Metacognition and Learning (2026) 21:2
https://doi.org/10.1007/s11409-025-09451-z

Rhythm is more than just an aesthetic experience: The rhyme effect on judgments of learning

Mengqi Hu[1] · Anran Li[1] · Chenyuqi Xu[2,3] · Hongyu Chen[4] · Aike Shi[5] · Wenbo Zhao[6] · Xiao Hu[7] · Liang Luo[1,8] · Chunliang Yang[1,7]

Received: 9 August 2024 / Accepted: 10 December 2025

Abstract
Previous research has extensively examined how various factors (e.g., font size, relatedness) influence Judgments of Learning (JOLs), yet comparatively little is known about whether phonological factors, such as rhyme, similarly influence metacognitive judgments. Across four experiments, the present study investigated how rhyme affects memory performance, JOLs, and metacognitive accuracy, and explored the underlying mechanisms involved. Experiments 1 and 2 revealed that rhyme yielded reliable mnemonic benefits only when study involved reading aloud, emphasizing the role of auditory encoding in enhancing memory. Moreover, participants consistently expected higher recall for rhyming than non-rhyming pairs. Experiment 3 confirmed that participants explicitly believed rhyme enhances memorability. Experiment 4 further showed that these beliefs partially mediated rhyme effects on JOLs and that learners updated these beliefs based on study experience. Importantly, despite context-dependent mnemonic benefits, rhyme reduced relative metacognitive accuracy, diminishing item-level discrimination between remembered and forgotten materials. Together, these findings suggest that phonological cues such as rhyme robustly shape metacognitive judgments while sometimes compromising relative accuracy. Practically, instructors and learners should use rhyme judiciously and combine it with more diagnostic cues to support both learning and monitoring accuracy.

Keywords Rhyme · Judgments of learning · Metamemory beliefs · Memory

Rhyme, defined as the repetition of similar sounds in final stressed syllables (Fabb, 1997), is a pervasive linguistic device with well-documented mnemonic benefits (e.g., Gupta et al., 2005; Király et al., 2017; Plastikwala, 2017). Its structured regularity enhances prosodic predictability, reduces phonological working memory load, and facilitates encoding, consolidation, and retrieval through chunking mechanisms (Lindstromberg & Boers, 2008; Nelson et al., 1987; Obermeier et al., 2013). While extensively used in poetry, advertising, and education, its contribution to metacognitive monitoring remains underspecified, particularly in learners' Judgments of Learning (JOLs). In self-regulated learning, accurate JOLs are crucial for selecting effective study strategies and achieving academic success (Metcalfe

Extended author information available on the last page of the article

Published online: 16 January 2026

& Finn, 2008; Yang et al., 2017). However, learners frequently rely on cues that are salient yet not consistently diagnostic of memory strength, such as perceptual features like font size, which can inflate JOLs without improving actual memory performance (Mueller et al., 2014; Rhodes & Castel, 2008; Yang et al., 2018). This tension motivates examining whether phonological structures such as rhyme influence memory and the accuracy of metacognitive judgments and, in particular, whether increases in JOLs under rhyme reflect subsequent memory or instead arise from a salient cue with limited diagnostic value.

Previous research has primarily investigated the influence of rhyme on JOLs within levels-of-processing contrasts. For instance, Bieman-Copland and Charness (1994) compared JOLs and recall performance across three encoding conditions: orthographic (letter-cued, e.g., ic-ice; shallow processing), phonological (rhyme-cued, e.g., hurt-dirt; intermediate processing), and semantic (meaning-cued, e.g., diagram-chart; deep processing). Their findings revealed a striking metacognitive misalignment: in the first study block, participants assigned significantly higher JOLs to rhyming word pairs than to orthographically similar items, yet recall rates were statistically equivalent across these conditions. This dissociation suggests that learners may misattribute mnemonic effectiveness to salient but weakly diagnostic phonological regularities under some task settings, much like the font-size effect in metamemory, where visually salient stimuli artificially boost JOLs without improving memory (Mueller et al., 2014). To better understand rhyme's impact on metacognitive accuracy, it is essential to systematically compare rhyming and non-rhyming stimuli within a unified experimental paradigm. Such comparisons will help determine whether learners' JOLs for rhyming materials accurately reflect their actual retention or if rhyme induces a mismatch between perceived and actual memory performance.

