The concreteness effect on judgments of learning: Evaluating the contributions of fluency and beliefs

Amber E. Witherby[1] · Sarah K. Tauber[1]

Abstract Researchers have often determined *how* cues influence judgments of learning (JOLs; e.g., concrete words are assigned higher JOLs than are abstract words), and recently there has been an emphasis in understanding *why* cues influence JOLs (i.e., the mechanisms that underlie cue effects on JOLs). The analytic-processing (AP) theory posits that JOLs are constructed in accordance with participants' beliefs of how a cue will influence memory. Even so, some evidence suggests that fluency is also important to cue effects on JOLs. In the present experiments, we investigated the contributions of participants' beliefs and processing fluency to the concreteness effect on JOLs. To evaluate beliefs, participants estimated memory performance in a hypothetical experiment (Experiment 1), and studied concrete and abstract words and made a pre-study JOL for each (Experiments 2 and 3). Participants' predictions demonstrated the belief that concrete words are more likely to be remembered than are abstract words, consistent with the AP theory. To evaluate fluency, response latencies were measured during lexical decision (Experiment 4), self-paced study (Experiment 5), and mental imagery (Experiment 7). Number of trials to acquisition was also evaluated (Experiment 6). Fluency did not differ between concrete and abstract words in Experiments 5 and 6, and it did not mediate the concreteness effect on JOLs in Experiments 4 and 7. Taken together, these results demonstrate that beliefs are a primary mechanism driving the concreteness effect on JOLs.

Students learn a large amount of information, some of which is concrete and some of which is abstract. For example, when learning about the structure of the brain, students can use model replicas to memorize where the brain structures are located. That is, they can use a concrete, imageable stimulus to aid their understanding. In contrast, when students attempt to learn abstract information, such as the differences between structuralism and functionalism, they cannot rely on a concrete representation. Importantly, concrete words (e.g., chair) are remembered more than are abstract words (e.g., equal), and participants predict (via judgments of learning; JOLs) that they will recall more concrete words relative to abstract words (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Tauber & Rhodes, 2012; Tullis & Benjamin, 2012). Although it is clear how concreteness influences JOLs (i.e., participants assign higher JOLs to concrete words relative to abstract words), little is known about *why* concreteness influences JOLs. In other words, the mechanisms that drive the concreteness effect on JOLs are not well understood, and filling this gap is the primary goal of the current experiments.

According to contemporary theory of monitoring of learning (i.e., the cue-utilization approach), JOLs are inferential and based on cues (e.g., concreteness) that are available at the time of their construction (Koriat, 1997). That is, during study participants attend to distinguishable features (i.e., cues) that can help them predict later memory performance. Importantly, any cue effect on JOLs may be driven by two non-mutually exclusive mechanisms: beliefs and fluency (e.g., Dunlosky, Mueller, & Tauber, 2015; Koriat, 1997). Recent evidence suggests that many cue effects on JOLs are

✉ Amber E. Witherby
  amber.witherby@tcu.edu

[1] Department of Psychology, Texas Christian University, 2800 S. University Dr, TCU Box 298920, Fort Worth, TX 76129, USA

Springer

largely driven by naïve beliefs (e.g., Hu et al., 2015; Jia et al., 2016; Kornell, 2015; Li, Jia, Li, & Li, 2016; Mueller, Dunlosky, & Tauber, 2016; Mueller, Dunlosky, Tauber, & Rhodes, 2014; Mueller, Tauber, & Dunlosky, 2013; Susser, Jin, & Mulligan, 2016). However, in some cases researchers have found that fluency partially mediates the relationship between a given cue and JOLs (e.g., Besken, 2016; Besken & Mulligan 2014; Undorf & Erdfelder, 2015), suggesting that processing fluency *and* beliefs contribute to cue effects on JOLs. Thus, it is possible that beliefs, fluency, or a combination of the two drive the concreteness effect on JOLs. We begin first by considering the contribution of beliefs to the concreteness effect on JOLs as suggested by the analytic-processing (AP) theory, followed by discussing the contribution of processing fluency.

According to AP theory and consistent with the cue-utilization approach to JOLs, people's beliefs about memory play a central role in the construction of JOLs. Specifically, AP theory states that when participants are instructed to make JOLs, they search for distinguishable characteristics about the study material that can help them reduce their uncertainty about future recall. They then use this information to create, or update, a belief about how that information will influence memory. Finally, they use that belief to construct their JOL (Dunlosky et al., 2015; Mueller et al., 2016). For instance, when participants are presented with a list containing concrete and abstract words, they use their pre-existing beliefs (or beliefs developed while studying) that concreteness will influence their memory as a basis for their JOLs. As a result, if participants believe that concrete words will be easier to recall than abstract words will be, they will assign higher JOLs to them.

AP theory has received support when examining the influence of other cues on JOLs (e.g., Hu et al., 2015; Jia et al., 2016; Li et al., 2016; Mueller et al., 2013; Mueller et al., 2014; Mueller et al., 2016). For instance, Mueller et al. (2013) had participants read a hypothetical experiment in which students studied a series of related (e.g., loaf – bread) and unrelated (e.g., dog – spoon) word pairs. Participants predicted the number of each type of word pair that students in the hypothetical experiment recalled. Participants demonstrated a strong belief that more related word pairs would be remembered relative to unrelated word pairs (mean predictions of 68% and 36%, respectively). To further evaluate beliefs about the influence of relatedness on memory, Mueller et al. (2013) had participants study related and unrelated word pairs and make a JOL for each. JOLs were made by estimating the likelihood of future recall on a scale from 0% (will not recall) to 100% (will absolutely recall). Importantly, one group of participants made their JOLs immediately after word pairs were presented (i.e., immediate JOL), whereas a second group made their JOLs immediately *before* studying the pairs (i.e., prestudy JOL; see Castel, 2008). The prestudy JOL group is of particular interest because participants do not have access to the pairs while making their JOLs. As such, it is likely that they relied on their beliefs about how that item type would influence memory. Participants' immediate JOLs were higher for related pairs relative to unrelated pairs. Most important, consistent with AP theory, participants' prestudy JOLs were higher for related pairs relative to unrelated pairs.

