Investigating the Effects of Item-Specific and Relational Encoding on JOL Reactivity for Categorized, Uncategorized, and DRM Word 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 analyses have been made available via OSF https://osf.io/t453a/.

Abstract

Judgments of Learning (JOLs) have been repeatedly shown to be reactive on learning. However, the processes underlying JOL reactivity likely differ as based on the type of stimuli participants study (e.g., cue-target word pairs vs. word lists) and test format (cued-recall vs. recognition testing). Recently, the item-order account (Zhao et al., 2022; 2023) posits that JOL reactivity on word lists reflects enhanced item-specific encoding of items, rather than increased relational encoding. The present study tests the account by comparing reactivity for item-level and global JOLs on categorized and uncategorized word lists (Experiments 1A and 1B) and DRM lists (Experiment 2). Overall, item-level JOLs were reactive on all list types, but only when memory was assessed using recognition testing (Experiments 1B and 2). When participants were tested via free-recall item JOLs were non-reactive, though global JOLs, which encourage relational encoding, improved memory for categorized but uncategorized lists. Separately, Experiment 2 found that while item-level JOLs improved correct recognition of studied items, global JOLs increased the DRM illusion. Taken together, these results suggest that item-level JOLs encourage item-specific processing at encoding, which benefits recognition via enhanced familiarity. However, when the JOL task emphasizes relational encoding (e.g., global JOLs), memory is only modified when items contain pre-existing relations (e.g., correct memory for categorized lists, false memory for critical lists). Thus, how JOLs influence memory is dependent upon both the type of stimuli and the method of testing.

Word Count: 234

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

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

When learning new information, it is critical to assess whether the knowledge one is acquiring has been sufficiently encoded. Metamemory, or the processes by which individuals regulate their learning, is crucial as it helps individuals identify which items need additional study (see 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 treated as neutral measures which have little influence on memory, particularly when they are elicited 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). Based on these findings, the act of making JOLs modifies 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).

Prior research has commonly tested for JOL reactivity using cue-target word pairs (e.g., Janes et al., 2018; Maxwell & Huff, 2022; Maxwell & Huff, 2023; Mitchum, Kelly, & Fox, 2016; Myers, Rhodes, & Hausman, 2020; Soderstrom et al., 2015; though see Senkova & Otani, 2021, who had participants study word lists). These studies have revealed a consistent pattern of memory changes: When pairs are semantically related (e.g., dog – paw), JOLs generally improve memory for the target item, though this memory benefit does not extend to unrelated pairs (e.g., dog – cup). To explain this pattern, Soderstrom et al. (2015) proposed that for JOLs to benefit memory, items must contain intrinsic cues which participants use to inform the magnitude of their JOLs (see Koriat, 1997) and that memory must be assessed using a method that is sensitive to these strengthened cues (i.e., *cue-strengthening account*). Recent work supports this cue-strengthening account, and later studies have explored the degree to which relatedness directly contributes to positive JOL reactivity on cue-target pairs (e.g., Janes et al., 2018; Halamish & Undorf, 2023; Maxwell & Huff, 2022; Maxwell & Huff, 2023; Rivers, Dunlosky, Janes, Witherby, and Tauber, 2023).

However, while previous reactivity studies have emphasized the link between cue-target relations and reactivity, the present study focuses instead on the potential for JOLs to also encourage *item-specific encoding* of stimuli. Based on the item-specific/relational framework (Einstein & Hunt, 1980; Hunt & Einstein, 1981), encoding tasks differ in the likelihood that they encourage processing of unique properties that differentiate individual items (i.e., item-specific encoding) or processing which emphasizes properties 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 item-specific encoding, relational encoding, or a combination of both processing types.

Importantly, whether JOLs encourage item-specific or relational processing likely depends upon the type of stimuli that participants study. Because previous studies have made extensive use of cue-target pairs, which may lend themselves to relational encoding (e.g., related cue-target pairs; see Maxwell & Huff, 2022), theories of JOL reactivity have often emphasized the role of relational encoding. However, JOLs may also encourage item-specific encoding of stimuli, particularly when they are elicited for words presented individually within a list rather than when they are provided on cue-target pairs. Like cue-target pairs, word lists can similarly contain pre-existing relationships (i.e., words within a list can be categorized around a specific topic or theme), though how these relationships are processed when participants make JOLs likely differs from how related cue-target pairs are processed (see Koriat, 1997). Furthermore, compared to cue-target pairs, word lists are more likely to encourage item-specific encoding, as JOLs are elicited on individual items rather than for items pairs in which pre-existing relations are more salient. Thus, making JOLs on individual words may instead encourage item-specific encoding rather than relational encoding.

