Investigating the Effects of Item-Specific and Relational Encoding on Judgment of Learning Reactivity for Categorized, Uncategorized, and DRM Lists

Nicholas P. Maxwell

Midwestern State University

Word Count: XXXX

**Author Note**

Correspondence concerning this article should be addressed to Nicholas P. Maxwell, Department of Psychology, Midwestern State University, 3410 Taft Blvd, Wichita Falls, TX, United States. Study materials, data files, and *R* code used for all analyses have been made available via OSF: https://osf.io/t453a/.

Abstract

Judgments of Learning (JOLs) have been repeatedly found to be reactive on learning. However, the specific processes underlying JOL reactivity likely differ based on the type of stimuli participants study and the method by which memory is assessed. Recently, item-specific encoding has been proposed as a mechanism explaining JOL reactivity on word list learning. The present study tests this account by comparing reactivity between item-level and global JOLs, which differentially emphasize item-specific and relational aspects of stimuli. Participants studied and made item-level JOLs, global JOLs, or silently read categorized and uncategorized word lists (Experiments 1A and 1B) or DRM lists (Experiment 2). Overall, item-level JOLs produced positive reactivity on all list types but only when memory was assessed via recognition testing (Experiments 1B and 2). When free-recall testing was used, item-level JOLs were non-reactive on categorized lists. Separately, global JOLs improved memory for categorized but not uncategorized lists. Finally, Experiment 2 found that while item-level JOLs improved correct recognition of studied items, global JOLs increased the DRM false memory illusion. Taken together, these findings suggest that item-level JOLs encourage item-specific processing at encoding, which benefits recognition of all item types. However, when the JOL task emphasizes relational encoding (e.g., global JOLs), memory is only improved when items contain pre-existing relations (e.g., correct memory for categorized lists, false memory for critical lists). Thus, how JOLs influence memory likely depends upon both the type of stimuli and the method of testing.

Word Count: XXX

*Keywords*: Judgments of Learning; Reactivity; Categorized Lists; DRM Lists; Recognition

Investigating the Effects of Item-Specific and Relational Encoding on Judgment of Learning Reactivity for Categorized, Uncategorized, and DRM Lists

When learning new information, it is critical to assess whether the knowledge being acquired has been sufficiently encoded. Metamemory, or the processes by which individuals monitor their learning, is crucial as it helps individuals identify which items need additional study while simultaneously letting them modify their actions accordingly (see T. O. Nelson & Narens, 1990). To investigate questions surrounding these processes, researchers commonly have participants make Judgments of Learning (JOLs), which involve studying items (commonly cue-target word pairs) and predicting the likelihood of correctly remembering them on a later test (see Rhodes, 2016; Schwartz & Metcalf, 2017, for reviews). Thus, JOLs provide researchers with a simple measure for assessing how various encoding manipulations affect the study process, including perceptual manipulations (e.g., font-size; Rhodes & Castel, 2008; Chang & Brainerd, 2022) and associations between cue-target pairs (e.g., Koriat & Bjork, 2005; Castel, McCabe, & Rhodes, 2007).

Historically, JOLs have been viewed as neutral measures with little influence on memory, particularly when participants are instructed to elicit them concurrently with or immediately following study (though see Spellman & Bjork, 1992, who argued that delayed JOLs were likely to impart a memorial benefit). However, recent studies have repeatedly demonstrated that JOLs are *reactive* on learning (e.g., Janes, Rivers, & Dunlosky, 2018; Maxwell & Huff, 2022; Soderstrom, Clark, Halamish, & Bjork, 2015; see Double, Birney, & Walker, 2018, for review). As such, making JOLs appears to modify participants’ memory for studied items, likely by making certain features of the stimuli more salient at encoding (Ericsson & Simon, 1993). Thus, providing JOLs at encoding could potentially produce memory benefits (i.e., *positive reactivity*) or costs (i.e., *negative reactivity*), which would be evident when comparing memory for participants making JOLs at encoding to a separate group engaging in a no-JOL control task (e.g., silent reading).

While several theories have been put forth to explain JOL reactivity, the changed-goal hypothesis (Mitchum, Kelley, & Fox, 2016) and cue-strengthening account (Soderstrom et al., 2015) are two of the most prominent accounts (see Chang & Brainerd, 2023). First, the changed-goal hypothesis proposes that JOLs alert participants to salient differences in item difficulty which causes participants to shift from mastering all items in a list to only mastering those they perceive as being easy to learn. As such, this hypothesis predicts positive reactivity on easy items (e.g., related study items) and negative reactivity on more difficult items (e.g., unrelated items). Separately, Soderstrom et al.’s cue-strengthening account proposes that JOLs strengthen intrinsic cues which participants use to inform the magnitude of their judgments (see Koriat, 1997) and, importantly, that these cues will benefit memory so long as the method of testing is sensitive to the specific cues that have been strengthened. Because relatedness cues are highly salient, the cue-strengthening account therefore predicts positive reactivity on related study materials (e.g., paired-associates). However, based on this account, unrelated stimuli may still show positive reactivity so long as the test is sensitive to other cues which may also be strengthened (e.g., familiarity cues; see Maxwell & Huff, 2024).

Overall, findings from studies investigating JOL reactivity with cue-target word pairs have largely supported the cue-strengthening account (e.g., Chang & Brainerd, 2023; Maxwell & Huff, 2023) and, moreover, suggest that JOLs specifically encourage relational encoding when elicited in this context (e.g., Halamish & Undorf, 2023; Maxwell & Huff, 2024; Rivers, Dunlosky, Janes, Witherby, & Tauber, 2023). Importantly, these studies have revealed that positive JOL reactivity is generally moderated by the presence of pre-existing relations between cue-target pairs. Specifically, when pairs are semantically related (e.g., dog – paw), JOLs generally improve memory for the target item; however, this memory benefit does not extend to unrelated pairs (e.g., dog – cup) and may even produce memory costs for this pair type (see Undorf, Schäfer, & Halamish, 2024). Additionally, participants often indicate on post-study questionnaires that the perceived relatedness was the most important factor which they considered when assigning their JOLs (Rivers et al., 2023). Taken together, these findings are consistent with a cue-strengthening account and, moreover, suggest that JOL encourage participants to reflect upon relational properties of stimuli (i.e., relational encoding).

**JOL Reactivity and Item-Specific Encoding**

Although previous JOL reactivity studies have often emphasized the link between cue-target relations and reactivity, there is growing evidence that JOLs also encourage *item-specific encoding* of stimuli, particularly when participants study words presented in individually in lists rather than cue-target pairs (e.g., Chang & Brainerd, 2024; Senkova & Otani, 2021; Zhao et al., 2023). Based on the item-specific/relational framework (Einstein & Hunt, 1980; Hunt & Einstein, 1981), encoding tasks differ regarding their likelihood of encouraging participants to process unique properties of stimuli that differentiate between individual items (i.e., item-specific encoding) or properties which are shared between items (i.e., relational encoding). While item-specific and relational encoding tasks emphasize different aspects of studied materials, both have been shown to improve memory performance relative to silent reading. As such, the memorial benefits of JOLs could potentially reflect enhanced item-specific encoding, relational encoding, or a combination of both processing types.

Importantly, whether JOLs are likely to encourage item-specific or relational encoding may partially depend upon the type of stimuli which participants study. Because previous JOL reactivity studies have extensively used cue-target word pairs, which easily lend themselves to relational encoding (i.e., related cue-target pairs have highly salient relatedness cues), theories of JOL reactivity have often emphasized the role of relational encoding as a factor underlying this effect (e.g., Soderstrom et al.’s, 2015 cue-strengthening account). However, JOLs may also encourage item-specific encoding, particularly when they are for words presented individually in a list rather than cue-target pairs. Like cue-target pairs, word lists can similarly contain pre-existing relations (i.e., words within a list can be categorized around a specific topic or theme). However, how individuals process listwise relations when making JOLs differs from how cue-target pair relations are processed (see Koriat, 1997). Furthermore, compared to cue-target pairs, word lists are by nature more likely to encourage item-specific encoding, as words are often presented in isolation in which list-wise relations are less salient relative to pairs. As a result, making JOLs on individual words (i.e., item-level JOLs) may be more likely to encourage item-specific rather than relational encoding compared to when they are elicited on cue-target pairs.

Recent studies have explored the potential for item-level JOLs to encourage item-specific encoding of individual words. For example, Senkova and Otani (2021) reported that making JOLs individually for each word in a study list improved free-recall of categorized word lists and, furthermore, demonstrated that the memorial benefits of JOLs approximated benefits observed when participants used other encoding tasks which similarly emphasized item-specific processing (e.g., pleasantness ratings in Experiment 1 and an imagery task in Experiment 2). Separately, Zhao et al. (2023) found that item-level JOLs were non-reactive on free-recall but improved recognition when participants studied uncategorized lists (which relies extensively upon item-specific processing). Importantly, Zhao et al. also demonstrated that item-level JOLs impaired temporal memory for list items (which was assessed via an order reconstruction task that is inherently relational). Thus, unlike reactivity observed with cue-target word pairs, which is thought to reflect heightened relational encoding (see Halamish & Undorf, 2023; Maxwell & Huff, 2024), there is growing evidence suggesting that item-level JOL reactivity observed on word lists primarily reflects item-specific encoding.