Although beliefs play an important role in cue effects on JOLs, it is also possible that JOLs are influenced by how fluently items are processed during study (e.g., Begg et al., 1989; Hertzog et al., 2003; Koriat, 1997), and research has supported this assertion (Besken, 2016; Besken & Mulligan, 2014; Susser & Mulligan, 2015, Experiment 2; Undorf & Erdfelder, 2015). For instance, related word pairs may be processed more fluently than unrelated word pairs, leading to higher JOLs for related word pairs relative to unrelated word pairs. Undorf and Erdfelder (2015) evaluated this possibility by presenting participants with related and unrelated word pairs and measuring the number of study trials necessary for participants to successfully recall the second word of the pair when given the first (i.e., trials to acquisition). From a fluency-based account of JOLs, it should take fewer trials to learn the related word pairs relative to unrelated word pairs, which is what they found. Moreover, the number of trials to acquisition partially mediated the relationship between relatedness and JOLs. Using the same logic, Undorf and Erdfelder (2015) further assessed processing fluency by allowing participants to self-pace their study of related and unrelated word pairs. Self-paced study times were faster for related word pairs than for unrelated word pairs, and for one list of pairs, study time partially mediated the relationship between relatedness and JOLs. Specifically, fluency accounted for a percentage (3–26%) of the relatedness effect on JOLs.

Thus, the fluency account of JOLs has received support; however, in some cases, researchers have found that fluency does not contribute to cue effects on JOLs (e.g., Jia et al., 2016; Li et al., 2016; Mueller et al., 2013; Mueller et al., 2014; Mueller et al., 2016). The conflicting results may occur for multiple reasons. One possibility is that fluency could be more important for one cue than it is for another. Another possibility is that the measure used to evaluate fluency can play an important role. For example, Undorf and Erdfelder (2015) found fluency to be an important contributor to the relatedness effect on JOLs when using trials to acquisition and self-paced study times; however, Mueller et al. (2013) did not find fluency to contribute to the relatedness effect on JOLs when using response times in a lexical decision task. To thoroughly evaluate the contribution of fluency to the concreteness effect on JOLs we measured fluency in four ways (i.e., response times in a lexical decision task, self-paced study times, trials to acquisition, and mental image latency).

In sum, the previous research that has evaluated the mechanisms that drive cue effects on JOLs has been mixed, with evidence to support both the belief-based and fluency-based accounts. In the present experiments we investigated the degree to which these mechanisms contribute to the concreteness effect on JOLs. To evaluate participants' beliefs, in Experiment 1 we used a hypothetical experiment, which has previously been used to establish memory beliefs (e.g., Koriat, Bjork, Sheffer, & Bar, 2004; Kornell, Rhodes, Castel, & Tauber, 2011). In Experiments 2 and 3, beliefs were evaluated when participants learned concrete and abstract words. To evaluate fluency, we measured lexical decision response times (Experiment 4), self-paced study times (Experiment 5), the number of trials to acquisition (Experiment 6), and the latency to form a mental image (Experiment 7).

Experiment 1

Experiment 1 was designed to evaluate participants' beliefs about the influence of concreteness on memory. To do so, participants read a description of a hypothetical experiment and estimated memory performance for concrete and abstract words. If participants hold accurate beliefs about concreteness, then their estimates should be higher for concrete words relative to abstract words.

Method

Participants

Forty undergraduate students from Texas Christian University participated for partial course credit in psychology courses. In Experiment 1 (and in the following experiments) sample sizes were selected based on previous research with similar methodology (e.g., Koriat et al., 2004; Mueller et al., 2013).

Materials and procedure

Participants read the following description of a hypothetical experiment:

> "In a previous experiment that we conducted, students were presented with a list of 40 words one after the other. Critically, half of the words (i.e., 20) were **concrete** and the other half (i.e., 20) were **abstract**. Each of the words was presented for 5 seconds. The students' task was to study these words so that they could remember as many words as possible on a later memory test. This memory test took place immediately after studying all of the words, and students were asked to recall as many words as possible from the list they had previously studied. We would like you to estimate how many **concrete** and **abstract** words the students remembered. Your estimates can range from 0 to 20 for each item type. An estimate of 0 means that the students did not remember any of the words, whereas an estimate of 20 means that students remembered all of the words of that type."

Word type (i.e., abstract or concrete) was counterbalanced such that each appeared equally often in the first and second positions. Additionally, to ensure that participants understood the terms abstract and concrete, half were given an example of both a concrete (i.e., table) and an abstract (i.e., loyalty) word in the description. The other half of participants were not provided with examples to prevent them from experiencing any processing fluency for abstract and concrete words. Participants were given unlimited time to write their estimates.

Results

Participants believed that concrete words would be more memorable than abstract words. Specifically, participants' estimates were higher for concrete words than for abstract words, $F(1, 38) = 51.28$, $p < .001$, $\eta_p^2 = .57$ (see Table 1). The interaction between word type and presence of examples was significant, $F(1, 38) = 6.78$, $p = .01$, $\eta_p^2 = .15$, because the difference in estimates was smaller when participants were given examples, $t(19) = 3.28$, $p = .004$, $d = .63$, relative to when they were not, $t(19) = 6.79$, $p < .001$, $d = 1.12$. Thus, participants may have been influenced by the examples provided in the hypothetical experiment (cf. Mueller et al., 2014). Even so, it is unclear whether participants may be influenced by the inclusion of examples generally, or by the specific examples provided; thus, the interaction should be replicated prior to drawing conclusions about it. More important, the interaction is in the opposite direction than would be predicted from the fluency hypothesis. The main effect of example was not significant, $F(1, 38) = 2.64$, $p = .11$, $\eta_p^2 = .07$.