Recent studies have explored the potential for JOLs to encourage item-specific encoding of individual words. For example, Senkova and Otani (2021) found that making JOLs improved free-recall of categorized lists versus uncategorized lists and, furthermore, demonstrated that the memorial benefits of JOLs approximated benefits observed with other encoding tasks which emphasized item-specific processing (e.g., pleasantness ratings in Experiment 1 and an imagery task in Experiment 2). Separately, Zhao et al. (2022) provided further evidence for an item-specific account of reactivity, as they found that JOLs improved recognition of uncategorized lists (which relies extensively upon item-specific processing) while simultaneously impairing temporal memory (which was assessed via an order reconstruction task and is inherently relational). These findings were subsequently replicated by Zhao et al. (2023), who similarly reported positive reactivity on correct recognition of word lists but negative reactivity on temporal memory for list items. Thus, unlike reactivity on cue-target pairs, which is thought to reflect heightened relational encoding (see Halamish & Undorf, 2023; Maxwell & Huff, in press), word list reactivity appears to reflect an item-specific process.

To explain positive JOL reactivity on word lists, Zhao et al. (2022; 2023) proposed an item-order account of reactivity (see McDaniel & Bugg, 2008). Based on this account, word lists contain both item-specific information regarding the individual words and relational information regarding potential inter-item relations. Importantly, this account proposes a dissociation between item-specific and relational processes, such that tasks which improve one type of encoding may be detrimental to the other. Thus, the item-order account predicts positive reactivity on word lists, particularly when the test is sensitive to item-specific cues (e.g., recognition testing). Finally, the item-order account makes diverging predictions about JOL reactivity based on test format. First, this account predicts positive reactivity when memory is tested via recognition, as this test format is particularly sensitive to item memory. However, because free-recall additionally relies upon intra-item relations (McDaniel & Bugg, 2008; Rawson & Zamary, 2019) which are impaired by JOLs (Zhao et al., 2023), the item-order account predicts weaker or no reactivity when testing occurs via free-recall.

Finally, while previous studies suggest that JOLs encourage item-specific encoding of list items, other findings suggest that JOL reactivity on word lists may instead reflect a relational encoding process. Recently, Chang & Brainerd (2024) tested for reactivity using two JOL types: item-level JOLs (i.e., JOLs elicited individually for each word within a list) and global JOLs (i.e., list-wise ratings of likely one is to remember all items presented within a study list). Overall, only the requirement to provide item-level JOLs improved free-recall, and these effects were moderated by list type. Specifically, item-level JOLs improved memory for randomized lists (i.e., lists containing exemplars from multiple categories) but not blocked lists (i.e., all list items were exemplars of one category). By fitting their free-recall data to the dual-retrieval model, the authors estimated the contributions of recollection and familiarity to reactivity. (see Brainerd, Reyna, & Howe, 2009). In doing so, Chang & Brainerd concluded that positive item-level reactivity on randomized lists was primarily tied to gist processing rather than item-specific encoding. However, given that reactivity effects of item-level JOLs are likely to be larger when memory is assessed via recognition testing, whether reactivity reflects item-specific or relational processes may be partially dependent on test type.

**The Present Study**

As noted above, previous research has demonstrated that while JOLs are generally reactive on word lists, the mechanisms driving this effect likely differ versus reactivity observed on cue-target word pairs. As such, the present study further examined the roles of item-specific and relational processing underlying reactivity on word lists. First, since previous studies have reported mixed findings regarding whether JOLs are reactive on free-recall of word lists, Experiment 1A sought to replicate findings from Senkova and Otani (2021) demonstrating that JOLs produce a greater memorial benefit on categorized versus uncategorized lists when memory is assessed via free-recall. Likewise, Experiment 1B sought to replicate findings showing that JOLs are reactive on uncategorized word lists when memory is assessed via recognition testing (e.g., Zhao et al., 2023) and test whether this effect extends to categorized lists. In doing so, Experiments 1A/1B provided additional tests of the item-order account, as this account makes diverging predictions regarding 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 also 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 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, leading to a reduction in the DRM illusion compared to silent reading (e.g., McCabe, Presmantes, Robertson, & Smith, 2004; Huff & Bodner, 2013; Huff & Bodner, 2019). Thus, Experiment 2 was designed to provide an additional test of the item-order account while also assessing the contributions of item-specific and relational processing to reactivity.

