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 [LINK]

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 has been sufficiently encoded. Metamemory, or the processes by which individuals regulate their learning, is crucial for successful learning, as these processes help 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; Rhodes & Castel, 2008; Maxwell, Perry, & Huff, 2022) and the presence of associations between items in cue-target pairs (e.g., Koriat & Bjork, 2005; Castel, McCabe, & Rhodes, 2007).

Historically, research involving JOLs considered these ratings neutral measures with no effect on memory, particularly when they were elicited concurrently with or immediately following study (though see Spellman & Bjork, 1992, who posited that JOLs made following a delay improved memory for judged items). As such, early JOL studies often focused on factors influencing the accuracy of these judgments, rather than their potential to influence memory. 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 memorial 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 investigating the mechanisms behind JOL reactivity have often used cue-target word pairs (e.g., Janes et al., 2018; Maxwell & Huff, 2022; Maxwell & Huff, 2023; Mitchum, Kelly, & Fox, 2016; Myers et al., 2020; Soderstrom et al., 2015; though see Senkova & Otani, 2021, who had participants study single word lists). These studies have revealed a consistent reactivity 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). To explain reactivity is moderated by relatedness, Soderstrom et al. proposed that two conditions must be met for reactivity to occur on cue-target pairs. First, studied items must contain intrinsic cues which participants use to inform their JOLs (see Koriat, 1997). Second, participants must be tested using a method that is sensitive to any cues that are strengthened due to making JOLs. Thus, the *cue-strengthening account* predicts a memory benefit on related pairs, particularly when testing is sensitive to strengthened cues (e.g., cued-recall testing).

Although the cue-strengthening account does not specify the exact cues strengthened by JOLs, it is likely that providing these judgments for cue-target pairs strengthens pre-existing pair relations given that relatedness is a highly salient cue for later remembering (Mueller, Tauber, & Dunlosky, 2013). Given this possibility, recent studies have explored the degree to which relatedness contributes to positive reactivity (e.g., Halamish & Undorf, 2023; Maxwell & Huff, 2022; Rivers, Janes, Dunlosky, Witherby, and Tauber, 2023), often by manipulating encoding tasks or using different types of associative cue-target pairs. A separate line of research has investigated whether JOLs are similarly reactive on the learning of word lists. Because previous research has shown that list relatedness similarly influences the magnitude of participants’ JOLs (e.g., Matvey, Dunlosky, & Schwartz, 2006) and that related (i.e., categorized lists) are recalled at greater rates versus unrelated (uncategorized lists; e.g., Chang and Brainard, 2023), participants similarly use relatedness cues to inform the magnitude of their JOLs for word lists. As such, relatedness effects observed with cue-target pairs may extend to categorized and uncategorized single word lists.

While relatedness has been shown to influence the magnitude of JOLs regardless of whether it is manipulated intra-item (i.e., related and unrelated cue-target pairs) or inter-item (i.e., categorized and uncategorized word lists), the mechanisms driving these effects likely differ based on the type of stimuli. Based on Koriat’s (1997) cue-utilization account, participants base their JOLs on *intrinsic* and *extrinsic* cues. For cue-target pairs, relatedness reflects an intrinsic cue, as the relation between items is inherent to the pair and easily processed at encoding. However, because by nature, single-word lists display study items to participants one item at a time, any relatedness cues within this context reflect *extrinsic cues* (i.e., cues reflecting the context in which encoding occurs rather than inherent properties of the stimuli). This is because when processing inter-item relatedness within a word list, participants must assess how the currently presented item relates back to each of the previously presented items within the same list. Thus, relatedness can serve as either an intrinsic or extrinsic cue, with the type of stimuli ultimately dictating how this cue is processed.

Regarding JOL reactivity on word lists, previous research has explored whether categorized and uncategorized lists show similar reactivity patterns as related and unrelated cue-target pairs. For example, Senkova and Otani (2021) demonstrated that making JOLs improved free-recall of categorized lists versus uncategorized lists and 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). Other studies, however, have found that word list reactivity is moderated by test type, with JOLs imparting a memorial benefit when testing occurs via recognition but not free-recall (e.g., Zhao et al., 2023; Rhodes & Tauber, 2011). Finally, Zhao et al. recently found that JOLs impaired temporal order memory for word lists but improved recognition memory, leading the authors to conclude that JOLs facilitate item-specific processing of word lists while simultaneously suppressing relational processing of list items. However, because Zhao et al.’s did not include categorized lists, it remains unclear whether these effects may be moderated by list relatedness.