To explain why item-level JOLs were reactive on word recognition but not free-recall of uncategorized lists, Zhao et al. (2023) proposed an item-order account (see McDaniel & Bugg, 2008). Per this account, word lists contain both item-specific information regarding the individual words and relational information regarding potential list-wise relations between words. Importantly, this account proposes a dissociation between item-specific and relational information, such that tasks emphasizing one aspect impair the other. Moreover, this account makes diverging predictions regarding reactivity based on test format. Specifically, the item-order account predicts that item-level JOLs would produce positive reactivity on word lists, particularly when the test is sensitive to both item-specific and familiarity-based cues (e.g., recognition testing; see Soderstom et al., 2015, for a similar argument regarding related cue-target pairs and cued-recall testing). However, because free-recall also relies upon relational cues (McDaniel & Bugg, 2008; Rawson & Zamary, 2019), the item-order account predicts no memorial benefit when participants are tested using this method. As such, item-level JOLs would be expected to be non-reactive or even produce negative reactivity on free-recall. Thus, the item-order account predicts that JOL reactivity is moderated by test type, with recognition but not free-recall showing a memorial benefit.

Finally, while previous studies suggest that JOLs encourage item-specific encoding on word lists, other evidence suggests that this reactivity effect may also reflect a relational encoding process. Critically, whether JOLs are more likely to encourage item-specific or relational encoding may depend upon how the JOL task is framed. Recently, Chang and Brainerd (2024) assessed reactivity for two types of JOLs: item-level JOLs (i.e., JOLs elicited individually for each word within a list) and global JOLs (i.e., list-wise ratings of how likely one is to remember all previously studied items within a list). Because global JOLs encourage participants to consider how each list item relates to all other items within the list, this JOL type is likely to encourage relational encoding, while the nature of item-JOLs is more likely to encourage item-specific encoding. Thus, comparisons between item-level and global JOLs may be particularly useful for investigating the roles of item-specific and relational encoding on JOL reactivity.

Overall, Chang and Brainerd (2024) reported that item-level JOLs improved free-recall, but only for randomized categorized lists (i.e., lists containing exemplars from multiple categories in which all items were intermixed). However, consistent with an item-order account, item-level JOLs were non-reactive on blocked categorized lists (i.e., list items were organized within categories), a finding which replicated patterns reported by Zhao et al. (2023). Global JOLs, however, were non-reactive on free-recall of blocked lists. However, this may have resulted from the free-recall test not being sensitive to relational cues strengthened from this JOL type (e.g., cue-strengthening). Considered alongside findings from Zhao et al. (2023), whether item-level JOLs encourage item-specific or relational encoding likely depends upon the stimuli used (e.g., blocked vs. categorized lists) and, importantly, whether the method of testing is sensitive to any potential item-specific or relational cues which may potentially be strengthened via JOLs.

**The Present Study**

As noted above, previous research has demonstrated that although item-level JOLs are generally reactive on word lists, the mechanisms driving this effect likely differ based on the stimuli and method of testing. As such, the present study compared JOL reactivity when memory was assessed via free-recall (Experiment 1A) and recognition testing (Experiments 1B and 2). Additionally, following the design of Chang & Brainerd (2024) each experiment compared between participants who provided item-level JOLs (i.e., JOLs elicited item-by-item for each word within a list), global JOLs (i.e., list-wise JOLs elicited following each word list) and a no-JOL control task. Unlike item-level JOLs which are thought to encourage item-specific encoding of word lists (e.g., Zhao et al., 2023), the global JOL task was designed to promote relational encoding by encouraging participants to process list-wise relations (i.e., processing all list items as a single unit, rather than emphasizing each individual word). As such, the inclusion of this additional encoding group allowed for a comparison between JOL tasks emphasizing item-specific and relational encoding.

Because previous studies have reported mixed findings regarding whether item-level JOLs are reactive on free-recall of word lists, Experiment 1A first sought to replicate findings from Senkova and Otani (2021) demonstrating that this JOL type produces a greater memorial benefit on categorized versus uncategorized lists when free-recall testing is used. Likewise, Experiment 1B sought to replicate findings showing that item-level JOLs are reactive on uncategorized word lists when recognition memory is assessed (e.g., Zhao et al., 2023) and, additionally, tested whether this effect would extend to categorized lists. As such, Experiments 1A/1B provided additional tests of the item-order account, as this account makes diverging predictions regarding item-level JOL reactivity when item memory is assessed via free-recall and recognition testing.

Next, Experiment 2 used Deese-Roediger-McDermott lists (DRM; Deese; 1959; Roediger & McDermott, 1995) to test the item-order account’s central claim that JOLs specifically encourage item-specific but not relational encoding of words. Like categorized lists, DRM lists similarly center around a specific category. However, items in DRM lists are also strongly related to a non-studied critical lure (e.g., *bed*, *rest*, *dream*, *nap*, and *pillow* are each related to the non-presented word *sleep*). At test, false memory for the critical lure is generally high, regardless of whether memory is assessed via free-recall or recognition (see Gallo, 2006). However, previous studies have found that item-specific and relational encoding tasks differentially affect the DRM illusion, with item-specific, but not relational tasks reducing this illusion compared to silent reading (e.g., McCabe, Presmantes, Robertson, & Smith, 2004; Huff & Bodner, 2013; Huff & Bodner, 2019). Thus, Experiment 2 provided an additional test of the item-order account while also assessing the contributions of item-specific and relational processing to JOL reactivity.

**Experiment 1A: Categorized versus Uncategorized Lists and Free-Recall Testing**

The goal of Experiment 1A was to test the item-order account of JOL reactivity on word lists. In doing so, this experiment also provided an additional test of global JOL reactivity within this context. As such, free-recall was compared between three groups: Participants who provided concurrent JOLs for each item (item-level JOLs), participants who made a single, list-wise JOL immediately following each list (global JOLs), and a no-JOL control group in which participants silently read each item without providing JOLs. Based on Senkova and Otani’s (2021) findings, item-level JOLs would be expected to benefit free-recall of categorized lists. However, the item-order account predicts that item-level JOLs would not improve free-recall of categorized lists, as this JOL type should inhibit the relational processes that facilitate memory when this test type is used. Separately, if global JOLs encourage relational encoding, this JOL type should be expected to improve free-recall of categorized lists, given that JOLs have been shown to encourage processing of pre-existing relations between stimuli (e.g., Halamish & Undorf, 2023; Maxwell & Huff, 2024). However, global JOLs would not be expected to improve free-recall of uncategorized lists, given the lack of pre-existing relations within this list type. Taken together, Experiment 1A provided an additional test of the item-order account while simultaneously testing for global JOL reactivity on categorized lists (Table 1 summarizes all anticipated JOL reactivity findings based on the item-order account).

**Method**

**Participants**

One hundred twenty-nine participants were recruited from Prolific (www.prolific.co) and were compensated at a rate of $4.00 per 20-minute session. This sample was informed by an a priori power analysis conducted with *G\*Power 3.1* (Faul, Erdfelder, Buchner, & Lang, 2009), which suggested that 111 participants would be needed to detect small-to-medium main effects/interactions (*α* = .05, 1 – *β* = .80, *d* = 0.30). However, data collection was extended to account for increased variability from online testing. Data were omitted from seven participants who failed to complete the filler task (which suggested failure to adhere to task instructions) and 9 participants with extreme recall rates (above 95% which suggested cheating or below 5% which suggested failure to attend to the study lists). This resulted in a total of 113 participants included in the final dataset (item-level JOLs, *n* = 36; global JOLs, *n* = 38; no-JOL control group, *n* = 39). All participants were native English speakers and were required to have obtained at least a high school degree or equivalent.

**Materials**

Ninety-six English words were served as stimuli. These words were split into eight 12-word lists. Of these lists, four were categorized, such that each word was an exemplar of a given category and were selected from Van Overschelde, Rawson, and Dunlosky’s (2004) categorical word norms. The four remaining lists were uncategorized based on these norms and, as such, all words within these lists were semantically unrelated. Uncategorized lists were created by randomly selecting words from unused categories in the Van Overschelde et al. norms. The four categorized and uncategorized lists were then arranged into two counterbalanced list sets, each of which consisted of two categorized and uncategorized lists (i.e., participants always studied two of each list type). Categorized and uncategorized word lists were matched between counterbalances on SUBTLEX frequency (Brysbaert & New, 2009), concreteness (D. L. Nelson, McEvoy, & Schrieber, 2004), and length, and a set of 2 (Counterbalance Version: A vs. B) × 2 (List Type: Categorized vs. Uncategorized) between subjects ANOVAs yielded no main effects or interactions with Counterbalance Version; *F*s ≤ 2.71, *p*s ≥ .10 (see Appendix Table A1 for mean lexical/semantic properties for each counterbalance). The full stimuli set used in Experiments 1A/1B has been made available at: https://osf.io/9qxne.