Finally, in addition to standard, item-level JOLs (i.e., JOLs elicited individually for each study item within a list), each experiment also included a separate group of participants who made global, list-wise JOLs following each study list (e.g., Chang & Brainerd, 2024). Unlike item-level JOLs, this global JOL task was designed to promote relational encoding by encouraging participants to process inter-list relations (i.e., processing all list items as a single unit, rather than emphasizing each individual word as is the case with item-level JOLs). As such, the inclusion of this additional control group allowed for a comparison between JOL tasks emphasizing item-specific and relational encoding.

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

The goal of Experiment 1A was to replicate findings initially reported by Senkova and Otani (2021) showing that JOLs produce a greater free-recall benefit on categorized lists versus uncategorized lists. In doing so, free-recall was compared between three groups: Participants who provided concurrent JOLs for each item, participants who made a single, global JOL immediately following the presentation of each study list, and a no-JOL control group in which participants silently read each item without providing JOLs. Based on Senkova and Otani’s (2021) findings, JOLs were expected to benefit free-recall of categorized lists. However, the item-order account makes diverging predictions for item-level and global JOLs. First, this account predicts that item-level JOLs would not be reactive, as making this type of JOL should inhibit relational processes that facilitate free-recall. However, global JOLs would be expected to improve recall, as this JOL task emphasizes inter-list relations by requiring participants to reflect on all items presented with a list. Thus, the inclusion of item-level and global JOL encoding groups allowed for a comparison between item-specific and relational focused judgments. Finally, because categorized lists contain pre-exiting relations, any benefits of global JOLs on this list type were expected to be greater than uncategorized lists.

**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 (*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 selected to serve as stimuli. These words were split into eight 12-word lists, which were based on Van Overschelde, Rawson, and Dunlosky’s (2004) categorical word norms. Of these lists, four were categorized, such that each word was an exemplar of a given category. The four remaining lists were uncategorized and, as such, all words within the list were semantically unrelated. Uncategorized lists were created by randomly selecting words from unused categories in the Van Overschelde et al. norms (see Appendix Table A1 for lexical properties of all lists).

**Procedure**

Experiment 1A was administered online using Collector, an online platform for conducting browser-based cognitive psychology experiments (Garcia & Kornell, 2015). Following informed consent, 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-based JOL group were instructed to rate their ability to remember each word, with JOLs elicited concurrent with study, such that JOLs were provided while each word was displayed. 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 items at test 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. After completing 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 would be expected to produce positive reactivity within this context. Furthermore, because this account makes no claims regarding the effects of list relatedness on reactivity, item-level JOLs were also expected to benefit recognition memory for both categorized and uncategorized lists. Finally, Experiment 1B similarly included a group of participants who made global JOLs following the presentation of each study list. However, because recognition testing is less sensitive to inter-item relations (Hunt & Einstein, 1981), it was unclear whether global JOLs would be reactive when memory was assessed using 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* suggested that the final sample had sufficient power to detect small-to-medium main effects/interactions (*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, 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 control items. Second, participants studied all lists 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 presentation of the final word list, participants completed a 60-second filler task in which they were given a random consonant and were tasked with generating as many words as possible which started with this letter. Immediately afterwards, participants began the recognition test, which presented 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 significant main effects and interaction, partial eta-squared (*ηp*2)is reported, and Cohen’s *d* effect sizes are reported for significant 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 and, importantly, 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 construction. For completeness, comparisons between list types are reported in Appendix Table A2.

***Experiment 1A.*** Free-recall responses were analyzed using a 3(Encoding Group: Item JOL vs. Global JOL vs. No-JOL) × 2(List Type: Categorized vs. Uncategorized) mixed-measures ANOVA. Overall, the main effect of Encoding Group was marginally significant, *F*(2, 110) = 2.46, *MSE* = .04, *p* = .09, *p*bic = .90. Collapsed across list types, free-recall was highest for participants in the global JOL group (.67), followed by participants in the item JOL group (.61) and the no-JOL control group (.60). Follow-up *t-*tests revealed no significant differences between groups (*t*s ≤ 1.56, *p*s ≥ .12, *p*BICs ≥ .72), except for the comparison between the global JOL and no-JOL groups, which was significant, *t*(72) = 2.03, *SEM* = .04, *d* = 0.46. Additionally, 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, *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, *ηp*2 = .07. Starting with categorized lists, providing global JOLs led to increased free-recall relative to the no-JOL group (.76 vs. .86; *t*(75) = 2.39, *SEM* = .03, *d* = 0.57). Free-recall for participants in the item JOL group, however, did not statistically differ from the no-JOL control group (.71 vs. .68) or the global JOL group, *t*s ≤ 1.58, *p*s ≥ .12, *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, *p*BICs ≥ .87) except for the comparison between the item-JOL and no-JOL groups, *t*(73) = 2.16, *SEM* = .05, *d* = 0.51. Thus the memorial benefits of global JOLs on categorized lists did not extend to uncategorized lists.