Because JOL reactivity theories derived from studies utilizing cue-target pairs often emphasize the roles of relational processing and intrinsic cues, they have difficulty explaining reactivity on word list learning. Recently, Zhao et al. (2023) proposed an item-order account of JOL reactivity on word-lists, which posits that JOLs facilitate memory for individual items within a word list via enhanced item-specific processing yet decrease memory for inter-item relations (see McDaniel & Bugg, 2008). Based on this account, the memorial benefits of JOLs on words lists observed with free-recall (Senkova & Otani, 2021) and recognition testing (Zhao et al., 2021; Zhao et al., 2023) occur because making JOLs encourages item-specific encoding of list items, rather than relational encoding as has been theorized for JOL reactivity observed with cue-target word pairs (e.g., Soderstrom et al.’s 2015 cue-strengthening account; see also Maxwell & Huff, in press). Thus, while JOLs are reactive on learning of both word lists and cue-target pairs, the underlying mechanisms likely differ.

**The Present Study**

Because previous studies have shown that JOLs are reactive on memory for word lists, one goal of the present study was to provide a further examination of list-wise relations as a potential factor contributing to reactivity effects. First, since previous research has demonstrated mixed findings regarding JOL reactivity on word lists and free-recall testing, 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 testing occurs via free-recall testing. Likewise, Experiment 2 sought replicate findings showing that JOLs are reactive on word lists when memory is assessed via recognition testing (e.g., Zhao et al., 2023). Next, Experiments 3 and 4 were designed to provide further tests of the item-order account using free-recall and recognition testing, respectively. Because a key aspect of the item-order account is that making JOLs promotes item-specific processing while decreasing relational processing, participants in Experiments 3 and 4 studied Deese-Roediger-McDermott lists (DRM; Deese; 1959; Roediger & McDermott, 1995), rather than categorized and uncategorized word lists. In doing so, this 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, which was designed to encourage relational encoding of lists rather than item-specific encoding. As such, each experiment provided a comparison of 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 provide a greater free-recall benefit on categorized lists compared to uncategorized lists. In doing so, free-recall rates were compared between three groups: Participants who provided concurrent, item-level JOLs, 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 (2021), JOLs were expected to benefit free-recall of categorized lists. However, if JOL reactivity on word lists reflects an item-specific process as posited by the item-order account, only item-level JOLs would be expected to show positive reactivity. Alternatively, if reactivity also reflects a relational process, global JOLs would similarly be expected to benefit recall, as global JOLs emphasize intra-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 direct comparison between item-specific and relational focused judgments.

**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 and interactions (ds ≥ 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 X 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 participants, 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 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 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. Experiment 1 additionally 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 was only benefited when the JOL task encouraged processing of intra-list relations, rather than item-specific encoding of individual words.

Overall, the finding that global JOLs benefitted recall is consistent with other reactivity studies showing 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 global JOLs on word lists, JOLs for cue-target pairs similarly encourage processing of intra-item relations (see Halamish & Undorf, 2023; Maxwell & Huff, 2022). Finally, the finding that item-JOLs produced no benefit on free-recall of word lists is similarly consistent with previous research suggesting that reactivity on word lists is not likely to occur with this test type (e.g., Zhao et al., 2023; though see Senkova & Otani, 2021 who demonstrated positive reactivity on free-recall of both categorized and uncategorized words lists). As such, Experiment 2 provided a further test of JOL reactivity effects on word list by replacing the free-recall tests with recognition testing.

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

Because Experiment 1 found that item-level JOLs were not reactive on free-recall of word lists, the goal of Experiment 2 was to provide a further test of the item-order account by assessing whether previously reported JOL 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, making item-level JOLs would be expected to produce a reactivity effect on study lists. Furthermore, because the item-order account makes no claims regarding the effects of list construction on reactivity, item level JOLs were expected to produce a memorial benefit on both list types. Additionally, because Experiment 1 demonstrated that global JOLs are reactive on free-recall of word lists, Experiment 2 similarly included a group of participants who made single, global JOLs following the presentation of each study list. However, given that recognition testing is less sensitive to relatedness cues, it was unclear whether global JOLs would be reactive when memory was assessed via this test type (see Maxwell & Huff, in press). Thus, Experiment 2 provided an additional test of the item-order account while also 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 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 in the final sample (item JOLs, *n* = 37; global JOLs, *n* = 40; no-JOL control group, *n* = 36). This sample was based on Experiment 1, and a sensitivity analysis conducted with *G\*Power* suggested that the final sample had sufficient power to detect small-to-medium main effects/interactions (*d*s ≥ 0.29). All participants were native English speakers.

**Materials and Procedure**

Experiment 2 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 1, Experiment 2 took approximately 20 minutes to complete.

**Results**

**Analysis of Hits and False Alarms**

Figure 2 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 item JOLs would produce positive reactivity on word lists when memory was assessed via recognition testing rather than free-recall. As with the previous experiment, Experiment 2 similarly included a separate group of participants who made global JOLs following each study list, in addition to a no-JOL comparison group. Overall, item 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. Considered alongside findings from Experiment 1, there is converging evidence that JOL reactivity on word lists is moderated by test 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 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 increased response bias.