**Procedure**

Experiment 1A was administered online using Collector, a platform for conducting online, browser-based cognitive psychology experiments (Garcia & Kornell, 2015). Following informed consent, participants were randomly assigned to one of the two list counterbalances and one of the three encoding groups (item-level JOL, global JOL, or no-JOL control task). All participants were told that they would be viewing a series of words and that their memory for each word would later be tested. Participants in the two JOL groups then received additional instructions regarding their respective judgments. Specifically, participants in the item-level JOL group were instructed to rate their ability to remember each word. Item-level JOLs were elicited concurrently with study, such that they were provided while each word was displayed on the screen. Separately, participants in the global JOL group were informed that following the last item in each list, they would be asked to provide a single JOL representing their ability to correctly remember the previously presented set of items on a later test. For both JOL groups, JOLs were framed as the percent likelihood of correctly remembering items and were elicited via a continuous 0-100 scale (i.e., 0 = definitely will not remember, 100 = definitely will remember). To mitigate potential anchoring effects, participants were instructed to be as accurate as possible when providing their JOLs and were encouraged to use the full range of the response scale. Participants in the no-JOL group did not receive additional judgment instructions and were instead instructed to read each pair silently

After receiving their respective encoding instructions, participants were presented with the first list. For all groups, encoding was self-paced, and participants pressed the ENTER key to advance to the next word within a list. Following the first list, participants completed a short filler task in which they were presented with a random consonant and instructed to list as many words which started with this letter as they could generate in a 30 second time span (e.g., list all words starting with the letter “M”). After the time limit had been reached, participants completed a free-recall task in which they were prompted to type as many words from memory as they could correctly recall from the previous list. Following the free-recall task, participants began the second list. This process then repeated until participants had completed all four-study lists. List presentation order was randomized for all participants, and all words were additionally randomized within lists. The full experiment took approximately 20 minutes to complete.

**Experiment 1B: Categorized versus Uncategorized Lists and Recognition Testing**

Experiment 1B provided an additional test of the item-order account by assessing whether previously reported reactivity patterns observed with recognition testing on uncategorized word lists (e.g., Zhao et al., 2023) would replicate and whether this effect would extend to categorized lists. Based on an item-order account, item-level JOLs were expected to produce positive reactivity within this context. Separately, prior work has found that JOL reactivity effects are generally in the positive direction and are larger for recognition memory versus free-recall (see Zhao et al., 2023). Additionally, Maxwell and Huff (2024) recently demonstrated that item-level JOLs improve recognition of cue-target word pairs, regardless of whether pairs are related or unrelated. Based on these findings, item-level JOLs were similarly expected to improve recognition memory for both categorized and uncategorized word lists compared to the no-JOL control group.

Finally, like the previous experiment, Experiment 1B similarly included a group of participants who provided global JOLs following the presentation of each study list. However, because recognition testing is generally less sensitive to item relations versus recall (Hunt & Einstein, 1981), it was unclear the extent to which global JOLs would be reactive on this test type. Thus, Experiment 1B provided an additional test of the item-order account while further exploring the effects of global JOLs on word list learning.

**Method**

**Participants**

A total of 119 undergraduate students were recruited from Midwestern State University and completed Experiment 1B online in exchange for partial course credit. Participants were randomly assigned to one of the three encoding groups described in Experiment 1A. Initial sample sizes for each group were based on Experiment 1A, and the same exclusion criteria were applied. This process removed six participants from the final dataset, leading to a total of 113 participants included in the following analyses (item-level JOLs, *n* = 37; global JOLs, *n* = 40; no-JOL control group, *n* = 36). A sensitivity analysis conducted with *G\*Power 3.1* suggested that the final sample had sufficient power to detect small-to-medium main effects/interactions (*α* = .05, 1 – *β* = .80, *d* = 0.29). All participants were native English speakers.

**Materials and Procedure**

Experiment 1B used the same stimuli and materials as Experiment 1A, with the following exceptions. First, two 48-item study lists were generated using the four categorized and four uncategorized study lists presented in Experiment 1A. Lists were constructed such that each one contained items from two categorized lists and two uncategorized lists. Participants were randomly presented with one list at encoding. Second, the four free-recall tests used in Experiment 1A were replaced with a single, 96-item old/new recognition test. This test included all 48 items from the previously studied lists, as well as the 48 items from the counterbalanced lists, which served as non-presented controls. Second, participants studied items back-to-back, rather than completing filler tasks in between lists. All other materials, including the categorized lists and the three sets of encoding instructions (item-JOL, global JOL, and silent reading), were identical to Experiment 1A.

Following the final list, participants completed a 60-second filler task in which they were tasked with generating as many words as possible for a random consonant. Immediately afterwards, participants began the recognition test, which presented them with all 96 items in a randomized order. Specifically, participants were informed that they would be viewing a series of words and were instructed to indicate whether each word had been previously studied (“old”) or had not been previously studied (“new”). Test performance was self-paced; however, participants were instructed to respond as quickly as possible without compromising their accuracy. Like Experiment 1A, Experiment 1B took approximately 20 minutes to complete.

**Results**

For all reported analyses, significance was set at *p* < .05. For all main effects and interaction, partial eta-squared (*ηp*2)is reported, and Cohen’s *d* effect sizes are reported for post-hoc comparisons. Finally, all non-significant main effects and post-hoc comparisons include a Bayesian strength estimation of the evidence supporting the null hypothesis (see Masson, 2011; Wagenmakers, 2007). This analysis compares two models, which one assume a significant effect and a null effect, respectively, and allows for the computation of a probability estimate (termed *p*BIC; Bayesian Information Criterion). Thus, *p*bics provide a probability estimate of the null hypothesis being retained, such that higher values denote greater strength of evidence in favor of the null hypothesis. Importantly, *p*BICs are sensitive to sample size, providing increased confidence in reported null effects. Finally, free-recall data collected in Experiment 1A was scored in *R* using the *lrd* package (Maxwell, Huff, & Buchanan, 2022), which provides a set of tools scoring lexical data from memory studies while correcting potential spelling errors. Figure 1 depicts changes in free-recall proportions (top panel) and hit-rates/false alarms (bottom panel) as functions of encoding group and list type. For completeness, all recall/recognition comparisons are reported in Appendix Tables A2 and A3, and mean JOLs are reported in Appendix Table A4.

***Experiment 1A.*** Free-recall responses were analyzed using a 3(Encoding Group: Item-Level JOL vs. Global JOL vs. No-JOL) × 2(List Type: Categorized vs. Uncategorized) mixed-measures ANOVA. Overall, the main effect of Encoding Group was non-significant, *F*(2, 110) = 2.46, *MSE* = .04, *p* = .09, *ηp*2 = .04, *p*bic = .90. However, a significant main effect of list-type indicated that across encoding groups, recall was higher for categorized lists versus uncategorized lists (.72 vs. 54, respectively; *F*(1, 110) = 172.70, *MSE* = .01,*p < .*001, *ηp*2 = .61). Importantly, a significant Encoding Group × Pair Type interaction confirmed the presence of a JOL reactivity pattern, *F*(2, 110) = 3.94, *MSE* = .01, *p* = .02, *ηp*2 = .07. Starting with categorized lists, providing global JOLs increased free-recall relative to the no-JOL group (.76 vs. .68; *t*(75) = 2.38, *SEM* = .03, *p* = .02, *d* = 0.57). However, free-recall for the item-level JOL group (.71) did not statistically differ from the no-JOL and global JOL groups (*t*s ≤ 1.58, *p*s ≥ .12, *d*s ≤ 0.36, *p*BICs ≥ .71). For uncategorized lists, free-recall was highest for participants in the global JOL group (.58), followed by the no-JOL control (.55) and the item-JOL group (.48). All comparisons were non-significant (*t*s ≤ 1.67, *p*s ≥ .10, *d*s ≤ 0.34, *p*BICs ≥ .87) except for the comparison between the item-level JOL and global groups, *t*(73) = 2.16, *SEM* = .05, *p* = .03, *d* = 0.51. Thus, the memorial benefits of global JOLs on categorized lists did not extend to uncategorized lists and, additionally, providing item-level JOLs resulted in a memory cost to this list type relative to global JOLs but was non-reactive compared to silent reading.