Experiment 2

Participants in Experiment 1 believed that concrete words are more memorable than are abstract words. In fact, 75% of the

Table 1 Mean estimated recall (out of 20) from Experiment 1

	Abstract	Concrete
Example	7.40 (.59)	9.50 (.85)
No Example	7.70 (.59)	12.20 (.85)
Overall	7.55 (.42)	10.85 (.60)

Note. Within-participant standard errors of the mean are in parentheses (Loftus & Masson, 1994)

participants provided higher estimates for concrete words relative to abstract words, demonstrating that participants apply their beliefs about concreteness on a global scale (i.e., when estimating overall performance). However, this does not reveal whether participants use beliefs about concreteness on an item-by-item basis. As such, in Experiment 2 participants made JOLs immediately before studying each word (i.e., prestudy JOLs; Castel, 2008). If participants apply beliefs about concreteness at the item level, then their prestudy JOLs will be higher for concrete words relative to abstract words. A second group of participants made JOLs immediately after studying each word (i.e., immediate JOLs) and participants in both groups took a free-recall test.

Method

Participants and design

A 2 (word type: abstract and concrete) × 2 (judgment group: immediate JOL vs. prestudy JOL) mixed design was used, with word type manipulated within-participant and judgment group manipulated between-participants. Eighty undergraduate students from Texas Christian University participated for partial course credit in psychology courses. Participants were randomly assigned to either an immediate JOL ($n = 40$) or a prestudy JOL ($n = 40$) group (cf. Mueller et al., 2013; Undorf & Erdfelder, 2015).

Materials and procedure

Participants studied a list of 30 words (15 concrete; 15 abstract) obtained from Paivio, Yuille, and Madigan (1968). Ratings of concreteness were greater for concrete words ($M = 6.73$) than for abstract words ($M = 2.34$), $t(28) = 23.46$, $p < .001$. Concrete and abstract words did not differ based on word frequency (Log_Freq_Hal; $M = 9.67$ and $M = 9.82$ respectively), length ($M = 6.13$ letters and $M = 6.20$ letters respectively), number of syllables ($M = 1.73$ and $M = 2.13$ respectively), orthography ($M = 2.20$ and $M = 1.73$ respectively), or familiarity ($M = 520.20$ and $M = 542.40$ respectively), $ts < 1.70$.

Each word was presented for 5 s, and presentation order was randomized anew for each participant with the constraint that no more than three of each word type were presented in a row. All participants made a self-paced JOL for each word. JOLs were made on a scale from 0–100%, with 0 indicating that the participant would definitely not recall the word and 100 indicating that the participant would absolutely recall the word. In the prestudy group participants made their JOL before seeing each word and were given the following prompt: "The word you are about to study is concrete (abstract). Please rate the likelihood that you will remember it on a later memory test." In the immediate JOL group, after studying each word

participants were told whether that word was concrete or abstract, and made their JOL. After study and JOLs, participants in both groups took a free-recall test on which they were given unlimited time to recall as many words as they could remember. Participants were told to recall the words in any order. Small spelling errors that did not change the meaning of the words (e.g., pluralizing a word) were counted as correct.

Results

Our primary interest was to evaluate JOLs; as such, analyses of JOLs are reported first followed by analyses of memory performance.

Judgments of learning

As evident in Fig. 1 (left panel), both judgment groups provided higher JOLs for concrete words than for abstract words; thus, participants in the prestudy group applied their beliefs about concreteness on an item-by-item basis when constructing their JOLs. Additionally, participants in the immediate group replicated the typical concreteness effect on immediate JOLs. These conclusions are supported by a main effect of word type, $F(1, 78) = 17.48$, $p < .001$, $\eta_p^2 = .18$, indicating that across groups, participants provided higher JOLs for concrete words compared to abstract words. There was also a main effect of judgment group, $F(1, 78) = 6.27$, $p = .014$, $\eta_p^2 = .01$, indicating that participants in the immediate group provided higher JOLs relative to the prestudy group. These effects were qualified by a significant two-way interaction, $F(1, 78) = 4.09$, $p = .047$, $\eta_p^2 = .05$, demonstrating that the difference in JOLs between abstract and concrete words was greater in the immediate group, $t(39) = 3.66$, $p = .001$, $d = .38$, relative to the prestudy group, $t(39) = 2.02$, $p = .048$, $d = .12$. Even so, the results from the prestudy JOL group provide support for the hypothesis that participants have beliefs about
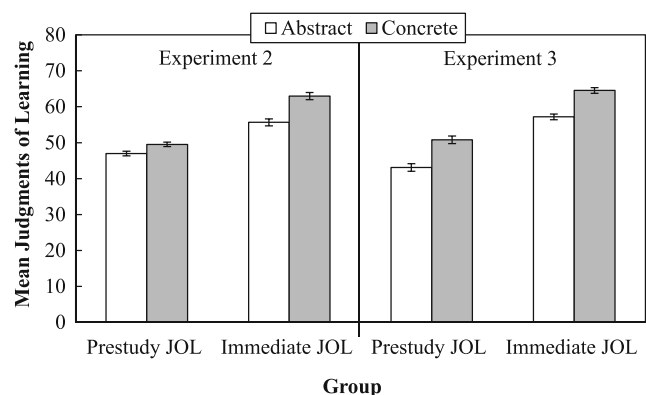

Fig. 1 Mean judgments of learning for each word type and group in Experiments 2 (left panel) and 3 (right panel). Error bars represent within-participant standard error (Loftus & Masson, 1994)

how concreteness will influence memory, and that they use these beliefs when constructing their JOLs.

Memory performance

Consistent with previous work (e.g., Tauber & Rhodes, 2012), participants in both groups recalled more concrete words than abstract words, $F(1, 78) = 27.34$, $p < .001$, $\eta_p^2 = .26$ (see Table 2). The main effect of group and the interaction were not significant, $Fs < 1$.