Overall, the finding that item JOLs produced positive reactivity on categorized and uncategorized lists is consistent with an item-order account of reactivity. However, as noted in the Introduction, the item order account assumes that JOLs specifically encourage item-specific encoding of words within each list. In Experiments 1 and 2, this assumption was tested by including a separate group of participants who made global JOLs. Thus, global JOLs were similar to JOLs elicited for cue-target pairs, as they required participants to consider all list items as a single unit. Thus, the global JOL task was designed to encourage relational, rather than item-specific, encoding of list items. However, because previous experiments did not assess the specific types of processing being encouraged by each JOL task, making it unclear whether item JOLs and relational JOLs were respectively encouraging item specific and relational processing. As such, Experiments 3 and 4 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, which is often falsely remembered at test (e.g., *bed*, *rest*, *dream*, *nap*, and *pillow* are centered around the non-presented word *sleep*; see Deese, 1959; Roediger & McDermott, 1995). By including DRM lists, Experiments 3 and 4 provided stronger tests of the mechanisms underlying both item and global JOLs, as relational but not item-specific tasks would be expected to increase false recall/recognition of critical lures. Thus, Experiments 3 and 4 provided additional tests of whether reactivity patterns observed in the previous experiments would replicate within this context while also testing the processes by which JOLs modify memory for word lists.

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

[GOAL OF EXPERIMENT 3] [WHY DRM LISTS – LINK TO ITEM SPECIFIC/RELATIONAL PROCESSING]

**Method**

**Participants**

[WORDS HERE] [PROBABLY PROLIFIC]

**Materials**

[WORDS HERE]

**Procedure**

[WORDS HERE]

**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.

**Results**

**Analyses of Correct and False Recognition**

Figure 4 shows mean hit rates for studied items and false recognition of critical lures for each encoding group (see Appendix Table AX for all comparisons). First, to test for potential JOL reactivity on correct recognition of list items, hit rates 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, with correct recognition 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 2018]

[DRM FINDINGS – TIE BACK TO PREVOUS WORK?]

[RECOGNITION TESTING FAMILIARITY ACCOUNT]

**JOL Reactivity and DRM Lists**

[PREVIOUS RESEARCH USING DRM FOR ISREL]

[FUTURE RESEARCH]

**Conclusion**

[WORDS HERE]

**Declarations**

**Open Practices Statement**

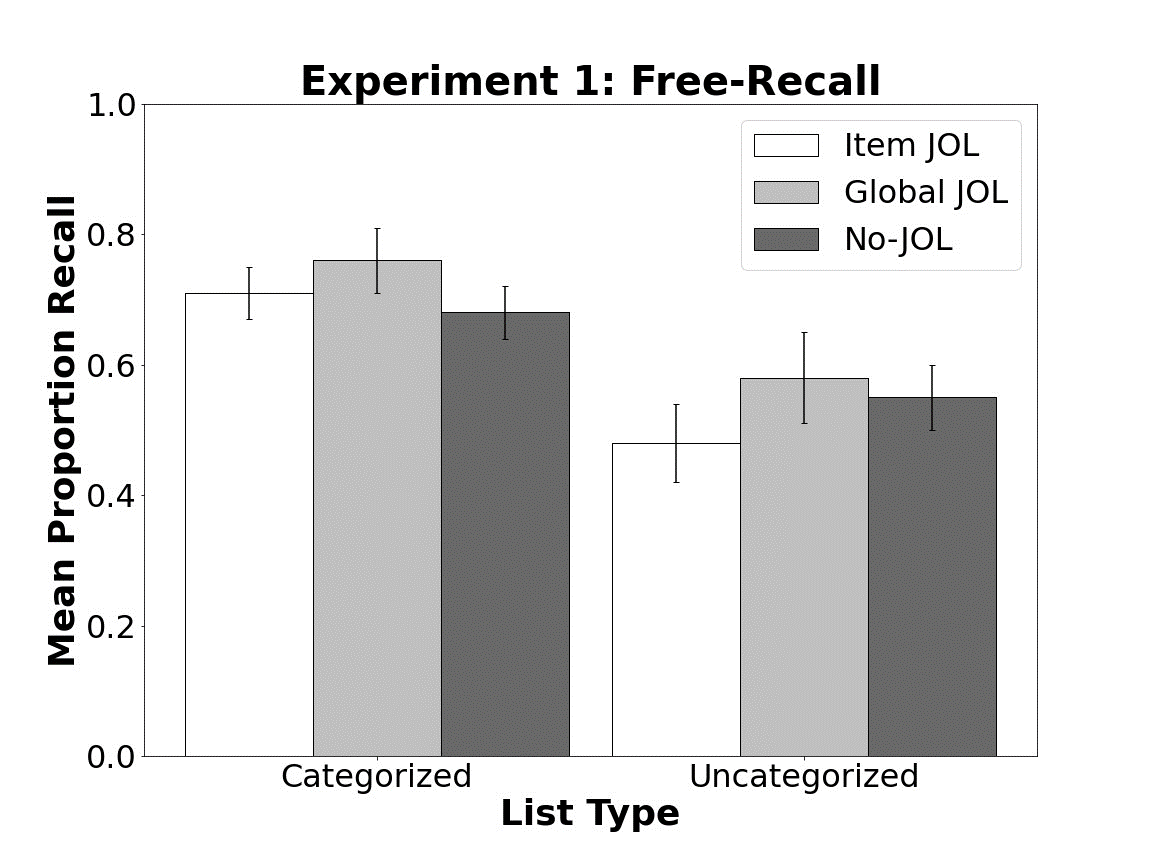
[WORDS HERE]