***Experiment 1B.*** Next, hit rates were analyzed via a 3 (Encoding Group: Item-Level JOL vs. Global JOL vs. No-JOL) × 2 (List Type: Categorized vs. Uncategorized) mixed ANOVA. First, this analysis yielded a significant main effect of Encoding Group, *F*(2, 110) = 16.39, *MSE* = .03, *p* < .001, *ηp*2 = .23, as collapsed across list types, hit rates were highest for participants in the item-level JOL group (.89), followed by the no-JOL (.73) and global JOL groups (.75). Post-hoc testing confirmed that all groups significantly differed (*t*s ≥ 4.61, *ps* < .001, *d*s ≥ 1.07), except for the comparison between the no-JOL and global JOL groups, which was non-significant, *t*(74) < 1, *SEM* = .03, *p* = .45, *d* = 0.15, *p*BIC = .87. Next, a significant main effect of List Type was detected *F*(1, 110) = 67.72, *MSE* = .01, *p* < .001, *ηp*2 = .38, as hit rates were higher for categorized lists versus uncategorized lists (.85 vs. .73, respectively). Finally, this analysis revealed a significant Encoding Group × List Type interaction, *F*(2, 110) = 8.23, *MSE* = .01, *p* < .001, *ηp*2 = .13. Post-hoc testing revealed that hit rates for the item-level JOL group exceeded the no-JOL group for categorized lists (.91 vs. .81) and uncategorized lists (.87 vs. .68; *t*s ≥ 3.03, *p*s≤ .003, *d*s ≥ 0.77). Hit rates for the item-level JOL group similarly exceeded hit rates for the global JOL group for both list types (categorized lists: .91 vs. .83; uncategorized lists: .87 vs. .68; *t*s ≥ 2.65, *p*s≤ .01, *d*s ≥ 0.64). However, no differences in hit rates were detected between the global JOL and no-JOL groups for either list type (categorized lists: .83 vs. .81; uncategorized lists: .68 vs. .65; *t*s ≤ 1, *p*s ≥ .51, *d*s ≤ 0.17, *p*BICs ≥ .87). Thus, the requirement to provide JOLs at encoding benefited hits for both list types, but only when JOLs were elicited individually for each item.

Regarding false alarms (i.e., false recognition of non-presented control items), a significant difference emerged between encoding groups, *F*(2, 110) = 3.91, *MSE* = .02, *p* = .02, *ηp*2 = .07. False alarms were highest for participants in the no-JOL group (.21), followed by the global JOL group (.14) and the item-level JOL group (.11). Post-hoc testing indicated that this effect was driven by differences between the item-level and no-JOL groups, *t*(71) = 2.47, *SEM* = .04, *p* = .02, *d* = 0.61. The difference between the global and no-JOL groups was marginal, *t*(74) = 1.94, *SEM* = .04, *p* = .06, *d* = 0.44, *p*BIC = .78, and no difference in false alarms was detected between the item and global JOL groups, *t*(75) < 1, *SEM* = .03, *p* = .40, *d* = 0.23, *p*BIC = .86.

Finally, following the designs of Myers et al. (2020) and Maxwell and Huff (2024), signal detection analyses were used to test for differences in discriminability (*d′*) and response criterion (*c*) as functions of encoding group[[1]](#footnote-1). Signal detection indices were computed in *R* via the *psycho* package (Makowski, 2018), which accounts for extreme scores via a log-linear correction based on Hautus’s (1995) guidelines. Overall, mean *d′* differed between encoding groups, *F*(2, 110) = 16.32, *MSE* = 0.75, *p* < .001, *ηp*2 = .23. Mean *d′* was greatest for the item-level JOL group (2.76), followed by the global JOL group (1.98), and the no-JOL control group (1.64). All comparisons differed significantly, *t*s ≥ 3.77, *p*s< .001, *d*s ≥ 0.85, except for the comparison between the global and no-JOL groups which did not reach significance, *t*(74) = 1.83, *SEM* = 0.19, *p* = .07, *d* = 0.42, *p*BIC = .62. Next, mean *c* was greatest for participants making global JOLs (.22), followed by the item-level group (.16) and no-JOL groups (.03). This effect, however, failed to reach significance, *F*(1, 110) = 2.51, *MSE* = 0.14, *p* = .09, *ηp*2 = .04, *p*BIC = .46.

**Discussion**

Previous research has reported mixed findings regarding the reactive effects of item-level JOLs on free-recall of word lists. As such, Experiment 1A first provided a further assessment of whether item-level JOLs would improve free-recall compared to a no-JOL control group. Experiment 1B then tested whether these effects would extend to recognition testing. Additionally, both experiments included a global JOL group, who provided a single judgment for each list following study, which allowed for a comparison between JOL tasks emphasizing item-specific and relational processes. Starting with Experiment 1A, item-level JOLs did not improve free-recall, regardless of list type. This finding is consistent with previous reactivity patterns reported by Zhao et al. (2023) and is likewise in line with Chang and Brainerd’s (2024) finding that item-level JOLs did not improve free-recall of blocked categorized lists. Interestingly, item-level JOLs also produced a memory cost on free-recall of uncategorized lists relative to global JOLs. While the recall difference between item-level JOLs and the no-JOL group failed to reach conventional significance, the finding that recall in the item-level group was numerically lower than the no-JOL group is consistent with studies showing that negative reactivity sometimes occurs when participants provide item-level JOLs for unrelated cue-target word pairs (e.g., Mitchum et al., 2016; see Undorf et al., 2024). Finally, global JOLs produced positive reactivity on categorized lists, as correct free-recall exceeded the no-JOL group. Thus, consistent with an item-order account, JOLs only benefited free-recall when they encouraged processing of list-wise relations, rather than item-specific encoding of individual words.

Next, Experiment 1B demonstrated that item-level JOLs improved recognition memory, regardless of whether items were studied in categorized or uncategorized lists. This replicates previous findings by Zhao et al. (2023) who similarly reported positive reactivity on uncategorized word lists while also extending these findings to include categorized lists. Global JOLs, however, were non-reactive, regardless of list type. Finally, signal detection analyses revealed that both types of JOLs improved discriminability compared to the no-JOL group, though this effect was larger for item-level JOLs. Response criterion also improved as a function of JOLs. As such, making JOLs improved participants’ ability to discriminate between studied and non-studied items, though the increased criterion levels for JOL participants suggests that these judgments led to more biased responding. Considered alongside Experiment 1A, there is converging evidence that item-level JOL reactivity is dependent upon test format, with recognition testing, but not free-recall, being particularly sensitive to reactivity from this JOL type.

Overall, the finding that item-level JOLs improved recognition but not free-recall is consistent with an item-order account of reactivity and provides further evidence that JOLs encourage item-specific processing when elicited individually for each item in a list. Additionally, the finding that global JOLs only benefitted free-recall aligns with previous studies demonstrating that JOLs improve memory for related but not unrelated cue-target pairs (e.g., Janes et al., 2018; Soderstrom et al., 2015; Maxwell & Huff, 2022), as like JOLs elicited for cue-target word pairs, global JOLs similarly encourage processing of pre-existing relations between items, benefiting later recollection. Considered alongside findings from Experiment 1A as well as previous findings reported by Zhao et al. (2023), there is converging evidence that item-level JOL reactivity on word lists is moderated by test type, likely due to this type of JOL encouraging item-specific processing of words.

While findings from Experiments 1A and 1B suggest that item-level JOLs encourage item-specific encoding, the previous set of experiments tested this by comparing different JOL types and test formats. However, a more complete test of this account would be to compare the effects of item-level and global JOLs using stimuli that are differentially affected by item-specific and relational encoding tasks. As such, Experiment 2 replaced the categorized and uncategorized lists used previously with DRM lists. Like categorized lists, items in DRM lists are associatively related. However, in addition to containing strong list-wise relations, all items are also strongly related to a non-presented critical lure (i.e., the DRM illusion; see Deese, 1959; Roediger & McDermott, 1995). The DRM illusion is highly robust, with previous research showing that false recognition of critical lures can approximate correct recognition rates (see Gallo, 2006). Importantly, previous research has demonstrated that the DRM illusion is differentially affected by item-specific and relational encoding tasks, with item-specific but not relational tasks generally leading to a strong reduction in the DRM illusion when recognition testing is used (McCabe et al., 2004; Huff & Bodner, 2013; see Huff & Bodner, 2019). Thus, Experiment 2 tested whether reactivity patterns observed with recognition testing Experiment 1B would replicate when participants studied DRM lists while testing the effects of both item-level and global JOLs on the DRM illusion.

**Experiment 2: DRM Lists and Recognition Testing**

Experiment 2 provided an additional test of the item-order account by investigating whether item-level and global JOLs would be differentially reactive on correct and false recognition in the DRM paradigm. Based on findings from Experiment 1B, item-level JOLs were expected to improve correct recognition of studied items. Global JOLs, however, were not expected to be reactive on recognition. Regarding false memory, the item-order account predicts that item-level JOLs would reduce false recognition of critical lures, as based on this account, the requirement to provide a JOL for each item should inhibit the processing of list-wise relations. This prediction is consistent with previous research showing that tasks which encourage item-specific encoding are generally effective at reducing the DRM illusion (e.g., Burns, Jenkins, & Deans, 2007; Huff & Bodner, 2013; McCabe et al., 2004). Separately, because relational encoding tasks emphasize shared connections between list items, these tasks may additionally increase false recognition of critical lures via spreading activation (see Roediger, Balota, & Watson, 2001). Thus, the DRM illusion was expected to be greater for participants making global JOLs versus item-level JOLs.