***Experiment 1B.*** Next, hit rates were analyzed via a 3 (Encoding Group: Item 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, *ηp*2 = .23, as collapsed across list types, hit rates were highest for participants in the item 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, *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, *p*BIC = .87. Next, a significant main effect of List Type was detected *F*(1, 110) = 67.72, *MSE* = .01, *ηp*2 = .14, 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, *ηp*2 = .04. Overall, hit rates for the item JOL group exceeded the no-JOL group for categorized lists (.91 vs. .83) and uncategorized lists (.87 vs. .68), *t*s ≥ 2.41, *d*s ≥ 0.59. Hit rates for the item JOL group similarly exceeded hit rates for the global JOL group for both list types (categorized lists: .91 vs. .83; uncategorized lists: .87 vs. 67; *t*s ≥ 2.65, *d*s ≥ 0.64). However, no differences in hit rates were detected between the global JOL and no-JOL groups for categorized (.83 vs. .81) and uncategorized lists (.68 vs. .65), *t*s ≤ 1, *p*s ≥ .51, *p*BICs ≥ .87. Thus, the requirement to provide JOLs at encoding benefited correct recognition of items from 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, *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 JOL group (.11). Post-hoc testing indicated that this effect was primarily driven by differences between the item JOL and no-JOL groups, *t*(71) = 2.47, *SEM* = .04, *d* = 0.61. The difference between the global and no-JOL groups was marginal, *t*(74) = 1.94, *SEM* = .04, *p* = .06, *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, *p*BIC = .86.

Finally, following the designs of Myers et al. (2020) and Maxwell and Huff (in press), signal detection analyses were used to test for differences in discriminability (*d′*) and response criterion (*c*) as functions of encoding group. Signal detection indices were computed in *R* via the *psycho* package (Makowski, 2018), which corrects for extreme scores following Hautus’s (1995) guidelines. Overall, mean *d′* differed as a function of encoding group, *F*(2, 110) = 16.32, *MSE* = 0.75, *ηp*2 = .23. Mean *d′* was greatest for the item 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, *d*s ≥ 0.85, except for the comparison between the global and no-JOL groups, which was marginally significant, *t*(74) = 1.83, *SEM* = 0.19, *p* = .07, *p*BIC = .62. Next, mean *c* was greatest for participants making global JOLs (.22), followed by the item JOL group (.16), and the no-JOL group (.03). This effect, however, failed to reach conventional significance, *F*(1, 110) = 2.51, *MSE* = 0.14, *p* = .09, *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 further assessed whether making item-level JOLs would improve free-recall compared to a no-JOL control group. Experiment 1B then tested for these effects using recognition testing. Additionally, both experiments included a global JOL group, who provided a single JOL for each list following study, which allowed for a comparison between encoding tasks which emphasized item-specific and relational processes. Starting with Experiment 1A, item-level JOLs did not produce a memorial benefit on free-recall, and no memory differences were observed between the item-JOL and control groups, regardless of whether memory was tested for categorized or uncategorized lists. This null-reactivity pattern is consistent with previous reactivity patterns reported by Zhao et al. (2022) and is likewise in line with Chang and Brainerd’s (2024) finding that item-level JOLs are not reactive on free-recall of blocked categorized lists. However, global JOLs produced a positive reactivity on categorized lists, as recall exceeded the no-JOL group. Thus, free-recall was only benefited when the JOL task encouraged processing of intra-list relations, rather than item-specific encoding of individual words.

Next, Experiment 1B demonstrated that item-level JOLs improved correct recognition of all studied items, regardless of whether they were presented in categorized or uncategorized lists. This replicates previous findings by Zhao et al. (2022) who similarly reported positive reactivity on uncategorized lists while extending these findings to include categorized lists. Global JOLs, however, were non-reactive, regardless of list type. Additionally, a set of 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 item-level JOLs.

Overall, the finding that item-level JOLs were non-reactive on free-recall is consistent with the item-order account and provides further evidence that item-level JOLs encourages item-specific processing of words. Additionally, the finding that global JOLs 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. Separately, the finding in Experiment 1B that item-level JOLs produced positive reactivity on both list types is similarly consistent with an item-order account of reactivity, as per this account, JOL reactivity is moderated by test type with memorial benefits occurring with recognition but not free-recall testing. 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 experiments tested this by comparing between different JOL tasks 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 utilized in the previous experiments with DRM lists. Like categorized study lists, items in DRM lists are associatively related. However, in addition to containing strong intra-list relations, all items are also strongly related to a non-presented critical lure (see Deese, 1959; Roediger & McDermott, 1995). Previous research has demonstrated that the DRM illusion is differentially affected by item-specific and relational encoding tasks, with item-specific encoding generally leading to a 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 JOLs for each item posits that inhibits the processing of inter-list 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 particpants were again randomly assigned to either the item-level, global, or no-JOL encoding groups. Data screening again used 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* suggested that the final sample was sufficient to detect a medium effect between encoding groups (*d* = 0.60). As with the previous experiments, all participants were native English speakers.