**Compliance with Ethical Practices**

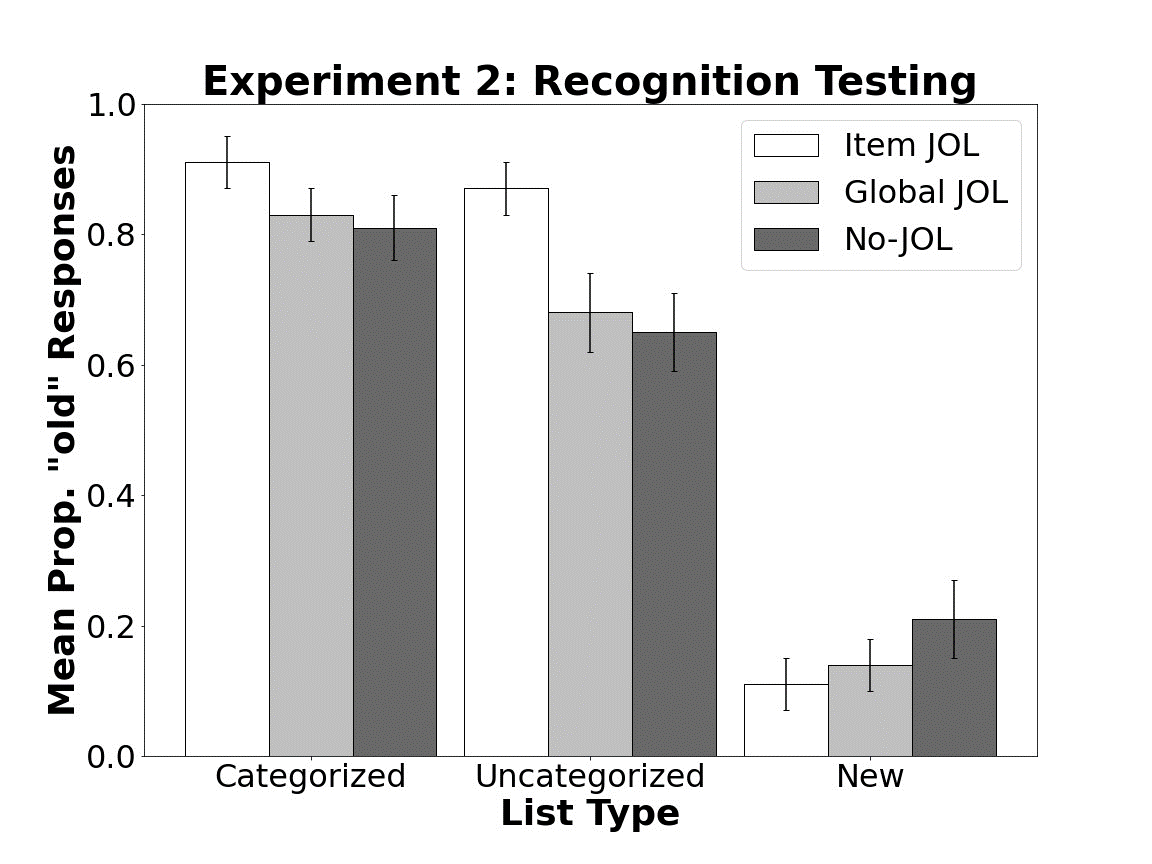
[WORDS HERE] [CONFLICTS OF INTERESTS AND FUNDING]

**References**

[FIRST REF HERE]



*Figure 1*. Mean proportion of correct free-recall in Experiment 1. Bars indicate 95% *CI*s.



*Figure 2.* Mean proportion of “old” responses in Experiment 2. “New” columns indicate “old” responses to distractor items. Bars indicate 95% *CI*s.

FIGURE 3

Figure 4

**Appendix**

[TABLE A1]

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

[DRM FREE RECALL]

Table A4

[DRM RECOGNITION]