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

[ABSTRACT WILL GO HERE]

Word Count: XXX

*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 important to assess whether the knowledge being acquired is being sufficiently encoded. Metamemory, or the processes by which individuals regulate their learning, is crucial for successful learning, as it helps individuals decide whether items need additional study or if they have been sufficiently learned (see Nelson & Narens, 1990). To investigate questions surrounding the specific metamemory processes individuals engage in at encoding, researchers commonly have participants make Judgments of Learning (JOLs), which involve studying items (commonly cue-target word pairs) while predicting the likelihood that these items will be correctly remembered 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 which do not influence memory, particularly when they were 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, more 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 accounts, 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 performance between participants making JOLs at encoding and a separate group of participants completing a no-JOL control task (e.g., silent reading).

Studies exploring the mechanisms behind JOL reactivity have often tested for reactivity effects on 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: When pairs are semantically related (e.g., mouse – cheese), JOLs improve memory for the target item, but this memorial benefit does not extend to unrelated pairs (e.g., mouse – cup). Several theories have been proposed to explain this pattern. One that has received significant attention in the literature is the cue-strengthening account (Soderstrom et al., 2015), which posits that pairs must contain intrinsic cues which participants use to inform the magnitude of their OLs (see Koriat, 1997) and that memory must be assessed using a method that is sensitive to any cues strengthened by JOLs (i.e., the *cue-strengthening account*). Thus, this account predicts positive reactivity on related pairs but not unrelated cue-target pairs. As a result, recent studies have explored the degree to which relatedness directly contributes to positive JOL reactivity (e.g., Janes et al., 2018; Halamish & Undorf, 2023; Maxwell & Huff, 2022; Maxwell & Huff, 2023; Rivers, Dunlosky, Janes, Witherby, and Tauber, 2023), often by manipulating encoding tasks or using different types of associative cue-target pairs.

While previous reactivity studies have often emphasized the link between cue-target relations and reactivity, the present study focuses instead on the potential for JOLs to 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 focuses on shared properties between studied 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. Thus, the memorial benefits of JOLs could potentially reflect item-specific encoding, relational encoding, or a combination.

Recent studies have explored the potential for JOLs to encourage item-specific encoding, particularly within the context of word list learning rather than cue-target pairs. 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 were similar to 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 making JOLs improved recognition memory (which relies extensively upon item-specific processing) while simultaneously impairing temporal memory (which was assessed via an order reconstruction task, which is inherently relational). These findings were subsequently replicated by Zhao et al. 2023, who similarly showed positive reactivity on recognition memory and negative reactivity on temporal memory. 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. (2023) proposed an item-order account (see McDaniel & Bugg, 2008). Based on this account, word lists contain both item-specific information regarding the individual words and relational information regarding relations between items within a list. 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). However, because free-recall testing also relies on relational memory (McDaniel & Bugg, 2008; Rawson & Zamary, 2019). Because the item-order account posits that JOLs primarily encourage item-specific processing of list items, positive reactivity would be expected to occur when testing occurs via recognition. However, because free-recall also relies upon relational processes, this account predicts weaker reactivity when memory is assessed with this test type. Thus, the item-order account can explain why word list reactivity is moderated by test-type.

**The Present Study**

Given the finding that JOLs are reactive on word list memory, the present study further examined the roles of item-specific and relational processing underlying this effect. First, since previous research has demonstrated mixed findings regarding whether JOLs are reactive on free-recall of word lists, Experiment 1 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 2 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) while also testing whether this effect extends to categorized lists. In doing so, Experiments 1 and 2 provided additional tests of the item-order account, as this account makes diverging predictions regarding reactivity when memory is assessed via free-recall and recognition.

Next, Experiments 3 and 4 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. Like categorized lists, DRM lists are similarly centered around a common theme. However, items in DRM lists are strongly related to a non-studied critical lure (e.g., *bed*, *rest*, *dream*, *nap*, and *pillow* are centered around 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). The use of DRM lists in Experiments 3 and 4 allowed for an assessment of item-specific and relational processes on JOL reactivity for word lists, as previous studies have found that item-specific and relational encoding tasks differentially affect the DRM illusion (e.g., McCabe, Presmantes, Robertson, & Smith, 2004; Huff & Bodner, 2013; Huff & Bodner, 2019).