Moreover, the cognitive mechanisms underlying the impact of rhyme on JOLs remain unclear. According to the dual-basis model, processing fluency and ingrained beliefs both play crucial roles in JOL formation (Koriat, 1997). It is reasonable to infer that rhyme could enhance JOLs through processing fluency, as learners are likely to experience greater ease of encoding when studying rhyming materials (McGlone & Tofighbakhsh, 2000). However, it remains unknown whether beliefs also contribute to the effect of rhyme on JOLs. Learners may consciously associate rhyme with enhanced recall, influenced by cultural exposure (e.g., nursery rhymes) or formal instructions, independent of their immediate processing experience. Yet, no study has directly tested whether such beliefs exist or whether they dynamically adjust in response to task feedback. Addressing these questions has both theoretical and practical implications. Theoretically, identifying belief-driven influences on JOLs would refine our understanding of how phonological features interact with metacognition. Practically, if beliefs shape JOLs, educators could implement targeted interventions to recalibrate learners' metacognitive monitoring: either by reinforcing the mnemonic benefits of rhyme or by mitigating overestimation biases.

Rhyme effect on memory

Rhyme is a fundamental linguistic feature commonly used in poetry, advertising, and everyday communication, often recognized for its potential to enhance recall (Avons et al., 1998; Bryant et al., 1990). Psychological research has explored its mnemonic effects, with studies suggesting that rhyme can aid memory by imposing structured patterns, lowering cogni-

tive load, and facilitating retrieval through phonological associations (Carr & Miles, 1997; Johnson & Hayes, 1987; Macaruso et al., 1996). These benefits are often attributed to the rhythmic and predictable nature of rhyming structures, which can reinforce memory traces through repeated phonological exposure. However, rhyme's effectiveness as a mnemonic device is not universal; its impact on recall varies depending on cognitive conditions and encoding strategies, rather than being an inherently robust memory aid.

One key factor shaping rhyme's mnemonic effectiveness is the degree of phonological engagement. When individuals deliberately attend to rhyme, it can serve as an effective retrieval cue (Hoorn, 1996; Nelson et al., 1992). For instance, Nelson et al. (1987) demonstrated that rhyme significantly improved recall performance when participants were instructed to process phonological similarities, whereas directing attention to semantic content eliminated this advantage. This pattern indicates that rhyme's mnemonic utility depends on task demands and encoding orientation: conditions that heighten awareness of phonological structure tend to reveal benefits, whereas tasks that emphasize meaning over sound do not.

Rhyme effect on JOLs

While the role of rhyme in memory has been extensively studied, its broader cognitive effects, particularly its influence on metacognitive monitoring, remain largely unexplored. In the field of metacognition, JOLs index learners' prospective evaluations of memory and guide self-regulated study (Koriat, 1997; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991). Because these item-by-item predictions of future recall inform strategy selection and time allocation, it is crucial to identify stimulus features that systematically influence them. Rhyme, as a salient form of phonological structure, is a prime candidate for such an influence and may shape prospective judgments even before objective memory benefits are observed. Although prior work has utilized rhyme, it has typically served as a comparative condition within broader manipulations rather than as the focal variable. Consequently, the independent influence of rhyme on JOLs and its theoretical status within metacognitive frameworks remain underspecified.

Rhyme is a salient phonological regularity whose predictability can increase subjective ease during encoding, making items seem easier to learn (Obermeier et al., 2016). Within the *cue-utilization framework*, JOLs are constructed from available cues and weighted by their perceived diagnostic, so phonological structure that feels informative is expected to elevate judgments (Koriat, 1997). *Dual-basis accounts* likewise hold that both experience-based fluency and theory-based beliefs contribute to JOLs; everyday exposure to rhymed language can therefore foster beliefs that rhyme aids memory, further elevating judgments (Mueller et al., 2016). Critically, phonological fluency is distinct from perceptual fluency, as it engages auditory-verbal processes rather than surface perceptual properties (e.g., Fu et al., 2006; Molinari & Leggio, 2016), potentially making its metacognitive consequences more context-dependent. These theoretical considerations converge on a clear prediction: rhyming materials should receive higher JOLs than non-rhyming ones, and this effect should be magnified when study conditions amplify phonological engagement.