**Method**

**Participants**

One hundred and twelve undergraduate students were recruited from Midwestern State University and completed Experiment 2 online for partial course credit. This sample size was based on the previous set of experiments, and participants were again randomly assigned to either the item-level, global, or no-JOL encoding groups. Data screening followed the same criteria as previously reported, and data from four participants were excluded. As such, the final dataset contained responses from 108 participants (item-level JOLs, *n* = 36; global JOLs, *n* = 37; no-JOL control group, *n* = 35), and a sensitivity analysis conducted using *G\*Power* 3.1 suggested that the final sample was sufficient to detect a medium effect between encoding groups (*α* = .05, 1 – *β* = .80, *d* = 0.60). As with the previous experiments, all participants were native English speakers.

**Materials and Procedure**

Twenty DRM lists previously used by Huff, Maxwell, and Mitchell (2022) which were initially selected from Roediger, Watson, McDermott, and Gallo (2001) served as stimuli. These lists contained 12 associates that were additionally related to a non-presented critical lure. Within each list, words were arranged in descending order of backward associative strength (BAS) based on the University of South Florida Free Association Norms (D. L. Nelson et al., 2004). Lists were then divided into two equal sets to serve as counterbalances, which were matched on BAS (see Table A5 for stimuli properties). Next, an 80-item old/new recognition test was generated based on these lists and consisted of 30 previously studied items (taken from positions 2, 8, and 10 from each studied list), 10 critical lures corresponding to each studied list, 30 non-presented items taken from non-studied, counterbalanced lists, and 10 critical lure controls which corresponded to the non-studied lists. The general procedure was identical to Experiment 1B such that participants were randomly assigned to one of the two counterbalances and studied all 10 lists back-to-back, with the order of list presentation randomized for each participant. All JOL instructions were identical to the previous experiments, and depending on their encoding group, participants either provided their JOLs concurrently with study, made list-wise JOLs following the completion of each list, or read each word silently. Like the previous set of experiments, encoding was self-paced for all groups. Experiment 2 took approximately 30 minutes to complete.

**Results**

**Analyses of Correct and False Recognition**

Figure 2 shows mean hit rates for studied items and false recognition of critical lures for each encoding group (see Appendix Table A6 for all comparisons including false alarms for control items). First, to test for potential JOL reactivity on correct recognition of list items, hits were analyzed via a one-way between-subjects ANOVA. This analysis yielded a significant difference between encoding groups, *F*(2, 105) = 12.44, *MSE* = .01, *p* < .001, *ηp*2 = .19, as hits were highest for participants in the item-level JOL group (.85), followed by the global JOL (.76) and no-JOL control groups (.72) Post-hoc *t*-tests confirmed that all comparisons differed significantly (*t*s ≥ 3.33, *p*s ≤ .001, *d*s ≥ 0.75), except for the comparison between the global JOL and no-JOL groups, which was non-significant, *t*(70) = 1.56, *SEM* = .03, *p* = .12, *d* = 0.32, *p*BIC = .74.

Turning to false recognition of critical lures, a separate one-way ANOVA revealed a significant difference between encoding groups, *F*(2, 105) = 3.89, *MSE* = .06, *p* = .02, *ηp*2 = .07, such that the DRM illusion was highest for participants making global JOLs (.76), followed by the no-JOL control group (.63), and the item-level JOL group (.62). All comparisons differed significantly (*t*s ≥ 2.39, *p*s ≤ .02, *d*s ≥ 0.55), except for the comparison between the item-level JOL and no-JOL control groups, *t*(69) < 1, *SEM* = .06, *p* = .88, *d* = 0.04, *p*BIC = .89.

**Analyses of *d′* and *c***

Following the design of Experiment 1B, signal detection analyses were used to test for changes in *d′* and *c* between encoding groups for both correct and false recognition. Starting with correct recognition, discriminability was highest for item-level JOLs (2.05), followed by global JOLs (1.71), and the no-JOL group (1.65). Although this pattern failed to reach conventional significance, *F*(2, 105) = 2.74, *MSE* = 0.61, *p* = .07, *ηp*2 = .05, *p*BIC = .40, a set of post-hoc analyses were conducted. These analyses revealed a significant difference in discriminability between the item-level and no-JOL groups, *t*(69) = 2.08, *SEM* = 0.20, *p* = .04, *d* = 0.49. The comparisons between the global and no-JOL groups and item-level and global JOL groups, however, did not reach significance (*t*s ≤ 1.99, *p*s *≥* .05,  *p*BICs ≥ .59, *d*s ≤ 0.47). Separately, response criterion differed between encoding groups, *F*(2, 105) = 6.59, *MSE* = .15, *p* < .01, *ηp*2 = .11. Mean *c* was greatest for participants in the no-JOL group (.23), followed by global JOLs (.11) and item-level JOLs (-.10). All comparisons differed significantly (*t*s ≥ 2.15, *p*s ≤ .03, *d*s ≥ 0.51), except for the comparison between the global JOL and no-JOL groups, *t*(70) = 1.42, *SEM* = .09, *p* = .16, *d* = 0.32, *p*BIC = .75.

Regarding false recognition, *d′* was numerically highest for the global JOL group (1.35), followed by the item-level (1.04) and no-JOL groups (1.04). Differences between encoding groups, however, were non-significant, *F*(2, 105) = 1.92, *MSE* = 0.62, *p* = .15, *ηp*2 = 0.03, *p*BIC = .94. Similarly, no significant differences in *c* emerged between groups (item-level JOL = .18, global JOL = -.04, no-JOL = .16; *F*(2, 105) = 2.24, *MSE* = .26, *p* = .11, *ηp*2 = 0.04, *p*BIC = .92.).

**Discussion**

Overall, findings from Experiment 2 are clear. Consistent with both Experiment 1B and the item-order account, the requirement to provide item-level JOLs at study improved correct recognition compared to participants in the no-JOL control group. Furthermore, global JOLs were non-reactive on correct recognition, additionally replicating Experiment 1B. Signal detection analyses similarly revealed that discriminability was greater for participants in the item-JOL group compared to the no-JOL control group. Regarding false recognition, making global but not item-level JOLs increased recognition of critical lures. Thus, a discrepancy emerged between JOL tasks and correct versus false recognition, further suggesting that item-level and global JOLs encourage separate types of processing. Finally, although item-level JOLs did not increase false recognition, this JOL type also did not reduce false recognition as initially predicted. Although item-level JOLs place a greater emphasis on item-specific properties relative to global JOLs, it is likely that this JOL type still encouraged some degree of relational encoding, given that relatedness is a highly salient cue for JOLs (see Koriat, 1997). However, considered alongside Experiments 1A/1B, there is converging evidence that item-level JOLs encourage item-specific encoding and, specifically, that these JOLs facilitate item-specific encoding to a greater extent relative to global JOLs.

**General Discussion**

The present study investigated whether JOL reactivity on word lists reflects the contributions of item-specific and relational encoding strategies. In doing so, each experiment provided additional tests of the item-order account by assessing whether JOL reactivity patterns previously reported on word lists would replicate when memory was assessed via free-recall (Experiment 1A) and recognition testing (Experiments 1B and 2) and whether recognition patterns reported on categorized/uncategorized word lists would extend to DRM lists (Experiment 2). Additionally, each experiment included a separate group of participants who made global JOLs following each list, which allowed for a comparison between JOL tasks which differentially emphasized item-specific and relational encoding of list items. As such, this allowed for a direct test of the item-order account’s claim that JOL reactivity on word lists is specifically driven by item-specific processes.

Consistent with an item-order account, Experiment 1A found no evidence that item-level JOLs are reactive when memory is assessed via free-recall testing. However, global JOLs, which emphasize relational encoding of list items, produced positive reactivity on categorized lists in which all items contained pre-existing list-wise relations. Importantly, Experiment 1B found that when recognition testing was used instead of free-recall, the inverse pattern was observed, such that item-level JOLs improved recognition for all list types while global JOLs were non-reactive. Thus, findings from Experiments 1A and 1B suggest a dissociation between JOL type and test format. While the finding in Experiment 1A that JOLs were non-reactive on free-recall contrasts with previous work by Senkova and Otani (2021), this null reactivity pattern is consistent with previous findings reported by Zhao et al. (2023). Similarly, studies investigating JOL reactivity with cue-target word pairs have also demonstrated that reactivity is moderated by test type, with item-level JOLs being reactive when memory is assessed via cued-recall and recognition but not free-recall testing (Myers et al., 2020; see also Soderstrom et al., 2015). As such, findings from Experiments 1A/1B further support are consistent with the item-order account’s central claim that positive item-level JOL reactivity on word lists reflects enhanced item-specific encoding while also being consistent with other studies showing that JOL reactivity is often moderated by test format.

