing mixed evidence, and creating novel solutions to reconcile conflicting findings), before reviewing and integrating the literature to propose their own original research questions. This remains an intriguing possibility to be pursued in future studies.

In addition, while our study has demonstrated the learning benefits of retrieval practice when paired with JOLs+, further studies ought to uncover the specific psychological mechanisms underlying these benefits. To more directly probe these mechanisms, future work may consider incorporating think-aloud procedures to examine potential qualitative changes in the strategies or approaches that learners adopt after having responded to JOLs+. The potential changes in learners' allocation and engagement of attention during restudy after JOLs+ can also be assessed through, for instance, eye-tracking methods or by measuring their reading times. The data arising from such measures would be useful in illuminating the changes in processing that undergird learners' subsequent performance, thus contributing to a fuller theoretical understanding of how exactly retrieval practice supplemented with JOLs+ enhances higher order learning outcomes.

Furthermore, the question of when JOLs+ should specifically be implemented during the learning process to optimize its benefits remains an open one. In particular, JOLs+ in the present research were administered after learners had read the argumentative text and had engaged in retrieval practice or notetaking. However, JOLs+ can potentially also be delivered at an earlier stage of the learning process, such as at the outset before reading a text (during which learners judge their expected rather than actual higher order learning) or prior to practicing retrieval after having read the text, such that the desired higher order learning objective is consistently reinforced throughout learning. Accordingly, investigating the effects of introducing JOLs+ at different learning stages, in conjunction with testing its efficacy across different age groups, learning domains, and educational materials, may shed further light on the parameters of the utility of JOLs+.

Conclusion

The present research has demonstrated that, when augmented with a metacomprehension monitoring intervention (JOLs+) that orients learners' attention toward the critical elements of the intended higher order learning outcome, retrieval practice can effectively and efficiently enhance learners' integrative argumentation. Our findings offer implications for the interplay of memory and attention in complex learning, while illuminating some limits of the well-established retrieval-based learning strategy. By further showing how these limits can be overcome with JOLs+, this research aims to serve as a stepping stone toward guiding educational practice in encouraging learners to thoughtfully consider both sides of an issue, and toward developing a more profound understanding of the diverse learning contexts in which retrieval practice can be augmented to become even more powerful.

References

Asterhan, C. S. C., & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist, 51,* 164–187. http://dx.doi.org/10.1080/00461520.2016.1155458

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., . . . Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74,* 379–439. http://dx.doi.org/10.3102/00346543074003379

Bloom, B. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: McKay.

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

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119,* 197–253. http://dx.doi.org/10.1037/0033-2909.119.2.197

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306–307. http://dx.doi.org/10.1207/s15327752jpa4803_13

Cacioppo, J. T., Petty, R. E., & Morris, K. J. (1983). Effects of need for cognition on message evaluation, recall, and persuasion. *Journal of Personality and Social Psychology, 45,* 805–818. http://dx.doi.org/10.1037/0022-3514.45.4.805

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., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology, 23,* 760–771. http://dx.doi.org/10.1002/acp.1507

Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III. (2006). retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135,* 553–571. http://dx.doi.org/10.1037/0096-3445.135.4.553

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16,* 228–232. http://dx.doi.org/10.1111/j.1467-8721.2007.00509.x

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Duschl, R. A. (2007). Quality argumentation and epistemic criteria. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 159–175). Dordrecht, The Netherlands: Springer. http://dx.doi.org/10.1007/978-1-4020-6670-2_8

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education, 88,* 915–933. http://dx.doi.org/10.1002/sce.20012

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. Retrieved from ERIC database. (ED315423).

Golanics, J. D., & Nussbaum, E. M. (2008). Enhancing collaborative online argumentation through question elaboration and goal instructions. *Journal of Computer Assisted Learning, 24,* 167–180. http://dx.doi.org/10.1111/j.1365-2729.2007.00251.x

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test . . . or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review, 26,* 307–329. http://dx.doi.org/10.1007/s10648-013-9248-9

Johnson, R. H., & Blair, J. A. (2006). *Logical self-defense*. New York, NY: International Debate Education Association.