Indirect evidence aligns with this prediction. For example, in levels-of-processing comparisons, Bieman-Copland and Charness (1994) found that participants initially assigned

higher JOLs to phonological (rhyme-cued) than to orthographic items, even when recall did not differ. Further support for an a priori belief in rhyme's effectiveness comes from Matvey et al. (2002), who reported that global predictions at the first study trial favored rhyming items. This pattern is also consistent with findings from paradigms using formal overlap, where learners often predict better memory for items with surface similarity over those with semantic links, even when recall favors the latter (Castel et al., 2007; Mueller et al., 2016). Taken together, these findings indicate that phonological structure can elevate mean JOLs and therefore motivate a direct test, within a unified paradigm, of whether rhyme reliably increases item-by-item judgments. Having established a rationale for a rhyme-driven elevation in mean JOLs, we next ask whether these judgments remain well-calibrated and item-diagnostic.

Absolute accuracy (calibration) captures how closely judgments track overall performance. In studies of font size, JOLs often rise while recall changes little, producing overconfidence when size varies from very small to intermediate (Chang & Brainerd, 2022; Rhodes & Castel, 2008). In contrast, when materials provide semantic support, JOLs and memory typically increase together and the calibration is preserved. For instance, relatedness elevates both JOLs and recall, yielding no reliable difference in absolute accuracy between related and unrelated items (Mueller et al., 2013). Prior work on rhyme indicates that its mnemonic benefits are conditional, emerging most strongly when phonological processing is emphasized but remaining weak under neutral study conditions (Hoorn, 1996; Nelson et al., 1992). Based on this profile, we expect rhymed items to receive higher JOLs than non-rhyming items, while gains in recall may be modest in many settings. This combination creates the potential for overestimation at the aggregate level for rhymed items.

Beyond absolute accuracy, relative accuracy (item-level discrimination; the extent to which JOLs distinguish remembered from forgotten items) is also crucial for self-regulated learning because it allows learners to separate well-learned from poorly learned items to guide study decisions (Dunlosky & Metcalfe, 2009; Metcalfe & Finn, 2008). Research indicates that relative accuracy improves when learners utilize diagnostic, item-specific cues. For example, additional practice and retention intervals that foster retrieval sampling reliably increase discrimination (Dunlosky & Nelson, 1994; Tauber & Rhodes, 2012). By contrast, salient but weakly diagnostic cues such as font size elevate JOLs yet typically do not enhance discrimination (see also mixed outcomes in Chang & Brainerd, 2023; Hu et al., 2015). Work on relatedness clarifies the mechanism: learners often base JOLs on mediator production during study, and because this cue is more diagnostic for unrelated than related pairs, discrimination is higher for unrelated items (Dunlosky & Matvey, 2001; Hertzog et al., 2002).

This principle leads to a clear prediction for rhyme. Rhyme is a salient, global phonological cue, but its item-level diagnosticity is often limited. If learners overweight this cue when forming their JOLs, their judgments will be uniformly elevated across all rhyming items. This over-reliance may in turn cause learners to neglect more diagnostic, item-specific cues (e.g., successful mediator formation or semantic linkage). Such a process would compress the JOL scale for rhyming items and reduce item-by-item differentiation, ultimately resulting in lower relative accuracy.

Role of beliefs in JOL formation

Rhyme can increase processing fluency, which should raise JOLs for rhymed relative to non-rhymed items. However, fluency alone may not fully account for rhyme effects on JOLs. Within the cue-utilization framework, JOLs are constructed from available cues and weighted by their perceived diagnosticity (Koriat et al., 2004; Susser & Mulligan, 2015). Dual-basis accounts further distinguish two sources—experience-based fluency and theory-based beliefs—both of which can inform judgments (Koriat, 1997). Accordingly, a central question is whether the rhyme effect on JOLs is driven purely by fluency or whether learners' explicit beliefs about rhyme's mnemonic utility also contribute.

Recent evidence indicates that beliefs often play a prominent role in JOL construction across several well-studied effects. For font size, learners routinely give higher JOLs to larger fonts while recall changes little, a pattern frequently attributed to beliefs about size and memorability, with mixed support for an additional fluency component (Chang & Brainerd, 2022; Murphy & Castel, 2022). A recent review similarly concludes that metamemory beliefs account for many cue-JOL regularities (e.g., word frequency, concreteness, identity), with fewer cases in which beliefs and fluency jointly contribute, and even fewer in which fluency alone suffices (Yang et al., 2021). This pattern is compatible with *analytic-processing perspectives* that emphasize conceptually guided judgment formation (Mueller et al., 2016), while still allowing that fluency cues may contribute under specific conditions.