Finally, in addition to standard, item-level JOLs, each experiment also included a separate group of participants who made global, list-wise JOLs following the completion of each list. Unlike the standard, item-level JOLs, the global JOL task was designed to encourage relational encoding of inter-list relatedness rather than item-specific encoding. As such, in addition to comparisons between participants making item-by-item JOLs and a control group, each experiment also included a comparison between JOL tasks which differentially encouraged item-specific or relational encoding of list items.

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

The goal of Experiment 1 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 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 predicts that item-level JOLs would not be reactive, given that making JOLs inhibits relational processes that are critical for free-recall. Separately, global JOLs were expected to benefit recall, given that this 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.prolic.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 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 1 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 (i.e., 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.

**Results**

For all reported analyses, significance was set at the *p* < .05 level, and 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 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 (top panel) depicts changes in free-recall proportions as functions of encoding group and list construction. For completeness, comparisons between list types are reported in Appendix Table A2. To test for reactivity effects, the data was 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.

**Discussion**

Previous research has reported mixed findings regarding the reactive effects of item-level JOLs on free-recall of word lists. As such, the goal of Experiment 1 was to further assess whether making item-level JOLs would improve free-recall compared to a no-JOL control group. Additionally, Experiment 1 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. Overall, item-level JOLs were not reactive on memory, as no recall differences were observed regardless of whether participants studied categorized or uncategorized lists. However, making global JOLs produced a positive reactivity on categorized lists, as recall exceeded the no-JOL group. Thus, free-recall only benefited when the JOL task encouraged processing of intra-list relations, rather than item-specific encoding of individual words.

Overall, the finding that item-level JOLs were non-reactive on free-recall of word lists is consistent the item-order account and provides further evidence that item-by-item JOLs specifically encourage item-specific processing of words. Separately, the finding that global JOLs benefitted free-recall aligns with other reactivity studies which have demonstrated JOLs improve memory for related but not unrelated cue-target pairs (e.g., Janes et al., 2018; Soderstrom et al., 2015; Maxwell & Huff, 2022). Like making JOLs on cue-target pairs, global JOLs similarly encourage relational encoding, which benefits free-recall. Thus, the finding that global but not item-level JOLs produce positive reactivity on free-recall suggests that these JOL tasks facilitates different types of processing.

Because Experiment 1 found that item-level JOLs were not reactive on word lists with free-recall testing, Experiment 2 further tested the item-order account by assessing whether these judgments would be reactive when memory was tested via recognition. In doing so, Experiment 2 sought to replicate findings reported by Zhao et al. (2023) showing that JOLs were reactive on recognition of uncategorized list items while testing whether these benefits would extend to categorized lists. Finally, Experiment 2 again included a group of participants who made global JOLs, which further tested the effects of item-specific and relational processes on reactivity.

**Experiment 2: Categorized versus Uncategorized Lists and Recognition Testing**

Experiment 2 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 expected to benefit recognition memory for both categorized and uncategorized lists. Additionally, because Experiment 1 demonstrated that global JOLs were reactive on free-recall of word lists, Experiment 2 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 assessing memory via this test type. Thus, Experiment 2 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 2 online in exchange for partial course credit. Participants were randomly assigned to one of the three encoding groups used in Experiment 1. Initial sample sizes for each group were based on the previous experiment, 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 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 2 used the same stimuli and materials as Experiment 1, with the following exceptions. First, the four free-recall tests used in Experiment 1 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 1.

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 1, Experiment 2 took approximately 20 minutes to complete.

**Results**

**Analysis of Hits and False Alarms**

Figure 1 (bottom panel) displays mean hit rates and false alarms as functions of list type and encoding group, and listwise comparisons are reported in Appendix Table A2. To test for reactivity, 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.

Finally, for 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.

**Signal Detection**

Following the design 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**

Experiment 2 tested whether item-level JOLs would produce positive reactivity on word lists when memory was assessed via recognition rather than free-recall. Like the previous experiment, Experiment 2 included a separate group of participants who made global JOLs following each study list, in addition to a no-JOL comparison group. Overall, item-level JOLs produced positive reactivity on both list types, replicating previous findings by Zhao et al. (2023) with uncategorized lists and 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 function of JOLs. Thus, making JOLs improved participants’ discriminability for targets, though the increased criterion levels for JOL participants suggest that providing JOLs at encoding led to more biased responding.