**Materials and Procedure**

Twenty DRM lists were taken from Roediger, Watson, McDermott, and Gallo (2001) to serve 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 (Nelson, McEvoy, & Schreiber, 2004). Lists were then divided into two equal sets of lists to serve as counterbalances, which were matched on BAS. 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 from 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 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 concurrent with study, made list-wise JOLs following the completion of each list, or read each word silently. Like the previous set of experiments, all encoding was self-paced. 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 A3 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, *ηp*2 = .19, as hits were highest for participants in the item 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, *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, *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, *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 JOL group (.62). All comparisons differed significantly (*t*s ≥ 2.39, *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, *p*BIC = .89.

**Signal Detection**

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, *p*BIC = .40, a set of planned post-hoc analyses revealed a significant difference in discriminability between the item-level and no-JOL groups, *t*(69) = 2.08, *SEM* = 0.20, *d* = 0.49. The comparison between the global and no-JOL groups, however, was non-significant, *t*(70) < 1, *SEM* = 0.19, *p* = 0.73, *p*BIC = .89, and the comparison between the item JOL and global JOL groups was marginal, *t*(71) = 1.99, *SEM* = 0.17, *p* = .05, *p*BIC = .55. Separately, response criterion differed between encoding groups, *F*(2, 105) = 6.59, *MSE* = .15, *ηp*2 = .11. Mean *c* was greatest for participants in the no-JOL group (.23), followed by global JOLs (.11) and item JOLs (-.10). All comparisons differed significantly (*t*s ≥ 3.65, *d*s ≥ 0.51), except for the comparison between the global JOL and no-JOL groups, *t*(70) = 1.42, *SEM* = .09, *p* = .16, *p*BIC = .75.

Turning to false recognition, *d′* was numerically highest for the global JOL group (1.35), followed by the item JOL (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, *p*BIC = .94. Similarly, no significant difference in *c* emerged between groups (item JOL = .18, global JOL = -.04, no-JOL = .16; *F*(2, 105) = 2.24, *MSE* = .26, *p* = .11, *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 but not global 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 particpants 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 dissociation emerged between JOL tasks and recognition types, providing further evidence that item-level and global JOLs encourage separate types of processing. Taken together, these findings provide further evidence that JOL reactivity on word lists primarily reflects an item-specific rather than a relational process, as only the JOL task emphasizing item-specific encoding produced a memorial benefit on correct recognition.

**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 the item-order account of JOL reactivity 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 these patterns reported with 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 intra-list 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, Experiments 1A and 1B demonstrated 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 patten is consistent with previous findings reported by Zhao et al. (2022). Similarly, studies investigating JOL reactivity with cue-target word pairs have similarly demonstrated that reactivity is moderated by test type, with item-JOLs being reactive when memory is assessed via cued-recall and recognition but not free-recall testing (Myers et al., 2020). As such, findings from Experiments 1A/1B further support the item-order account’s claim that item-level JOL reactivity reflects enhanced item-specific encoding of list items.

Next, Experiment 2 replicated reactivity patterns observed in Experiment 1B as again item-level but not global JOLs were reactive on correct recognition. Importantly, because participants in Experiment 2 studied DRM lists, this experiment also tested whether each JOL task differentially affected false recognition of critical lures. 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 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. However, only the requirement to provide global JOLs increased the DRM illusion. Thus, compared to item-level JOLs, the relational nature of the global JOL task facilitated activation of the critical lure rather (see Roediger et al. 2001), leading to increased false recognition at test.

Overall, the finding that item-level JOLs improved correct recognition but not free-recall is consistent with prior research suggesting that JOLs encourage item-specific encoding of word list items (e.g., Zhao et al., 2022; 2023). Additionally, this finding is in-line with studies showing that JOLs are similarly reactive on recognition of cue-target word pairs. For example, Maxwell and Huff (in press) recently demonstrated that while JOL reactivity effects were moderated by pre-existing relations when memory was tested via cued-recall, JOLs improved recognition of all studied word pairs, regardless of whether the cue and target were related. They reasoned that the requirement to provide JOLs at encoding encourages processing of both relational and item-specific cues regarding an item’s familiarity, both of which benefit subsequent memory (i.e., cue-strengthening; Soderstrom et al., 2015; see also Koriat, 1997). However, because recognition testing is more sensitive to familiarity cues rather than relational cues (see Koriat & Goldsmith, 1996; Yonelinas, 2002), item-specific processing is more beneficial to this test type than relational encoding. Findings from the present study further support this account, global JOLs, which place more emphasis on relational encoding, were non-reactive when memory was assessed via free-recall.