Karpicke, J. D. (2012). retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science, 21,* 157–163. http://dx.doi.org/10.1177/0963721412443552

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. T. Wixted (Ed.) & J. H. Byrne (Series Ed.), *Cognitive psychology of memory, Vol. 2 of Learning and memory: A comprehensive reference* (pp. 487–514). Oxford, United Kingdom: Academic Press. http://dx.doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27,* 317–326. http://dx.doi.org/10.1007/s10648-015-9309-3

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331,* 772–775. http://dx.doi.org/10.1126/science.1199327

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163–182. http://dx.doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes, 39,* 125–128. http://dx.doi.org/10.1207/s15326950dp3902&3_2

Kobayashi, K. (2010). Strategic use of multiple texts for the evaluation of arguments. *Reading Psychology, 31,* 121–149. http://dx.doi.org/10.1080/02702710902754192

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370. http://dx.doi.org/10.1037/0096-3445.126.4.349

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15,* 309–328. http://dx.doi.org/10.1016/S0885-2014(00)00030-7

Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition, 43,* 85–98. http://dx.doi.org/10.3758/s13421-014-0453-7

Loh, K. K., Tan, B. Z. H., & Lim, S. W. H. (2016). Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Computers in Human Behavior, 63,* 943–947. http://dx.doi.org/10.1016/j.chb.2016.06.030

Manz, E. (2015). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research, 85,* 553–590. http://dx.doi.org/10.3102/0034654314558490

Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief

change. *Contemporary Educational Psychology, 29,* 103–128. http://dx.doi.org/10.1016/j.cedpsych.2004.01.001

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16,* 492–509. http://dx.doi.org/10.1016/j.learninstruc.2006.09.007

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice, 41,* 226–232. http://dx.doi.org/10.1207/s15430421tip4104_4

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science, 20,* 516–522. http://dx.doi.org/10.1111/j.1467-9280.2009.02325.x

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

McGuire, W. J. (1997). Creative hypothesis generating in psychology: Some useful heuristics. *Annual Review of Psychology, 48,* 1–30. http://dx.doi.org/10.1146/annurev.psych.48.1.1

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology, 68,* 465–489. http://dx.doi.org/10.1146/annurev-psych-010416-044022

Millis, K. K., Simon, S., & tenBroek, N. S. (1998). Resource allocation during the rereading of scientific texts. *Memory & Cognition, 26,* 232–246. http://dx.doi.org/10.3758/BF03201136

Muller Mirza, N., & Perret-Clermont, A.-N. (Eds.). (2009). *Argumentation and education: Theoretical foundations and practices*. Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-0-387-98125-3

Nguyen, K., & McDaniel, M. A. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied, 22,* 59–71. http://dx.doi.org/10.1037/xap0000066

Noroozi, O., Weinberger, A., Biemans, H. J. A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer-supported collaborative learning (ABCSCL): A synthesis of 15 years of research. *Educational Research Review, 7,* 79–106. http://dx.doi.org/10.1016/j.edurev.2011.11.006

Novak, J. D., & Gowin, B. (1984). *Learning how to learn*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139173469

Nussbaum, E. M. (2005). The effect of goal instructions and need for cognition on interactive argumentation. *Contemporary Educational Psychology, 30,* 286–313. http://dx.doi.org/10.1016/j.cedpsych.2004.11.002

Nussbaum, E. M. (2008). Using argumentation vee diagrams (AVDs) for promoting argument-counterargument integration in reflective writing. *Journal of Educational Psychology, 100,* 549–565. http://dx.doi.org/10.1037/0022-0663.100.3.549

Nussbaum, E. M. (2011). Argumentation, dialogue theory, and probability modeling: Alternative frameworks for argumentation research in education. *Educational Psychologist, 46,* 84–106. http://dx.doi.org/10.1080/00461520.2011.558816

Nussbaum, E. M., & Edwards, O. V. (2011). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences, 20,* 443–488. http://dx.doi.org/10.1080/10508406.2011.564567

Nussbaum, E. M., & Kardash, C. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology, 97,* 157–169. http://dx.doi.org/10.1037/0022-0663.97.2.157

Nussbaum, E. M., & Schraw, G. (2007). Promoting argument-counterargument integration in students' writing. *Journal of Experimental Education, 76,* 59–92. http://dx.doi.org/10.3200/JEXE.76.1.59-92

Nussbaum, E. M., Winsor, D. L., Aqui, Y. M., & Poliquin, A. M. (2007). Putting the pieces together: Online argumentation vee diagrams enhance thinking during discussions. *Computer-Supported Collaborative Learning, 2,* 479–500. http://dx.doi.org/10.1007/s11412-007-9025-1

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144,* 710–756. http://dx.doi.org/10.1037/bul0000151

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation* (J. Wilkinson & P. Weaver, Trans.). Notre Dame, IN: University of Notre Dame Press.

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 83–105). Hillsdale, NJ: Erlbaum.

Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science, 11,* 481–518. http://dx.doi.org/10.1207/s15516709cog1104_4

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17,* 382–395. http://dx.doi.org/10.1037/a0026252

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). 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

Roediger, H. L., III, & Karpicke, J. D. (2006b). 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, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory & Cognition, 1,* 242–248. http://dx.doi.org/10.1016/j.jarmac.2012.09.002

Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *Computer-Supported Collaborative Learning, 5,* 43–102. http://dx.doi.org/10.1007/s11412-009-9080-x

Schwartz, B. L., & Metcalfe, J. (2017). Metamemory: An update of critical findings. In J. Wixted (Ed.) & J. H. Byrne (Series Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference* (2nd ed.; pp. 423–432). Oxford, United Kingdom: Academic Press. http://dx.doi.org/10.1016/B978-0-12-809324-5.21056-0

Schwarz, B. B., & Baker, M. J. (2017). *Dialogue, argumentation and education: History, theory and practice*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/9781316493960

Smith, M. A., Blunt, J. R., Whiffen, J. W., & Karpicke, J. D. (2016). Does providing prompts during retrieval practice improve learning? *Applied Cognitive Psychology, 30,* 544–553. http://dx.doi.org/10.1002/acp.3227

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory, 22,* 784–802. http://dx.doi.org/10.1080/09658211.2013.831454

Stapleton, P. (2001). Assessing critical thinking in the writing of Japanese university students. *Written Communication, 18,* 506–548. http://dx.doi.org/10.1177/0741088301018004004

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition, 32,* 696–710. http://dx.doi.org/10.3758/BF03195860

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66–73. http://dx.doi.org/10.1037/0022-0663.95.1.66

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85–106). New York, NY: Routledge.

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *The British Journal of Educational Psychology, 81,* 264–273. http://dx.doi.org/10.1348/135910710X510494

U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies.* Washington, DC. Retrieved from www.ed.gov/about/offices/list/opepd/ppss/reports.html

van Eemeren, F. H. (2010). *Strategic maneuvering in argumentative discourse: Extending the pragma-dialectical theory of argumentation*. Amsterdam, the Netherlands: John Benjamins Publishing. http://dx.doi.org/10.1075/aic.2

van Eemeren, F. H., Garssen, B., Krabbe, E. C. W., Henkemans, A. F. S., Verheij, B., & Wagemans, J. H. M. (2014). *Handbook of argumentation theory*. Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/978-90-481-9473-5

van Eemeren, F. H., & Grootendorst, R. (1992). *Argumentation, communication, and fallacies: A pragma-dialectical perspective*. Hillsdale, NJ: Erlbaum.

van Eemeren, F. H., Grootendorst, R., Henkemans, F. S., Blair, J. A., Johnson, R. H., Krabbe, E. C. W., . . . Zarefsky, D. (1996). *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Routledge.

van Eemeren, F. H., & Houtlosser, P. (2003). The development of the pragma-dialectical approach to argumentation. *Argumentation, 17,* 387–403. http://dx.doi.org/10.1023/A:1026338402751

Vorobej, M. (2006). *A theory of argument*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511498879

Voss, J. F., & Van Dyke, J. A. (2001). Argumentation in psychology: Background comments. *Discourse Processes, 32,* 89–111. http://dx.doi.org/10.1207/S15326950DP3202&3_01

Walton, D. N. (1996). *Argumentation schemes for presumptive reasoning*. Mahwah, NJ: Erlbaum.

Walton, D. (2007). *Dialog theory for critical argumentation*. Philadelphia, PA: John Benjamins. http://dx.doi.org/10.1075/cvs.5

Walton, D. (2010). A dialogue model of belief. *Argument and Computation, 1,* 23–46. http://dx.doi.org/10.1080/19462160903494576

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132,* 408–428. http://dx.doi.org/10.3200/GENP.132.4.408-428

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28,* 193–219. http://dx.doi.org/10.1177/0741088311399236

Wolfe, C. R., & Britt, M. A. (2008). Locus of the myside bias in written argumentation. *Thinking & Reasoning, 14,* 1–27. http://dx.doi.org/10.1080/13546780701527674

Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication, 26,* 183–209. http://dx.doi.org/10.1177/0741088309333019

Wolfe, C. R., Britt, M. A., Petrovic, M., Albrecht, M., & Kopp, K. (2009). The efficacy of a Web-based counterargument tutor. *Behavior Research Methods, 41,* 691–698. http://dx.doi.org/10.3758/BRM.41.3.691

Wong, S. S. H., & Lim, S. W. H. (2019). Prevention–permission–promotion: A review of approaches to errors in learning. *Educational Psychologist, 54,* 1–19. http://dx.doi.org/10.1080/00461520.2018.1501693

Wong, S. S. H., Ng, G. J. P., Tempel, T., & Lim, S. W. H. (2017). Retrieval practice enhances analogical problem-solving. *Journal of Experimental Education*. Advance online publication. http://dx.doi.org/10.1080/00220973.2017.1409185

Yong, P. Z., & Lim, S. W. H. (2016). Observing the testing effect using Coursera video-recorded lectures: A preliminary study. *Frontiers in Psychology: Cognition, 6,* 2064. http://dx.doi.org/10.3389/fpsyg.2015.02064

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 August 21, 2018
Revision received February 26, 2019
Accepted February 28, 2019 ■