Next, Experiment 2 replicated reactivity patterns observed in Experiment 1B as again item-level but not global JOLs were reactive on correct recognition. Because participants in Experiment 2 studied DRM lists, this experiment also tested whether each JOL task differentially affected false recognition of critical lures. Importantly, because the DRM illusion on recognition memory is differentially affected by item-specific and relational encoding tasks (Huff & Bodner, 2013), the use of DRM lists in Experiment 2 provided an additional test of whether item-level and global JOLs encourage item-specific and relational encoding. Consistent with Experiment 1B, making item-level JOLs improved recognition of studied items and, again, this positive reactivity effect did not extend to global JOLs. Moreover, only the requirement to provide global JOLs increased the DRM illusion. Thus, compared to item-level JOLs, the enhanced relational nature of the global JOL task relative to the item-level JOL task likely increased activation of the critical lure (see Roediger et al., 2001), leading to increased false recognition at test relative to item-level JOLs.

Overall, the finding that item-level JOLs improved correct recognition but not free-recall is consistent with an item-order account of JOL reactivity on word lists. Additionally, the finding that item-level JOLs in Experiment 1B improved memory for all list types when memory was assessed using recognition testing is also consistent with previous research by Zheng et al. (2024), who reported that item-level JOLs enhance both the familiarity and recollection processes underlying recognition (see Yonalinas, 2002). Because recognition tests require participants to correctly discriminate between old and new items, this test format places greater emphasis on item-specific properties rather than relational properties of stimuli. As a result, when memory is assessed via recognition, familiarity and recognition cues are both likely to assist correct recognition, regardless of the context in which items were studied (i.e., categorized vs. uncategorized lists). However, because free-recall testing is less sensitive to familiarity and utilizes more recollective cues, item-level JOLs do not produce positive reactivity when this test type is used. As a result, item-level JOLs appear to be particularly effective at improving recognition but not free-recall.

Separately, because global JOLs place a greater emphasis on list-wise relations relative to item-level JOLs, global JOLs may be particularly effective whenever items are presented in categorized lists in which intra-list relations are highly salient. However, for reactivity to occur within this context, the method of testing must be similarly sensitive to these strengthened relations (i.e., cue-strengthening). In the present study, this likely occurred in Experiment 1A, as participants completed a series of free-recall tests following study of each word list. However, as noted by an anonymous reviewer, participants encoded all pairs in Experiment 1B prior to completing a single recognition test assessing memory for all 96 OLD/NEW items. While this test format was utilized in Experiment 1B to account for potential ceiling effects on recognition that would occur from using short study/test cycles, this presentation format may have inadvertently diminished any potential relatedness cues which would have otherwise benefitted recall of categorized lists such as those reported by Senkova and Otani (2021). Separately, discrepancies in presentation format may also explain why global JOLs improved free-recall in Experiment 1A but were non-reactive on blocked categorized lists in Chang and Brainerd’s (2024) study. As such, more work is needed to fully explore the interaction between test format and JOL type on reactivity.

Moreover, the finding that item-level JOLs did not improve free-recall in Experiment 1A is consistent with an item-order account and suggests that item-level JOLs impair relational encoding. In the present study, this was tested by comparing memory between item-level and global JOL tasks. However, future research may wish to provide additional tests of this account using temporal clustering or serial position analysis. If item level-JOLs indeed have reduced relational encoding versus global JOLs, the item-JOL task would be expected to impair serial position accuracy. However, note that item although item-level JOLs are likely to bias participants towards item-specific encoding, this JOL type likely still encourages some degree of relational encoding, albeit to a lesser degree relative to global JOLs. For example, studies utilizing cue-target word pairs have repeatedly demonstrated that reactivity reflects a relational process when recall testing is used (e.g., Halamish & Undorf, 2023; Maxwell & Huff, 2024; Rivers et al., 2023). Importantly, this positive reactivity pattern often approximates memory benefits observed when participants encode pairs via other tasks which similarly encourage relational encoding (e.g., frequency of co-occurrence judgments, judgments of associative memory; Maxwell & Huff, 2022). Thus, whether item-level JOLs are more likely to encourage item-specific or relational encoding appears to be strongly dependent on both the type of stimuli that participants study and the specific framing of the JOL task.

This notion may partially explain why item-level JOLs did not reduce the DRM illusion in Experiment 2 as initially anticipated. Because JOLs are particularly sensitive to pre-existing item relations (e.g. Halamish & Undorf, 2023; Maxwell & Huff, 2024, Soderstrom et al., 2015; Rivers et al., 2023) these judgments may have still encouraged participants to process intra-list relations to some degree, even when they were elicited for individual items (e.g., item-level JOLs). This additional relational processing likely reduces any potential benefits of item-specific encoding that would otherwise reduce the DRM illusion, and which are typically detected with other item-specific encoding tasks (e.g., McCabe et al., 2004; Huff & Bodner, 2013). Thus, although both JOL tasks were designed to encourage one specific type of processing over the other, it is unlikely that either JOL task was truly process pure (see Huff & Bodner, 2019). For example, when participants are instructed to apply item-specific tasks while studying related word lists (e.g., categorized lists, DRM lists), they likely also engage in some degree of relational encoding due to the inherent relational properties of the study lists (Hunt & Seta, 1984). However, given that the present study is the first to explore JOL reactivity within the context of the DRM lists, more research will be needed to fully explore the differential effects of both JOL types on correct and false recognition.

Finally, given the benefits of item-level JOLs on recognition testing that were observed in the present study as well as similar benefits reported with cue-target word pairs (e.g. Maxwell & Huff, 2024; Myers et al., 2020), future research may wish to explore whether this effect extends to educationally relevant stimuli. Although previous research has found that JOLs are non-reactive when participants’ study general knowledge facts (Schäfer & Undorf, 2024) and text passages (Ariel, Karpicke, Witherby, & Tauber, 2021), these studies have primarily assessed memory using recall testing, rather than recognition (though see Schäfer & Undorf, who found no positive reactivity on general knowledge when using a multiple-choice paradigm). As previously noted, recognition testing is generally more sensitive to familiarity relative to free-recall testing. Thus, if item-level JOLs consistently enhance both familiarity and recollection as proposed by Zheng et al. (2024), these judgments would be expected to improve memory for educationally relevant study materials, so long as learning is assessed using recognition-based approaches. Ultimately, however, more research is needed to fully explore the extent to which item-level JOLs can improve memory within applied contexts.

**Conclusion**

In summary, the present study further evaluates JOL reactivity effects on word list learning. By comparing between item-level and global JOLs, which differentially emphasize item-specific and relational encoding, each experiment tested recent work suggesting that JOL reactivity observed with word lists primarily reflects enhanced item-specific encoding. In doing so, the present study provides further support for an item-order account of JOL reactivity, as item-level JOLs were reactive on recognition but not free-recall testing. Thus, the present study provides further evidence that item-level JOLs encourage item-specific encoding when these judgments are provided individually for study list items. Separately, when the JOL tasks was reframed to encourage relational encoding, memory was only influenced when list items contained pre-existing relations (e.g., categorized lists, the DRM illusion). Taken together, the present study demonstrates that while JOLs are likely to encourage item-specific encoding, both the potential memorial benefits of JOLs and the processes driving reactivity are likely determined by the stimuli participants study and the method of testing.

**Declarations**

**Open Practices Statement**

Study materials, data files, and *R* code used for analyses have been made available via OSF (https://osf.io/t453a/).

**Compliance with Ethical Practices**

The reported studies were approved by the Institutional Review Board at Midwestern State University (Protocol #24022005). All participants provided informed consent prior to participating in the experiments. The author reports no conflicts of interest.

**Funding Information**

Data collection was partially supported by an intramural research grant awarded to the author by Midwestern State University.

**References**

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. K. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review, 33*, 693-712.

Brysbaert, M. & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Burns, D. J., Jenkins, C. L., & Dean, E. E. (2007). Falsely recalled items are rich in item-specific information. *Memory & Cognition, 35*, 1630–1640.

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14* (1), 107–111.

Chang, M. & Brainerd, C. J. (2024). Judgments of learning reactivity on item-specific and relational processing. *Journal of Intelligence, 12*(1), 1–14.

Chang, M. & Brainerd, C. J. (2022). Association and dissociation between judgments of learning and memory: A meta-analysis of the font-size effect. *Metacognition and Learning, 17*, 443–476.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17–22.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgments of learning. *Memory, 26*(6), 741–750.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology, 6*(5), 588–598.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (Rev. ed.)*. Cambridge, MA: Bradford Books/ MIT Press.

Faul, F., Erdfelder, E., Buchner, A, & Lang, A. G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks.* Psychology Press.

Garcia, M. & Kornell, N. (2015). Collector [Computer software]. Retrieved April 3rd, 2020 from https://github.com/gikeymarica/Collector

Halamish, V. & Undorf, M. (2023). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(4), 547–556.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d′*. *Behavior Research Methods, Instruments, & Computer, 27*, 46–51.

Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1246–1256.

Huff, M. J., & Bodner, G. E. (2019). Item-specific and relational processing both improve recall accuracy in the DRM paradigm. *Quarterly Journal of Experimental Psychology*, 72(6), 1493–1506.

Huff, M. J., Maxwell, N. P., & Mitchell, A. (2022). Distinctive Sans Forgetica font does not benefit memory accuracy in the DRM paradigm. *Cognitive Research: Principles and Implications, 7*, 102.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*(5), 497–514.

Hunt, R. R., & Seta, C. E. (1984). Category size effects in recall: The roles of relational and individual item information. J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 454–464.

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both? *Psychonomic Bulletin & Review, 25*(6), 2356–2364.

Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experiment Psychology: General, 126*(4), 349–370.

Koriat, A. & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187–194.

Koriat, A. & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490–517.

Makowski, D. (2018). The *psycho* package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470.

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*, 679–690.

Maxwell, N. P. & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: Evidence from associative memory and frequency judgments. *Metacognition and Learning, 17*, 589–625.

Maxwell, N. P. & Huff, M. J. (2023). Is discriminability a requirement for reactivity? Comparing the effects of mixed vs. pure list presentations on judgment of learning reactivity*. Memory & Cognition, 51*(5), 1198–1213.

Maxwell, N. P. & Huff, M. J. (2024). Judgment of learning reactivity reflects enhanced relational encoding on cued-recall but not recognition tests. *Metacognition and Learning, 19*, 189-213.

Maxwell, N. P., Huff, M. J., & Buchanan, E. M. (2022). The *lrd* package: An *R* package and Shiny application for processing lexical data. *Behavior Research Methods, 54*, 2001-2024.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review, 11*, 1074–1079.

McDaniel, M. A. & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15,* 237-255.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145*(2), 200–219.

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition, 48*, 745-758.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In: *The psychology of learning and motivation*, ed. G. Bower. American Psychologist.

Rawson, K. A. & Zamary, A. (2019). Why is free recall practice more effective than recognition practice for enhancing memory? Evaluating the relational processing hypothesis. *Journal of Memory and Language, 105*, 141–152.

Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 65–80). Oxford University Press.

Rhodes, M. G. & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615–625.

Rivers, M. L., Dunlosky, J., Janes, J. L., Witherby, A. E., & Tauber, S. K. (2023). Judgments of learning enhance recall for category-cued but not letter-cued items. *Memory & Cognition, 51*, 1547–1561.

Roediger, H. L. III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). American Psychological Association.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented lists. J*ournal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407.

Schäfer, F. & Undorf, M. (2024). On the educational relevance of immediate judgment of learning reactivity: No effects of predicting one’s memory for general knowledge facts. Journal of Applied Research in Memory and Cognition, 13(1), 113-123..

Schwartz, B. L. & Metcalfe, J. (2017). Metamemory: An update of critical findings. In J. H. Bryne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 423–432). Academic Press.

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition, 49*, 955-967.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 553–558.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3*(5), 315-317.

Undorf, M., Schäfer, F., & Halamish, V. (2024). Making judgments of learning either enhances or impairs memory: Evidence from 17 experiments with related and unrelated word pairs. *Collabra:* Psychology, 10(1): 117108

Van Overschelde, J. P., Rawson, K. A., & Dunloskey, J. Category Norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*(3), 289-335.

Wagenmakers, E. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14*, 779-804.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517.

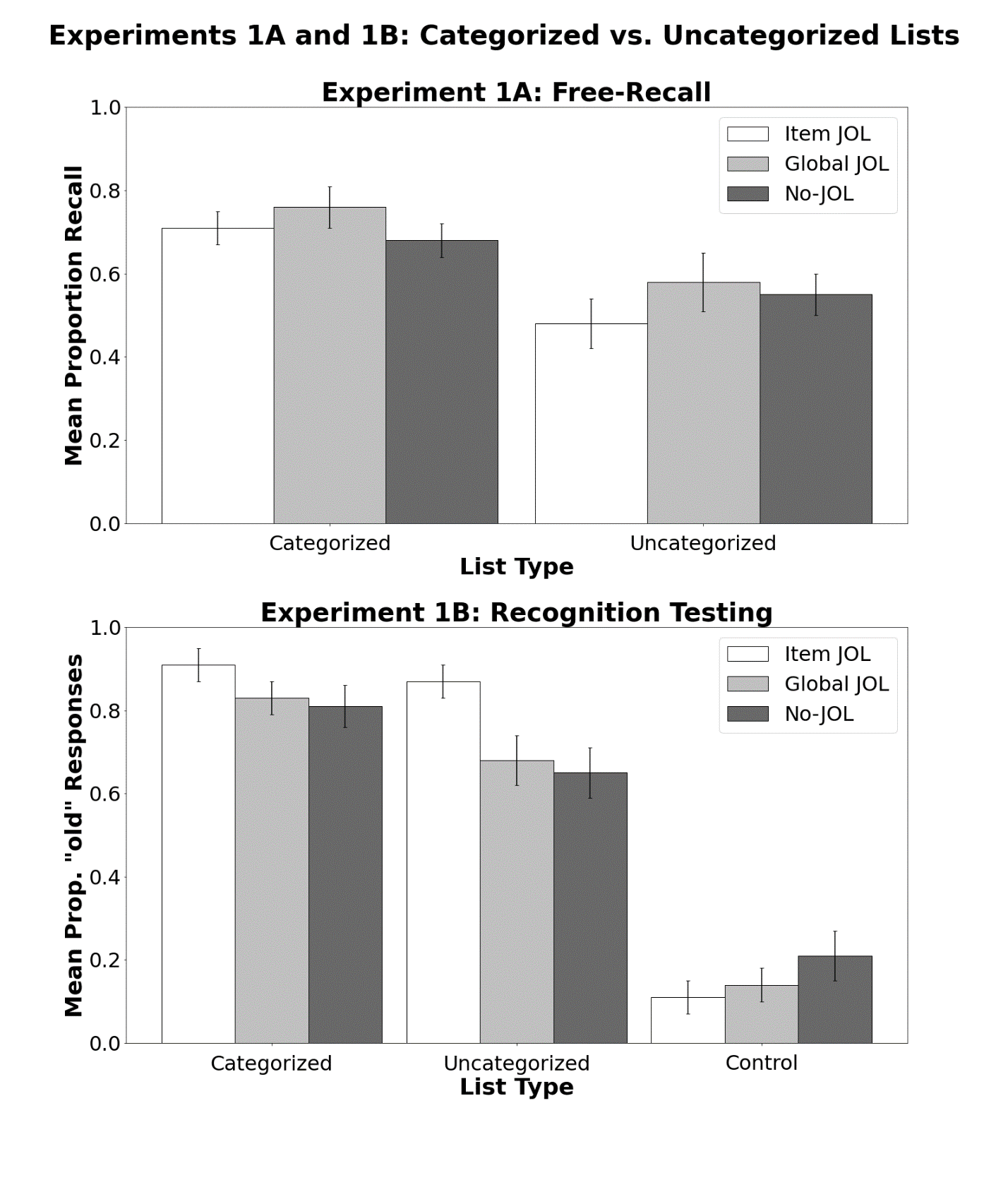
Zhao, W., Li, J., Shanks, D. R., Li, B., Hu, X., Yang, C., & Luo, L. (2023). Metamemory judgments have dissociable reactivity effects on item and interitem relational memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 49*(4), 557-574.