While the present study provides further evidence that item-level JOL reactivity on word lists reflects an item-specific process, it should be noted that item-level JOLs likely also encourage relational encoding, albeit to a lesser extent than global JOLs. For example, Chang and Brainerd (2024) recently reported that item-level JOLs were reactive on free-recall of categorized word lists, but only when items were studied in randomized lists. By fitting free-recall responses to the dual-retrieval model, the authors determined that the memorial benefits of item-level JOLs on free-recall reflected increased familiarity via gist processing rather than purely via an item-specific process. These findings are consistent with the present study, in which all categorized lists contained exemplars of a single category (i.e., blocked lists) and item-JOLs were non-reactive. Separately, research investigating JOL reactivity on cue-target pairs has demonstrated that reactivity also reflects a relational process within this context, as related but not unrelated pairs show a memorial benefit (e.g., Janes et al., 2018; Soderstrom et al., 2015) and that these memory benefits approximate benefits observed using other encoding tasks which encourage relational encoding (e.g., frequency of co-occurrence judgments, judgments of associative memory; Maxwell & Huff, 2022). Taken together, whether item-JOLs ultimately encourage item-specific or relational encoding appears to be strongly dependent on the type of stimuli that participants study and the method of testing.

Moreover, though the item-level and global JOL tasks were each 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/relational tasks while studying related word lists (e.g., categorized lists, DRM lists), they are likely to engage in some relational encoding. This is because related word lists typically encourage relational encoding of items, regardless of the specific encoding task being applied (Hunt & Seta, 1984). As such, when participants provide item-level JOLs on categorized lists, the nature of the list encourages relational encoding, in addition to item-specific encoding afforded by the task. This issue may be particularly relevant in Experiment 2, as although item-level JOLs did not increase false recognition of critical lures, this JOL type also did not lead to the expected reductions in the DRM illusion compared to the no-JOL control group.

Finally, given the benefits of item-level JOLs on recognition testing that were observed in the present study and similar benefits reported with cue-target word pairs (e.g. Maxwell & Huff, in press; 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 on memory for general knowledge facts (Schäfer & Undorf, in press) and comprehension of text passages (Ariel, Karpicke, Witherby, & Tauber, 2021), these studies have assessed memory using recall testing, rather than recognition. As previously noted, recall and recognition test types differentially emphasize familiarity cues, such that recognition testing is generally more sensitive to familiarity. Thus, if item-level JOLs consistently enhance familiarity via enhanced item-specific encoding, these judgments would be expected to improve memory for educational materials, so long as memory is assessed using recognition-based approaches versus tests assessing recollection. 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 provides an additional evaluation of JOL reactivity on word lists. By comparing between item-level and global JOLs, which differentially emphasize item-specific and relational encoding, each experiment further tested recent work suggesting that JOLs on elicited for word list items primarily reflect enhanced item-specific encoding. Overall, 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. Separately, when the JOL tasks was reframed to encourage relational encoding, memory was only influenced when study items contained pre-existing relations (e.g., categorized lists, the DRM illusion). Taken together, this set of experiments 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 interaction between stimuli and test format. For item-level JOLs, the present, study provides further evidence that reactivity from this type of JOLs reflects the use of item-specific encoding when these judgments are provided individually for study list items.

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

Brainerd, C. J., Reyna, V. F, & Howe, M. L.(2009.) Trichotomous processes in early memory development, aging, and neurocognitive impairment: A unified theory. *Psychological Review,* *116*, 783–832.

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.

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. (in press). Judgment of learning reactivity reflects enhanced relational encoding on cued-recall but not recognition tests. *Metacognition and Learning,* 1-25.

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.

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.

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.

Schäfer, F. & Undorf, M. (in press). 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.

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.