Experiment 3

The results from Experiment 2 demonstrated that participants used beliefs about concreteness to inform their JOLs on an item-by-item basis. However, the concreteness effect on participants' prestudy JOLs was relatively small ($d = .12$), and the effect was larger on participants' immediate JOLs ($d = .38$). Thus, it is possible that the effect is partially driven by other mechanisms (e.g., fluency). Specifically, the effect may be larger on immediate JOLs relative to prestudy JOLs because participants could use their belief that concrete words would be better remembered *and* information gained from processing fluency during learning. In contrast, participants in the prestudy JOL group were only able to rely on their beliefs as a basis for their judgments.

Experiment 2 was the first experiment to evaluate the concreteness effect with prestudy JOLs; thus, our primary goal in Experiment 3 was to replicate the concreteness effect on them (for detailed discussions on the importance of replication, see Lishner, 2015; Pashler & Harris, 2012). Thus, Experiment 3 was identical to Experiment 2, with the exception that

Table 2 Mean proportion correctly recalled in Experiment 2 through Experiment 7

	Abstract	Concrete	Non-word
Experiment 2			
Prestudy JOL Group	.29 (.01)	.36 (.01)	-
Immediate JOL Group	.31 (.01)	.40 (.01)	-
Experiment 3			
Prestudy JOL Group	.34 (.01)	.42 (.01)	-
Immediate JOL Group	.29 (.01)	.40 (.01)	-
Experiment 4	.14 (.01)	.24 (.01)	.03 (.01)
Experiment 5	.35 (.01)	.40 (.01)	-
Experiment 6	.62 (.01)	.69 (.01)	-
Experiment 7	.37 (.01)	.43 (.01)	-

Note. Within-participant standard errors of the mean are in parentheses (Loftus & Masson, 1994). Only immediate JOLs were used in Experiments 4–7. Non-words were only used in Experiment 4

JOL judgment of learning

participants in Experiment 3 were instructed to make their judgments as accurate as possible. These instructions were used to ensure that participants focused on their JOLs, as they were the primary measure of interest.

Method

Participants and design

As in Experiment 2, a 2 (word type: abstract and concrete) × 2 (judgment group: immediate JOL vs. prestudy JOL) mixed design was used, with word type manipulated within-participant and judgment group manipulated between-participants. Eighty undergraduate students from Texas Christian University participated for partial credit in psychology courses. Participants were randomly assigned to either an immediate JOL ($n = 40$) or a prestudy JOL ($n = 40$) group.

Materials and procedure

The materials and procedure were identical to Experiment 2, except participants in both groups received the following additional instructions prior to beginning the experiment:

> "It is very important that you make accurate judgments. When you make each judgment think about the likelihood that you will remember that word in the future. Please take your time and try to make your judgments as accurate as possible."

The instructions were presented on the computer screen for participants to read.

Results

As in Experiment 2, analyses of JOLs are reported followed by analyses of memory performance.

Judgments of learning

Replicating Experiment 2, participants' JOLs in both judgment groups were higher for concrete words relative to abstract words (see Fig. 1, right panel). This was supported by a significant effect of word type, $F(1, 78) = 32.69$, $p < .001$, $\eta_p^2 = .30$. Additionally, participants' JOLs in the immediate group were higher relative to the prestudy group, $F(1, 78) = 10.36$, $p = .002$, $\eta_p^2 = .12$. The interaction was not significant, $F < 1$.

Memory performance

As evident from Table 2, participants recalled more concrete words than abstract words, $F(1, 78) = 34.84$, $p < .001$, $\eta_p^2 =$

.31. The main effect of group and the interaction were not significant, Fs < 1.

Experiment 4

Results from the first three experiments provide converging evidence that participants believe that concrete words are more memorable than are abstract words, and they use this belief to inform their JOLs. The effect on prestudy JOLs in Experiment 3 was much larger than the effect on prestudy JOLs in Experiment 2 (ds of .36 and .12 respectively), which could have occurred for a number of reasons. The larger effect in Experiment 3 could be attributable to participants' additional attention to their JOLs, to demand characteristics produced by the instructions, or to a sampling effect. In particular, participants were not randomly assigned to experiment, which makes the across-experiment comparison challenging to interpret. Even so, the results from the first three experiments provide evidence for the belief-based account of JOLs, supporting the AP theory.

Though beliefs clearly contribute to the concreteness effect on JOLs, it is possible that processing fluency contributes too, which may be evident with direct measures of processing fluency. From this perspective, concrete words are given higher JOLs than are abstract words because they are processed more easily. In Experiment 4, we evaluated this possibility using a lexical decision task, which has been used previously to measure processing fluency in JOL tasks (e.g., Mueller et al., 2013; Mueller et al., 2014). To do so, participants were presented with a series of words and non-words (e.g., vaded) and decided as quickly as possible whether each was a word or not. Participants then made an immediate JOL for each item and took a final free-recall test. Of most interest, if concrete words are processed more fluently than are abstract words, then response times should be shorter for them. Moreover, participants' response times should mediate the effect of concreteness on JOLs.

Method

Participants and design

Forty undergraduate students from Texas Christian University participated for partial course credit in their psychology classes. Word type (abstract, concrete, and non-word) was manipulated within-participant.

Materials and procedure

Materials included the same 30 abstract and concrete words from Experiment 2 as well as 30 non-words (e.g., vaded; taken from Mueller et al., 2014, Experiment 1). Non-words, abstract

words, and concrete words were equated on length (Ms = 6.03, 6.20, and 6.13 letters, respectively) and orthography (Ms = 2.03, 1.73, 2.20 respectively), ts < 1.

Participants were first told that they would study a series of words that would contain concrete words, abstract words, and non-words, and their task was to determine whether each was a word or non-word. Then, they were presented with each word one-at-a-time and were told to press "z" if it was a word and "m" if it was a non-word as quickly as possible while still being accurate. Prior to seeing each item participants were given 3 s to prepare by placing their fingers on the "z" and "m" keys, followed by a 1-s fixation point. During the lexical decision task, participants were not told whether each individual word was abstract or concrete to avoid distracting them from making their lexical judgments quickly and accurately. After making each lexical decision, participants made an immediate JOL (self-paced) and took a free-recall test as in Experiment 2.