Overall, the finding that item JOLs produced positive reactivity on categorized and uncategorized lists is consistent with an item-order account of reactivity. Considered alongside findings from Experiment 1, there is converging evidence that JOL reactivity on word lists is moderated by test type, likely due to the JOL task encouraging item-specific processing of words. However, in Experiments 1 and 2, the assumption that item-level JOLs specifically encourage item-specific encoding was assessed by comparing between different JOL tasks and test types. A more complete test would be to assess the effects of item-level and global JOLs using stimuli that are differentially affected by item-specific and relational encoding tasks. As such, Experiment 3 replaced the categorized and uncategorized lists studied used in the previous experiments with DRM lists, which are associatively related and centered around a non-presented, but strongly related 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 (McCabe et al., 2004; Huff & Bodner, 2013), though the effects may be moderated by test-type (see Huff & Bodner, 2019). Thus, Experiment 3 tested whether reactivity patterns observed in Experiment 1 with free-recall testing would replicate when participants studied DRM lists while testing the effects of both item-level and global JOLs on the DRM illusion.

**Experiment 3: DRM Lists and Free-Recall Testing**

Experiment 3 provided a further test of whether item-level and global JOLs would be reactive on word list learning. Unlike the previous experiments, which utilized categorized and uncategorized word lists, participants in Experiment 3 studied DRM lists, which allowed for a comparison of item-specific/relational encoding effects on both correct and false recall. Based on an item-order account, item-level JOLs were not expected to be reactive on correct recall, given this test type’s emphasis on inter-item relations. However, based on findings from Experiment 1, global JOLs were expected to produce positive reactivity. Regarding false recognition, the item-order account predicts that item-JOLs would produce negative reactivity. This is because previous research has consistently found that tasks which encourage item-specific encoding are effective at reducing the DRM illusion (e.g., Burns, Jenkins, & Deans, 2007; Huff & Bodner, 2013; McCabe et al., 2004). Regarding global-JOLs, previous research is mixed regarding their effects on the DRM illusion, with some studies showing that relational encoding can produce reduction in false memories relative to silent reading, though the effect is smaller relative to item-specific encoding (Huff & Bodner, 2019). However, because relational encoding emphasizes shared connections between list items, these tasks may increase false recognition of critical lures via spreading activation (see Roediger, Balota, & Watson, 2001). Thus, making global JOLs may instead produce positive reactivity on false recall via enhanced relational encoding of critical lures.

**Method**

**Participants**

An additional XXX participants were recruited via Prolific and completed Experiment 4 in exchange for $4.00 per 30-minute session, and participants were again randomly assigned to either the item-JOL, global JOL, or no-JOL groups. Data screening followed the same procedure utilized in the previous experiments, which led to the exclusion of XX participants. As a result, data from XXX participants is reported in the following set of analyses (item JOLs, *n* = xx; global JOLs, *n* = xx; no-JOL control group, *n* = xx). The final sample was consistent with the previous experiments, and a sensitivity analysis conducted with *G\*Power* suggested that this sample was sufficient to detect XX main effects/interactions (*d* = 0.xx). Finally, all participants were native English speakers and were additionally required to have obtained at least a high school degree or equivalent.

**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. All other materials, including the filler and free-recall tasks, were identical to Experiment 1. Finally, the general procedure was identical to Experiment 1, with the exception that participants studied DRM lists instead of categorized and uncategorized lists and that the total number of studied lists was increased from four to 10. The full experiment took approximately 30 minutes to complete.

**Results**

[WORDS HERE]

**Experiment 4: DRM Lists and Recognition Testing**

[WORDS HERE] [PREDICTIONS – ITEM JOLS SHOULD LOWER FALSE RECOGNITION] [GLOBAL JOLS SHOULD INCREASE?]