Learners may also hold explicit beliefs that rhymed materials are easier to remember, given their pervasive presence in songs, poetry, advertising, and instruction. Such exposure can foster the view that rhyme reduces processing demands and facilitates encoding, which may in turn elevate JOLs (e.g., Menninghaus et al., 2014; Mueller et al., 2013). Crucially, the mere existence of a belief does not guarantee its use in judgment. Beliefs influence JOLs when they are activated and applied to the items at hand (Tauber & Witherby, 2019). The present work therefore asks three empirical questions: (a) how common these rhyme-related beliefs are, (b) whether they are recruited when learners assign item-by-item JOLs, and (c) whether they change with task experience. Answering these questions clarifies the contribution of beliefs to the rhyme effect on JOLs and specifies when phonological structure shapes metacognitive monitoring through conceptually driven mechanisms rather than through momentary fluency cues alone.

Belief measurements

Prior work has typically assessed metamemory beliefs with pre-task prediction instruments that simulate a memory paradigm and elicit participants' a priori assumptions. Such questionnaires are administered before study to minimize contamination by task experience, and the resulting predictions track belief-based expectations about cue-memory relations (e.g., Kornell et al., 2011; Mueller et al., 2014). For example, Hu et al. (2015) asked participants to forecast recall for words in large versus small fonts prior to learning; these predictions correlated with the subsequent font-size effect on JOLs, indicating that pre-study beliefs contribute to judgment formation. This tradition motivates our inclusion of a pre-study belief measure tailored to rhyme.

Beliefs, however, are not static and can be reshaped by processing experience during the task. In a learner-observer design, Undorf et al. (2017) showed that direct task engagement led learners to assign higher JOLs when recognition was faster, whereas observers, who lacked that experience, treated longer exposure as more beneficial. These dissociations indicate that experience can reweight the cues that are treated as diagnostic, shifting judgments from naive heuristics toward signals derived from the unfolding task. To capture such updates, it is informative to assess beliefs again after study as an index of experience-shaped belief change.

Accordingly, the present study measures both pre-study and post-study beliefs about the mnemonic utility of rhyme. This dual assessment allows us to examine (a) how prevalent rhyme-related beliefs are before learning, (b) whether these beliefs align with or diverge from item-by-item JOLs, and (c) whether beliefs shift after exposure to the specific materials and procedures (Frank & Kuhlmann, 2017). Comparing pre-study and post-study beliefs provides a direct test of belief updating and supplies mediators for evaluating how conceptually driven expectations, alongside fluency, contribute to the rhyme effect on JOLs.

Overview of the current study

The current study conducted four experiments to investigate the effect of rhyme on JOLs and the role of metamemory beliefs in this effect. In Experiment 1, participants studied rhyming and non-rhyming word pairs and made item-by-item JOLs. The rhyme effect on JOLs was quantified as the difference in JOLs between rhyming and non-rhyming pairs. In Experiment 2, participants were instructed to read the word pairs aloud before making JOLs. Reading aloud was expected to increase the salience of the rhyme cue, thereby enhancing participants' explicit awareness of rhyme. This manipulation aimed to amplify the magnitude of the rhyme effect on both JOLs and memory, allowing for a clearer assessment of its impact. In Experiments 3 and 4, we explored the role of beliefs in the rhyme effect on JOLs. Experiment 3 utilized a standard questionnaire to investigate whether people hold pre-existing beliefs that rhyming word pairs are more memorable than non-rhyming ones. Experiment 4 assessed two types of beliefs: pre-existing beliefs (measured by a pre-study questionnaire) and post-study beliefs (measured by a post-study questionnaire). Subsequently, multilevel mediation analyses were conducted to determine whether these beliefs mediate the rhyme effect on JOLs.

Experiment 1

Experiment 1 was designed to explore whether JOLs are sensitive to rhyme. More specifically, Experiment 1 aimed to determine whether people can metacognitively appreciate the beneficial effect of rhyme on memory.

Method

Participants

Based on the result (Cohen's $d = 0.49$ for the rhyme effect on JOLs) of a pilot study (with 30 participants), a power analysis, conducted via G*power (Faul et al., 2007), showed that a total of 56 participants were required to observe a significant (2-tailed, $\alpha = 0.05$) rhyme effect on JOLs at 0.95 power. Finally, 64 participants (10 male; *M* of age = 20.08 years, *SD* = 1.65) were recruited from Beijing Normal University (BNU) participant pool. All of them were native Chinese speakers, reported normal or corrected-to-normal vision, were tested individually in a sound-proofed cubicle, and received monetary compensation. All experiments in the present study were approved by the Ethics Committee of BNU Faculty of Psychology.