Results

Our focus was on lexical decision response times, which are presented first, followed by analyses of immediate JOLs and memory performance.

Lexical decision response times

Cases in which participants made an incorrect lexical decision (e.g., designated a word as "non-word") were rare (6.80% of the trials), and were removed from analyses. Median response times were calculated for each participant.[1] Further, 39% of participants' responses were outliers (±2.5 SD of the mean). Even so, outcomes from analyses with the outliers included and with the outliers excluded were identical; thus we report analyses with the full set of data. As evident from Fig. 2, participants' response times for abstract words were shorter (i.e., faster) than for non-words, $t(39) = 6.23$, $p < .001$, $d = .95$. Likewise, response times were shorter for concrete words than for non-words, $t(39) = 6.91$, $p < .001$, $d = 1.06$. More importantly, response times were significantly shorter for concrete words than for abstract words, $t(39) = 3.63$, $p = .001$, $d = .32$. These conclusions were supported by a significant one-way repeated-measures ANOVA, $F(2,78) = 42.44$, $p < .001$, $\eta_p^2 = .52$. This indicates that processing fluency differed between abstract and concrete words; thus, mediational analyses can be conducted.

Researchers have traditionally evaluated within-participant mediation by comparing zero-order correlations with partial

---

[1] For all experiments that included response time measures (i.e., Experiments 4, 5, and 7), response times were calculated by averaging participants' mean responses and median responses. All results maintain with both measures, so median response times are reported.
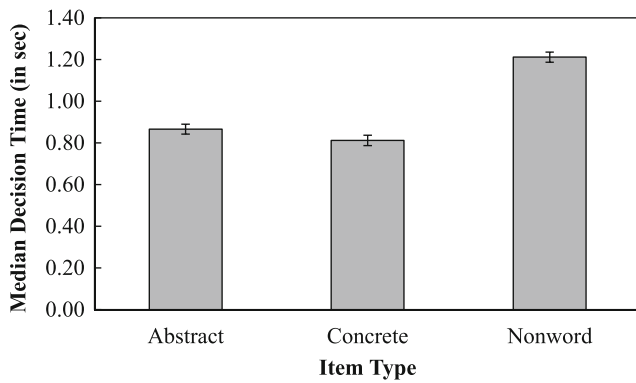
Fig. 2 Median response times in a lexical decision task (in sec) for abstract words, concrete words, and non-words in Experiment 4. Error bars represent within-participant standard error (Loftus & Masson, 1994)

correlations that control for the mediator (e.g., Hertzog et al., 2003; Mueller et al., 2013; Mueller et al., 2014). However, this can be problematic because the likelihood of committing a Type 1 error increases with additional tests (e.g., MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Montoya & Hayes, in press; Preacher & Hayes, 2004). To avoid this issue, statisticians recommend doing a single test of the indirect effect of the independent variable (i.e., concreteness) on the dependent variable (i.e., JOLs) through the mediator (i.e., response times). Thus, to evaluate whether fluency mediated the relationship between concreteness and JOLs, we conducted a within-participant mediation analysis with a newly developed macro (MEMORE; Montoya & Hayes, in press). Importantly, this macro also applies the bootstrapping procedure. Bootstrapping allows researchers to resample from the data to generate a sampling distribution for which an indirect effect and corresponding confidence interval can be attained. This is particularly beneficial for

within-participant designs because it can be a more powerful assessment of mediation with smaller samples (e.g., Lockwood & MacKinnon, 1998; Preacher & Hayes, 2004).

Thus, 1,000 bootstrap samples were performed and the 95% confidence interval obtained for the indirect effect of concreteness on JOLs through response times contained zero (−1.04, .83) and was thus non-significant (see Table 3). Therefore, processing fluency, as assessed with a lexical decision task, did not mediate the effect of concreteness on JOLs.

Judgments of learning

Participants' immediate JOLs were higher for concrete words ($M = 82.87$, $SE = 1.32$) relative to abstract words ($M = 79.77$, $SE = 1.02$), $t(39) = 3.54$, $p = .001$, $d = .18$, and for concrete words relative to non-words ($M = 35.41$, $SE = 2.20$), $t(39) = 13.82$, $p < .001$, $d = 1.53$. Further, participants' immediate JOLs were higher for abstract words relative to non-words, $t(39) = 14.19$, $p < .001$, $d = 1.49$. These observations were supported by a significant one-way ANOVA, $F(2,78) = 189.49$, $p < .001$, $\eta_p^2 = .83$.

Memory performance

A one-way ANOVA revealed a significant effect of word type, $F(2,78) = 65.24$, $p < .001$, $\eta_p^2 = .63$ (see Table 2). Specifically, recall was greater for concrete words relative to abstract words, $t(39) = 5.02$, $p < .001$, $d = .86$, and for concrete words relative to non-words, $t(39) = 11.21$, $p < .001$, $d = 1.54$. Finally, recall was greater for abstract words relative to non-words, $t(39) = 6.81$, $p < .001$, $d = 1.17$.

Table 3 Within-participant mediation models for Experiment 4 and Experiment 7

	b	SE	95% CI
Experiment 4			
The effect of concreteness on lexical decision times	.05*	.02	.02, .08
The effect of concreteness on JOLs	−3.05*	1.04	−5.16, −.94
The effect of lexical decision times on JOLs	−.95	9.73	−20.67, 18.77
Total indirect effect	−.05	.47	−1.04, .83
Experiment 7			
The effect of concreteness on image latencies	.33*	.10	.12, .53
The effect of concreteness on JOLs	−21.90*	2.32	−26.60, −17.21
The effect of image latencies on JOLs	5.75	3.47	−1.28, 12.78
Total indirect effect	1.88	1.53	−.17, 5.69

Note. Only immediate JOLs were used in Experiments 4 and 7. Each model included three direct effects and one indirect effect. The indirect effect in Experiment 4 evaluated whether concreteness influenced response times in a lexical decision task, which in turn influenced JOLs. The indirect effect in Experiment 7 evaluated whether concreteness influenced image latencies, which in turn influenced JOLs