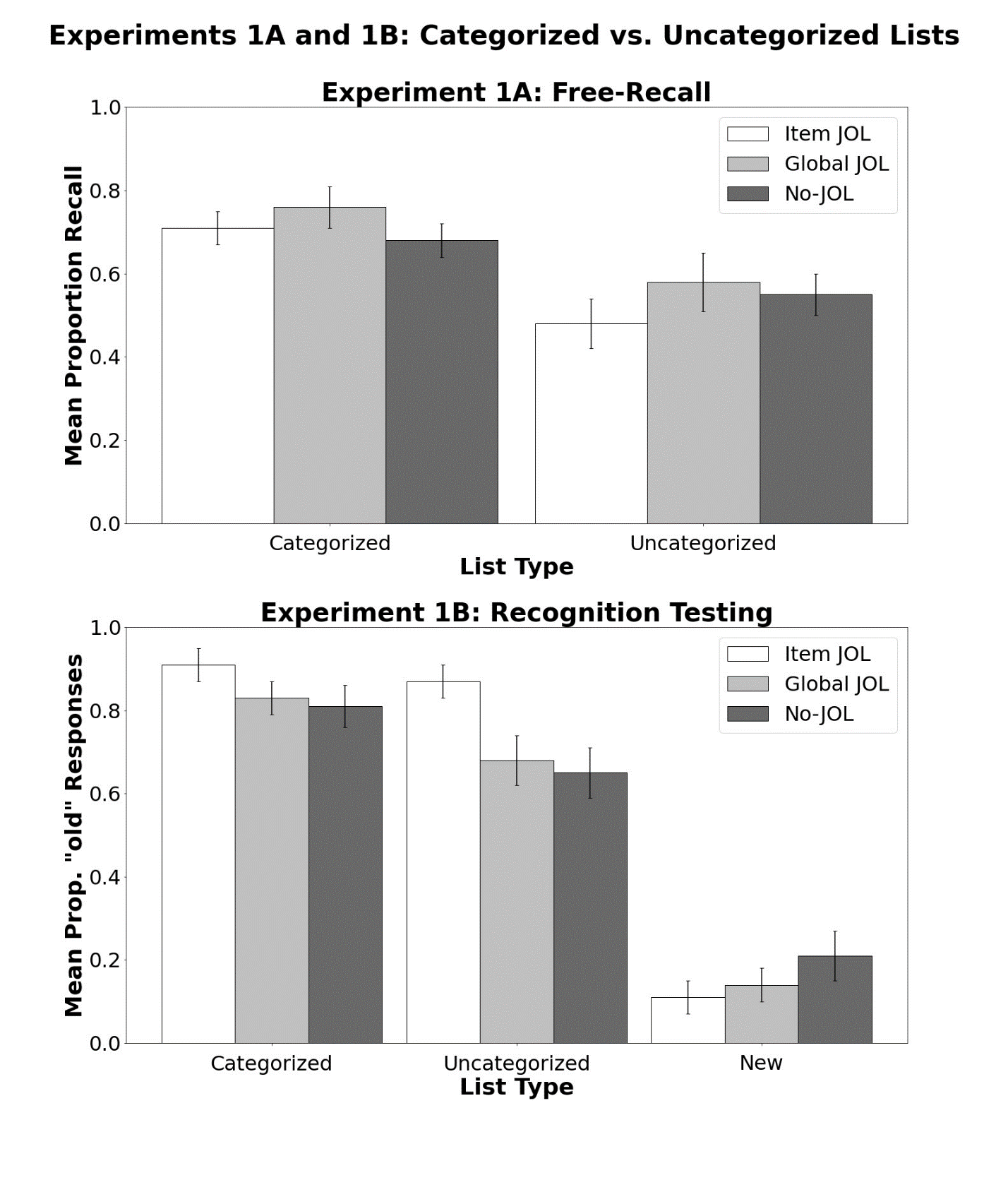
Van Overschelde, J. P., Rawson, K. A., & Dunloskey, J. Category Norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*(3), 289-335.