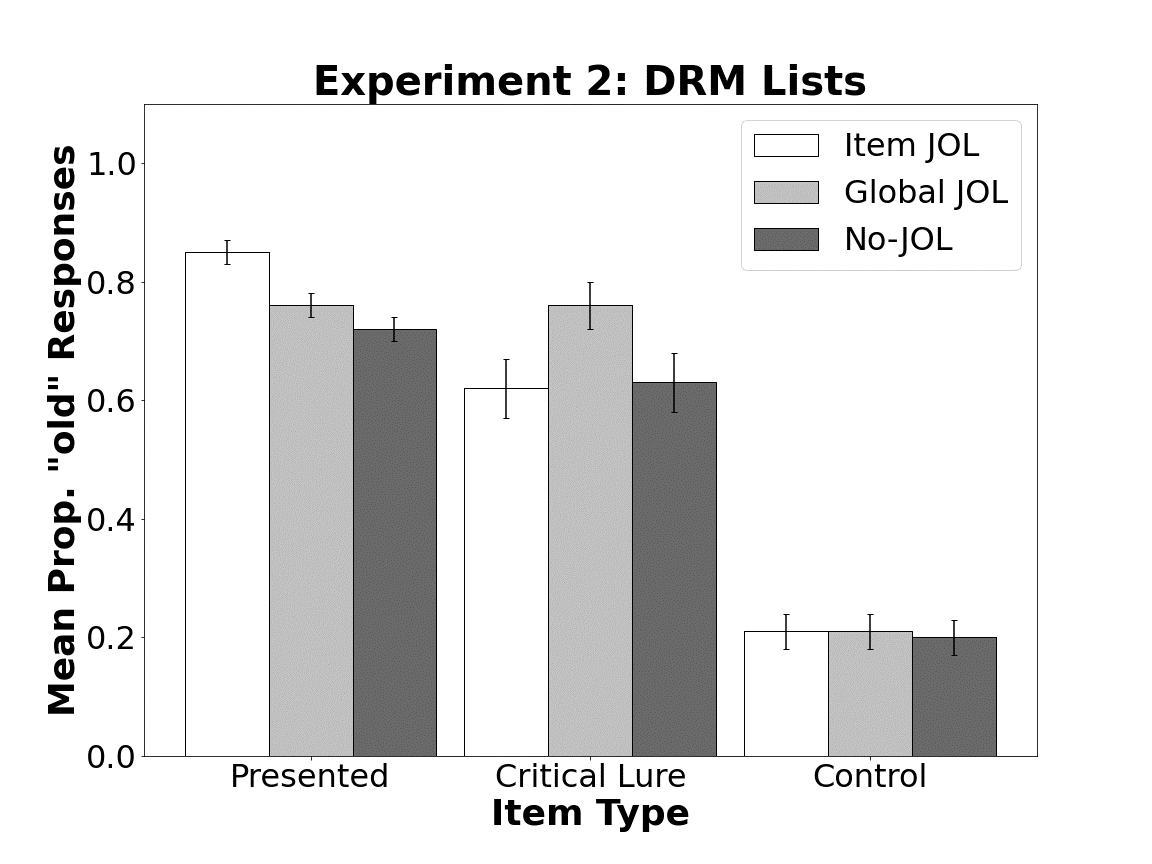
Zheng, J., Baike, L., Zhao, W., Su, N., Fan, T., Yin, Y., Hu, Y., Hu, X., Yang, C., & Luo, L. (2024). Soliciting judgments of learning reactivity facilitates both recollection- and familiarity- based recognition memory. *Metacognition and Learning, 19,* 609-633.

Table 1.

*Expected Reactivity Patterns Based on the Item-Order Account for each List and JOL Type for each Experiment*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment | Test Type | Item/List Type | JOL Type | Predicted Reactivity Pattern |
| Ex. 1A | Free-Recall | Categorized | Item JOL | Null |
|  |  |  | Global JOL | Positive |
|  |  | Uncategorized | Item JOL | Null |
|  |  |  | Global JOL | Null |
| Ex. 1B | Recognition | Categorized | Item JOL | Positive |
|  |  |  | Global JOL | Null |
|  |  | Uncategorized | Item JOL | Positive |
|  |  |  | Global JOL | Null |
| Ex. 2 | Recognition | DRM Studied Items | Item JOL | Positive |
|  |  |  | Global JOL | Null |
|  |  | DRM Critical Lures | Item JOL | Negative |
|  |  |  | Global JOL | Positive |

*Figure 1*. Mean proportion of correct free-recall in Experiment 1A (top panel) and mean proportion of “old” responses in Experiment 1B (bottom panel). “Control” columns indicate “old” responses to control items, which were taken from the non-studied counterbalance. Bars indicate 95% *CI*s.



*Figure 2.* Mean proportion of “old” responses to presented items and critical lures in Experiment 2. Bars indicate 95% *CI*s. “Control” columns indicate “old” responses to control items, which were taken from the non-presented counterbalance.

**Appendix**

Table A1

*Lexical Properties of Categorized and Uncategorized Lists in Experiments 1A and 1B.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| List Type | Counterbalance | Frequency | Length | Concreteness |
| Categorized | A | 2.91 (0.53) | 5.29 (1.57) | 6.24 (0.41) |
|  | B | 2.85 (0.62) | 5.71 (1.60) | 6.09 (0.44) |
| Uncategorized | A | 2.94 (0.68) | 5.25 (1.39) | 5.75 (0.65) |
|  | B | 3.01 (0.73) | 4.71 (1.08) | 5.53 (0.94) |

*Note*: Cells denote mean values; parenthesis indicate *SD*s. Frequency ratings were derived from SUBLTEX (Brysbaert & New, 2009). Concreteness ratings were taken from D. L. Nelson et al. (2004). Values are collapsed across individual study lists for each list type/counterbalance.

Table A2

*Comparison of Mean Recall Proportions in Experiment 1A as functions of Encoding Group and List Type*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Encoding Group | List Type | *M* | *± 95% CI* | U |
| Item JOL | Categorized | .71 | .04 |  |
|  | Uncategorized | .48 | .06 | 1.46\* |
| Global JOL | Categorized | .76 | .05 |  |
|  | Uncategorized | .58 | .07 | 0.99\* |
| No-JOL | Categorized | .68 | .04 |  |
|  | Uncategorized | .55 | .05 | 0.86\* |

*Notes*: Right-most column denotes Cohen’s *d* effect sizes for post-hoc comparisons between list types within encoding groups. \* = *p* < .05. U = Uncategorized lists.

Table A3

*Comparison of Mean Hit Rates, False Alarms, d′*, *and c in Experiment 1B as functions of Encoding Group and List Type.*

|  |  |  |  |
| --- | --- | --- | --- |
| Item Type/Measure | Item JOL | Global JOL | No JOL |
| Categorized Items | .91 (.04) | .83 (.04) | .81 (.05) |
| Uncategorized Items | .87 (.04) | .67 (.06) | .65 (.06) |
| Controls | .11 (.04) | .14 (.04) | .21 (.06) |
| *d′* | 2.76 (0.31) | 1.97 (0.27) | 1.64 (0.25) |
| *c* | 0.03 (0.12) | 0.22 (0.10) | 0.16 (0.14) |

*Notes*: Categorized and Uncategorized item rows denote proportion of hits. Control row denotes proportion of false alarms to non-studied items. *d′* and *c* values were computed using hits that were collapsed across list types. Parentheses indicate ± 95% *CI*s.

Table A4

*Comparison of Mean JOLS in Experiments 1A, 1B, and 2 as functions of Encoding Group and List Type*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment | Encoding Group | List Type | *M* | *± 95% CI* |
| Ex. 1A | Item JOL | Categorized | 62.01 | 4.96 |
|  |  | Uncategorized | 54.89 | 5.11 |
|  | Global JOL | Categorized | 55.36 | 8.40 |
|  |  | Uncategorized | 43.95 | 7.72 |
| Ex. 1B | Item JOL | Categorized | 67.99 | 5.01 |
|  |  | Uncategorized | 61.66 | 5.50 |
|  | Global JOL | Categorized | 56.14 | 7.49 |
|  |  | Uncategorized | 54.11 | 8.13 |
| Ex. 2 | Item JOL | DRM | 62.48 | 6.71 |
|  | Global JOL | DRM | 62.06 | 7.79 |

*Notes*: Right-most column denotes Cohen’s *d* effect sizes for post-hoc comparisons for each list type within encoding groups.

Table A5

*Lexical Properties for DRM lists in Experiment 2*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Counterbalance | Frequency | Length | Concreteness | BAS |
| A | 2.76 (0.74) | 5.63 (1.61) | 5.02 (1.22) | .207 (.20) |
| B | 2.89 (0.84) | 5.23 (1.50) | 5.31 (1.08) | .218 (.20) |

*Note*: Cells denote mean values; parenthesis indicate *SD*s. Frequency ratings were derived from SUBLTEX (Brysbaert & New, 2009). Concreteness ratings were taken from D. L. Nelson et al. (2004). BAS = Backward Associative Strength between item and critical lure. Values are collapsed across individual study lists for each list type/counterbalance.

Table A6

*Comparison of Mean Hit Rates, False Alarms, d′*, *and c in Experiment 2 as functions of Encoding Group and Item Type.*

|  |  |  |  |
| --- | --- | --- | --- |
| Item Type/Measure | Item JOL | Global JOL | No JOL |
| List Items | .85 (.02) | .76 (.02) | .72 (.02) |
| List Item Controls | .20 (.03) | .20 (.03) | .18 (.04) |
| Critical Items | .62 (.04) | .76 (.04) | .63 (.05) |
| Critical Item Controls | .25 (.04) | .27 (.03) | .25 (.03) |
| *d′* | 2.05 (0.24) | 1.71 (0.23) | 1.65 (0.29) |
| *c* | -0.10 (0.13) | 0.11 (0.13) | 0.23 (0.11) |

*Notes*: List Items row denotes proportion of hits on studied items. Control rows indicate false alarms on non-studied items taken from control lists and critical lures corresponding to non-studied lists. False alarms were collapsed across both control item types when computing *d′* and *c.*

1. False alarms were collapsed across categorized and uncategorized list types. As such, *d′* and *c* were computed for all list items rather than separately for list types. [↑](#footnote-ref-1)