* $p < .05$

JOL judgment of learning, *b* beta (the unstandardized regression coefficient), *SE* standard error, *CI* confidence interval

Experiment 5

In Experiment 4, response times in the lexical decision task were faster for concrete words relative to abstract words. However, this difference did not mediate the relationship between concreteness and JOLs. Although the inclusion of non-words is essential when using a lexical decision task, doing so resulted in a small concreteness effect on JOLs ($d = .18$), which could have constrained our ability to observe mediation. Thus, in Experiment 5 we evaluated fluency by allowing participants to self-pace their study. Self-paced study times have been established as a measure of processing fluency (e.g., Koriat, 2008; Undorf & Erdfelder, 2015), although it has not previously been used to assess the concreteness effect on JOLs. Participants made an immediate JOL after studying each word and received a free-recall test. Of most interest, if concrete words are processed more fluently than are abstract words then self-paced study times should be shorter for them. Further, if a fluency advantage for concrete words contributes to the concreteness effect on JOLs, then participants' self-paced study times should mediate the effect of concreteness on JOLs.

Method

Participants and design

Forty undergraduate students from Texas Christian University participated for partial credit in their psychology classes. Word type (abstract and concrete) was manipulated within-participant.

Materials and procedure

Materials were identical to Experiment 2. Each word was presented one at a time and presentation order was randomized anew for each participant, with the same constraints as in Experiment 2. Prior to beginning the experiment, participants were told that some of the words would be concrete and some would be abstract, and that their task was to study the words for a future test. Participants were given unlimited time to study each word and told to press the spacebar when they were finished. After studying each word, participants made a self-paced immediate JOL, after which they were given 3 s to prepare to study the next word by placing their fingers on the spacebar. The word type (concrete/abstract) was presented at the top of the screen while participants studied and made a JOL for each word. Finally, participants completed a self-paced free-recall test as in Experiment 2.

Results

Our primary interest was in self-paced study times, which are presented first. Next were analyses of JOLs and memory performance.

Self-paced study times

Median study times were calculated for each participant. There were few outliers, and the results were consistent after excluding them. Thus, we report on the full set of data. Study time for concrete ($Mdn = 4.14$ sec, $SE = .21$) and abstract ($Mdn = 4.26$ sec, $SE = .24$) words did not differ, $t(39) = .71$, $p = .48$. Thus, processing fluency, as measured via study times, did not differ based on concreteness.

Judgments of learning

As expected, participants' immediate JOLs were higher for concrete words ($M = 57.19$, $SE = .66$) relative to abstract words ($M = 51.54$, $SE = .66$), $t(39) = 4.37$, $p < .001$, $d = .24$.

Memory performance

Recall was greater for concrete words than for abstract words, $t(39) = 2.59$, $p = .013$, $d = .33$ (see Table 2).

Experiment 6

In Experiment 5, self-paced study times did not differ between concrete and abstract words. As such, mediational analyses exploring the indirect effect of study time on the concreteness effect on JOLs could not be conducted. These results are consistent with Mueller et al. (2014) and Mueller et al. (2016) who did not find self-paced study times to mediate the relationship between other cues (font size and identical word pairs, respectively) and JOLs (but see Undorf & Erdfelder, 2015). Undorf and Erdfelder (2015, Experiment 1) found an alternative measure of processing fluency to contribute to a cue effect on JOLs. Specifically, they demonstrated that it took fewer study trials for participants to recall related word pairs relative to unrelated word pairs, and this difference partially mediated the relationship between relatedness and JOLs. Adopting this logic, in Experiment 6, we assessed the contribution of fluency to the concreteness effect on JOLs by measuring how many study-test trials it took for participants to recall concrete and abstract words (i.e., trials to acquisition). If concrete words are processed more fluently than are abstract words, then it should take participants fewer trials to recall concrete words relative to abstract words. Further, this difference should mediate the relationship between concreteness and JOLs.

Method

Participants and design

Forty undergraduate students from Texas Christian University participated for partial credit in psychology courses. Word type (concrete and abstract) was manipulated within-participant.

Materials and procedure

Materials were identical to Experiment 2. Prior to beginning the task, participants were told that they would study a series of concrete and abstract words presented one at a time for 5 s each. Participants were informed that they would receive a test immediately after studying the words and that they would repeat the study-test cycle until they had correctly recalled each word. Presentation order was randomized anew for each participant with the same constraints as in Experiment 2, and participants were not told what type of word they were studying. After the initial study phase, participants took a self-paced free-recall test by verbally recalling each word that they could remember, and a research assistant recorded their responses. A verbal protocol was used to reduce participants' spelling mistakes. Spelling mistakes would have been problematic because correctly recalled words that were misspelled would be counted as incorrect in the program. Once the test was completed, all of the words that were successfully recalled were removed from the subsequent study list. Participants repeated this study-test procedure until they had recalled each word. Once participants had successfully recalled each word, they completed a final study phase in which each word was presented for 5 s, with presentation order randomized anew for each participant, and then made a self-paced immediate JOL. Word type was not presented during study or when participants made JOLs. Finally, participants took a self-paced free-recall test.

Results

Our primary interest was in the number of trials to acquisition, which are discussed first. Next were analyses of JOLs and memory performance.

Trials to acquisition

The number of trials to acquisition did not differ for concrete ($M = 2.04$, $SE = .03$) and abstract words ($M = 2.09$, $SE = .03$), $t < 1$. Thus, mediational analyses could not be conducted.

Judgments of learning

Consistent with the previous experiments, participants' JOLs were higher for concrete words ($M = 61.44$, $SE = 1.04$) than for abstract words ($M = 51.37$, $SE = 1.04$), $t(39) = 4.92$, $p < .001$, $d = .67$.

Memory performance

Recall was greater for concrete words relative to abstract words, $t(39) = 2.67$, $p = .01$, $d = .36$ (see Table 2).