**Method**

**Participants**

One hundred twelve undergraduate students were recruited from Midwestern State University and completed Experiment 4 for partial course credit. This sample size was based on the previous experiments, and particpants were again randomly assigned to either the item JOL, global JOL, or no-JOL encoding groups. Data screening again used the same criteria as the previous experiments, and data from four participants were excluded. As such, the final dataset contained responses from 108 participants (item 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**

Experiment 4 followed the same general procedure outlined in Experiment 3, except that the free-recall tests were replaced with a single, 80-item old/new recognition test, which occurred following completion of the final word list. This test consisted of 30 previously studied items (taken from positions 2, 8, and 10 from each list), 10 critical lures from each studied list, 30 non-presented items taken from counterbalanced lists, and 10 critical lure controls which were taken from the non-studied set of lists. 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 made JOLs concurrent with study, made list-wise JOLs following the completion of each list, or read each word silently. Like the previous experiments, all encoding was self-paced. Experiment 4 took approximately 30 minutes to complete.

**Results**

**Analyses of Correct and False Recognition**

Figure 2 (bottom panel) 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, with false recognition highest for participants making global JOLs (.76), followed by the no-JOL control group (.63), and the item JOL group (.63). All comparisons differed significantly (*t*s ≥ 2.39, *d*s ≥ 0.55), except for the comparison between the item JOL and no-JOL control groups, *t*(69) < 1, *SEM* = .06, *p* = .88, *p*BIC = .89.

**Signal Detection**

Next, a set of signal detection analyses tested 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 JOL 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.

For 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**

[COMPARE FINDINGS TO EX 3] [DISSOCIATION BETWEEN JUDGMENT TYPES AND REACTIVITY] – [GLOBAL JOLS REACTIVE ON FALSE RECOGNITION, ITEM JOLS REACTIVE ON CORRECT RECOGNITION]

**General Discussion**

[RECAP GOAL OF THE EXPERIMENT] [SUMMARY OF EXPERIMENTS]

[CONSISTENCY WITH PREVIOUS EXPERIMENTS]

[SENKOVA & OTANI]

[ZHAO ET AL] [ITEM SPECIFIC PROCESSES PRIMARILY ENHANCE RECOLLECTION – SEE HUFF & BODNER 2019]

[CHANG AND BRAINARD 2024 INTELLIGENCE PAPER]

[DRM FINDINGS – TIE BACK TO PREVOUS WORK?]

[RECOGNITION TESTING FAMILIARITY ACCOUNT]

**JOL Reactivity and DRM Lists**

[PREVIOUS RESEARCH USING DRM FOR ISREL]

[ISREL TASKS NOT PURE – JACOBY, 1991]

[FUTURE RESEARCH]

**Conclusion**

[WORDS HERE]

**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 #XXXX). All participants provided consent prior to participating in the experiments. The author reports no conflicts of interest. Data collection was partially funded by an intramural research grant awarded to the author by Midwestern State University.

**References**

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. (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.

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.

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., & 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.

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.

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.

A graph of different types of testing

Description automatically generated with medium confidence

*Figure 1*. Mean proportion of correct free-recall in Experiment 1 (top panel) and mean proportion of “old” responses in Experiment 2 (bottom panel). “New” columns indicate “old” responses to distractor items. Bars indicate 95% *CI*s.

FIGURE 2

A graph of different sizes and colors

Description automatically generated with medium confidence

*Figure 4.* Mean proportion of “old” responses to presented items and critical lures in Experiment 4. Bars indicate 95% *CI*s.

**Appendix**

Table A1

*Lexical Properties of Categorized and Uncategorized Lists in Experiments 1 and 2.*

|  |  |  |  |
| --- | --- | --- | --- |
| 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. 1 | 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 2. | 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 1 denote proportion of correct recall. Means for Experiment 2 denote proportion of correct recognition.

Table A3

*Comparison of Mean Recall Proportions in Experiment 3 and Hit Rates in Experiment 4 for each Encoding Group as Functions of Item Type.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment | Item Type | Item JOL | Global JOL | No JOL |
| Ex. 3 | List items |  |  |  |
|  | Critical Items |  |  |  |
|  | Intrusions Per List |  |  |  |
| Ex. 4 | 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) |

*Notes*: Cells for Experiment 3 denote proportion of correct recall. Cells for Experiment 4 denote proportion of correct recognition. Parentheses indicate 95% CIs.