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

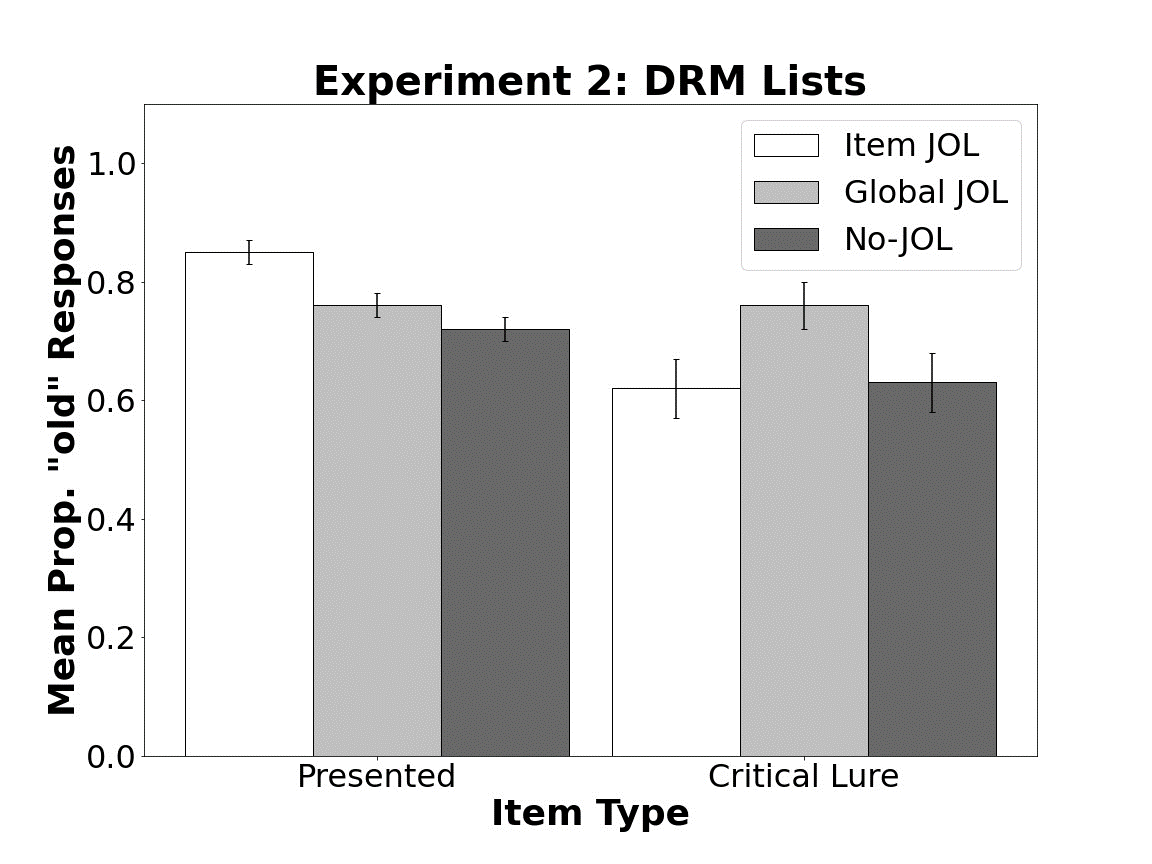
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517.

Zhao, W., Baike, L., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., Ningxin, S., Fan, T., Yin, Y., Luo, L., & Yang, C. (2022). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children’s learning. *Child Development, 93*(2), 405–417.

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.



*Figure 1*. Mean proportion of correct free-recall in Experiment 1A (top panel) and mean proportion of “old” responses in Experiment 1B (bottom panel). “New” columns indicate “old” responses to distractor items. 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.

**Appendix**

Table A1

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

|  |  |  |  |
| --- | --- | --- | --- |
| List Type | Variable | *M* | *SD* |
| Categorized | Concreteness | 6.17 | 0.43 |
|  | Length | 5.50 | 1.58 |
|  | Frequency | 2.97 | 0.57 |
| Uncategorized | Concreteness | 5.64 | 0.80 |
|  | Length | 4.97 | 1.27 |
|  | Frequency | 2.88 | 0.70 |

*Note*: Frequency ratings were derived from SUBLTEX (Brysbaert & New, 2009). Concreteness ratings were taken from Nelson et al. (2004). Values are collapsed across individual study lists for each list type. The full stimuli set has been made available at https://osf.io/4tgj6.

Table A2

*Comparison of Mean Recall Proportions in Experiment 1 and Hit Rates in Experiment 2 for each Encoding Group as Functions of List Type.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Experiment | Encoding Task | List Type | *M* | *± 95% CI* | U |
| Ex. 1A | 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\* |
| Ex 1B. | Item JOL | Categorized | .91 | .04 |  |
|  |  | Uncategorized | .87 | .04 | 0.31 |
|  | Global JOL | Categorized | .83 | .04 |  |
|  |  | Uncategorized | .68 | .06 | 0.92\* |
|  | No-JOL | Categorized | .81 | .05 |  |
|  |  | Uncategorized | .65 | .06 | 1.03\* |

*Notes*: Right-most column denotes Cohen’s *d* effect sizes for post-hoc comparisons. \* = *p* < .05. U = Uncategorized lists. Means for Experiment 1A denote proportion of correct recall. Means for Experiment 1B denote hit rates.

Table A3

*Comparison of Mean Recognition in Experiment 2 for each Encoding Group as Functions of Item Type.*

|  |  |  |  |
| --- | --- | --- | --- |
| Item Type | 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) |

*Note*: Parentheses indicate 95% CIs.