Experiment 7

Experiment 6 demonstrated that the number of trials to acquisition was equivalent for concrete and abstract words; thus, processing fluency did not differ for concrete and abstract words with this measure. These results are consistent with Experiment 5, and Experiments 4–6 converge on the conclusion that processing fluency does not contribute to the concreteness effect on JOLs. In Experiment 7, we explored one final measure of processing fluency.

With regard to concreteness, the time it takes to generate a mental image could be a particularly important measure of fluency because the primary difference between concrete words and abstract words is that concrete words are given higher imageability ratings than are abstract words (e.g., Paivio et al., 1968). If concrete words are processed more fluently than are abstract words, then participants should be faster to generate mental images for them, and this should mediate the relationship between concreteness and JOLs. This possibility was investigated by Hertzog et al. (2003), who had participants form interactive mental images of word pairs that were either concrete (e.g., table-flag) or abstract (e.g., justice-distinguish). Participants pressed a key as soon as they formed a mental image for each word pair, after which they made a JOL. Consistent with the fluency hypothesis, participants were faster to create a mental image for the concrete word pairs relative to the abstract word pairs (see also Hertzog, Fulton, Mandviwala, & Dunlosky, 2013). However, the latency to generate mental images did not mediate the relationship between concreteness and JOLs.

Even so, image latency may mediate the relationship between concreteness and JOLs when participants study single words. Studying single words is qualitatively different than studying word pairs because word pairs offer additional cues that participants use as a basis for their JOLs (e.g., the associative relationship between the two words). In contrast, when participants study single words that vary in concreteness, they are likely to only focus on the relative concreteness of the words. In Experiment 7 we evaluated this issue by presenting participants with concrete and abstract words with the task to

generate a mental image for each as quickly as possible. For each word, participants reported whether they were successful in forming a mental image and made a JOL. Finally, participants took a free-recall test.

Method

Participants and design

Forty undergraduate students from Texas Christian University participated for partial credit in psychology courses. Word type (concrete and abstract) was manipulated within-participant.

Materials and procedure

Materials were identical to Experiment 2. Words were presented one at a time and presentation order was randomized anew for each participant, with the same constraints as in Experiment 2. Prior to beginning the experiment, participants were given two practice words. Participants were told that they would be presented with a series of concrete and abstract words and that their goals were to generate a mental image for each and to remember each word for a later test. They were given unlimited time to generate a mental image for each word and pressed the spacebar as soon as they were done. Next, participants were asked whether they were successful in forming a mental image, after which they made a self-paced immediate JOL. Participants were given 3 s to get ready for the next word by moving their fingers to the spacebar. Word type was not presented during image generation or when making JOLs. Finally, participants completed a self-paced free-recall test as in Experiment 2.

Results

Analyses on image latency are presented first, followed by analyses of JOLs and memory performance.

Image latency

Participants reported that they were more successful generating images for concrete words (98.5% of the words) relative to abstract words (73.7% of words), $t(39) = 6.77$, $p < .001$, $d = 1.17$. Cases in which participants reported that they failed to generate a mental image were excluded from image latency analyses.

There were few outliers, and the results were maintained after excluding them. Thus, we report on the full set of data. Consistent with the fluency hypothesis, participants were quicker to generate mental images of concrete words ($Mdn = 2.16$ sec, $SE = .05$) relative to abstract words ($Mdn = 2.49$ sec,

$SE = .06$), $t(39) = 3.23$, $p = .003$, $d = .22$. To evaluate whether image latency mediated the relationship between concreteness and JOLs, within-participant mediational analyses were conducted (Montoya & Hayes, in press). One thousand bootstrap samples were performed and the 95% confidence interval obtained for the indirect effect of concreteness on JOLs when controlling for image latency contained zero (−.17, 5.69), and was thus non-significant (see Table 3). Thus, image latency did not mediate the relationship between concreteness and JOLs.

Judgments of learning

As in the previous experiments, participants' JOLs were higher for concrete words ($M = 71.92$, $SE = 1.03$) relative to abstract words ($M = 51.90$, $SE = 1.03$), $t(39) = 9.81$, $p < .001$, $d = 1.05$.

Memory performance

Recall was greater for concrete words relative to abstract words, $t(39) = 2.62$, $p = .01$, $d = .44$ (see Table 2).

General discussion

The present experiments were the first to evaluate the mechanisms that produce the concreteness effect on JOLs. As predicted by the AP theory, participants had (or developed) beliefs about how concreteness influences memory. Specifically, when estimating global memory performance in Experiment 1, participants predicted that more concrete words would be remembered than would abstract words. This belief was also evident in Experiments 2 and 3 because participants assigned higher prestudy JOLs to concrete words than to abstract words. However, it is possible that as the word list progressed participants' prestudy JOLs were also influenced by processing fluency. If so, the concreteness effect on prestudy JOLs should be small at the start of the study list and should become larger as the list progressed. In contrast, analyses on prestudy JOLs across serial position revealed that the magnitude of the concreteness effect remained stable across serial position (i.e., concreteness did not interact with serial position in Experiments 2 or 3, $Fs < 1$). Thus, participants' beliefs produced the concreteness effect on prestudy JOLs (cf. Dunlosky et al., 2015; Hu et al., 2015; Jia et al., 2016; Li et al., 2016; Mueller et al., 2013; Mueller et al., 2014; Mueller et al., 2016; Susser et al., 2016). Taken together, these experiments provide support for the AP theory.

The results of the fluency experiments did not support the fluency account of JOLs with four measures of processing fluency that have been used in previous research (e.g., Hertzog et al., 2003; Mueller et al., 2013; Undorf &

Erdfelder, 2015). Specifically, the results revealed that self-paced study times (Experiment 5) and the number of trials to acquisition (Experiment 6) did not differ between concrete and abstract words. Thus, processing fluency did not differ between concrete and abstract words with either measure. Further, response times in a lexical decision task (Experiment 4) and latencies to generate mental images (Experiment 7) did not mediate the relationship between concreteness and JOLs. One possibility is that mediation was not observed because the analyses were low powered. To explore this, simulation studies were conducted, which revealed that the power for bootstrapping was low in both Experiment 4 (0.019) and Experiment 7 (0.043), in part because the relationship between each mediator and JOLs was not strong. In fact, to achieve 80% power approximately 20,000 participants would be needed in Experiment 4 and 5,000 participants would be needed in Experiment 7. Thus, even though the mediation analyses were underpowered, it is clear that fluency as measured via response times in a lexical decision task and image latency will not be a primary factor driving the concreteness effect on JOLs.

In contrast with the current experiments, fluency does partially mediate other cue-effects on JOLs (e.g. Besken, 2016; Besken & Mulligan, 2014; Undorf & Erdfelder, 2015). Even so, all of these outcomes are consistent with the cue-utilization approach to JOLs (Koriat, 1997). Specifically, participants' JOLs can be influenced by beliefs about how a cue influences memory and by processing fluency. Further, in our view our outcomes and those previously reported are also similar in that they are consistent with the AP theory of JOLs. The magnitude of fluency effects in mediational analyses has been small, and other research has found that beliefs play an important role in each cue effect on JOLs. Thus, participants' beliefs are meaningful for cue effects on JOLs even in instances in which fluency partially mediates the relationship between the cue and JOLs.

An important direction for future research will be to continue to develop new ways to assess fluency and determine which measures are most sensitive to detect fluency effects on JOLs. For example, it may be fruitful to explore multiple study-test trials as a measure of processing fluency (e.g., Undorf & Erdfelder, 2015). Additionally, the study lists in the present experiments were relatively short (i.e., 30 words) and it may be important to replicate the results with longer lists. Research would also benefit from qualitative assessments of participants' beliefs. For instance, future research could include post-experiment questionnaires for a comprehensive view of people's beliefs about the cue (e.g., Mueller et al., 2014). Finally, to fully understand the concreteness effect on JOLs, it may be fruitful for researchers to investigate factors that influence the magnitude of the effect on immediate JOLs. For example, researchers could manipulate whether participants are informed of the word type at the time of immediate JOLs and evaluate the concreteness effect on them.

Although not central to our primary question, it is noteworthy that Experiments 2 through 7 replicated the robust concreteness effect on memory (e.g., Begg et al., 1989; Hertzog et al., 2003; Hertzog et al., 2013; Roche, Tolan, & Tehan, 2011; Tauber & Rhodes, 2012; Tullis & Benjamin, 2012). Specifically, participants recalled more concrete words relative to abstract words. This effect has important consequences for student learning, as students are required to learn an array of information that varies in concreteness. Specifically, students will typically find it easier to learn information that is concrete, and more difficult to learn information that is abstract. As such, educators should take this into account and provide students with extra support (e.g., concrete examples, demonstrations, etc.), particularly when they are learning abstract information.

To conclude, researchers have often focused on how cues influence monitoring of learning; however, less research has examined *why* those cues influence monitoring of learning. The present research is the first to answer this question with respect to word concreteness. In support of the AP theory, concreteness influences participants' monitoring of learning because people have (or develop) the belief that concrete words are more memorable than are abstract words, and they use that information as a basis for their judgments. By contrast, processing fluency is not a primary contributor to the concreteness effect on JOLs.

References

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610–632.

Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* doi:10.1037/xlm0000246

Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 429–440. doi:10.1037/a0034407

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36,* 429–437.

Dunlosky, J., Mueller, M. L., & Tauber, S. K. (2015). The contribution of processing fluency (and beliefs) to people's judgments of learning.

In D. S. Lindsay, A. P. Yonelinas, H. I. Roediger, D. S. Lindsay, A. P. Yonelinas, & H. I. Roediger (Eds.), *Remembering: Attributions, processes, and control in human memory: Essays in honor of Larry Jacoby* (pp. 46–64). New York, NY, US: Psychology Press.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 22–34.

Hertzog, C., Fulton, E. K., Mandviwala, L., & Dunlosky, J. (2013). Older adults show deficits in retrieving and decoding associative mediators generated at study. *Developmental Psychology, 49,* 1127–1131.

Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PLoS ONE, 10,* 1–11. doi:10.1371/journal.pone.0142351

Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology, 6,* 1–10.

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.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36,* 416–428.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643–656.

Kornell, N. (2015). If it is stored in my memory I will surely retrieve it: Anatomy of a metacognitive belief. *Metacognition and Learning, 10,* 279–292.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 2,* 787–794.

Li, P., Jia, X., Li, X., & Li, W. (2016). The effect of animacy on metamemory. *Memory & Cognition.* doi:10.3758/s13421-016-0598-7

Lishner, D. A. (2015). A concise set of core recommendations to promote the dependability of psychological research. *Review of General Psychology, 19,* 52–68. doi:10.1037/gpr0000028

Lockwood, C. M., & MacKinnon, D. P. (1998). Bootstrapping the standard error of the mediated effect. *Proceedings of the 23rd Annual Meeting of SAS Users Group International* (pp. 997–1002). Cary, NC: SAS Institute, Inc.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476–490.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104. doi:10.1037/1082-989X.7.1.83

Montoya, A. K., & Hayes, A. F. (in press). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods.* doi: 10.1037/met0000086

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language, 70,* 1–12.

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *The Quarterly Journal of Experimental Psychology, 69,* 781–799. doi:10.1080/17470218.2015.1058404

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review, 20,* 378–384.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76,* 1–25.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7,* 531–536. doi:10.1177/17456916124634

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36,* 717–731. doi:10.3758/BF03206553

Roche, J., Tolan, G. A., & Tehan, G. (2011). Concreteness effects in short-term memory: A test of the item-order hypothesis. *Canadian Journal of Experimental Psychology, 65,* 245–253.

Susser, J. A., Jin, A., & Mulligan, N. W. (2016). Identity priming consistently affects perceptual fluency but only affects metamemory when primes are obvious. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 657–662.

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review, 22,* 1014–1019. doi:10.3758/s13423-014-0768-1

Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *Quarterly Journal of Experimental Psychology, 65,* 1376–1396.

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review, 19,* 743–749.

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition, 43,* 647–658.