Materials

To develop study stimuli (i.e., rhyming and non-rhyming word pairs), we first selected 60 two-character Chinese words as targets (e.g., 马路, for which the corresponding Pinyin is *Malu*, and its English translation is *road*). Next, for each target word, a rhyming cue (e.g., 跳舞, for which the corresponding Pinyin is *Tiaowu*, and its English translation is *dancing*) was selected to be paired with the target to form a rhyming pair (e.g., 跳舞 - 马路, for which the final characters of the cue and target shared the same compound vowel "u"). To make it easy for readers to understand, here is an English example of rhyming word pairs: *force - horse*. Besides the rhyming cue, a non-rhyming cue (e.g., 电话, for which the corresponding Pinyin is *Dianhua*, and its English translation is *phone*) was selected to be paired with the same target to form a non-rhyming word pair (e.g., 电话 - 马路, for which the final characters of the cue and target have different vowels). Take *pear - horse* as an English example of non-rhyming word pairs. In total, we developed 60 rhyming and 60 non-rhyming word pairs. The rhyming and non-rhyming word pairs shared the same target words.

Before the formal experiment, we recruited 70 college students (65 female; *M* of age = 19.14 years, *SD* = 1.36) to rate the concreteness level of each of cue and target words on a 9-point scale (1 = very abstract, 9 = very concrete). They also rated the level of rhyme (1 = not rhyming at all, 9 = very rhyming) of each word pair. Based on their ratings, 40 target words were selected. For these 40 target words, their paired rhyming and non-rhyming cues did not differ in the number of strokes ($p = .678$), word frequency ($p = .247$) and concreteness ratings ($p = .926$). The only difference between rhyming and non-rhyming pairs was that rhyme ratings of rhyming pairs ($M = 7.88$, $SD = 0.43$) were substantially greater than those of non-rhyming ones ($M = 1.64$, $SD = 0.27$), $t(78) = -77.63$, $p < .001$, $d = 17.38$. All rhyming and non-rhyming word pairs, along with their item-level characteristics, are provided in the *Stimuli* file on Open Science Framework (OSF: https://osf.io/5a9cz).

To prevent any item-selection effects, for each participant, the computer randomly chose 20 targets to be paired with their rhyming cues, with the other 20 targets paired with their non-rhyming cues. Therefore, each participant studied 20 rhyming and 20 non-rhyming pairs. Besides these 40 word pairs, another four (two rhyming and two non-rhyming) pairs were presented as practice trials, helping participants get familiar with the experimental procedure. In addition, another four (two rhyming and two non-rhyming) pairs were pre-

sented at the beginning and end of the study phase, serving as the primacy and recency buffer items. The data of practice, primacy, and recency trials were excluded from analyses.

Experimental design and procedure

The experiment involved a within-subjects design (material type: rhyming vs. non-rhyming). The experiment was programmed using *jsPsych* 6.1.0 (de Leeuw, 2015).

Participants were briefly informed of the procedure at the beginning of the experiment. A practice task consisting of four trials was implemented before the main experiment to ensure that participants fully understood the task requirements.

The main experiment consisted of three phases: a study phase, a distractor phase and a test phase. During the study phase, 44 word pairs (including two pairs as primary buffers and two as recency buffers) were presented one-by-one on the screen. Except for the primary and recency buffers, the remaining 40 word pairs were presented in a random order. Each word pair was displayed for 5 s, preceded by a cross sign shown for 0.5 s. After studying each word pair, participants made a JOL on a slider presented at the center of the screen. Specifically, participants were asked "*During the following test*, *if you are given the word on the left*, *what is the likelihood that you will recall the word on the right?*" Participants made JOLs by dragging and clicking a slider scale ranging from 0 (*Sure I will not recall it*) to 100 (*Sure I will recall it*). There was no time limitation for making item-by-item JOLs.

The distractor task was implemented after the study phase. During this task, a three-digit number (e.g., *200*) appeared at the center of the screen, and participants were asked to subtract 3 consecutively for 90 s, during which they reported their calculated answers aloud. This task served to prevent participants from rehearsing the studied word pairs.

The final test began after the distractor task. The 40 cue words (without primary and recency buffers) appeared one-by-one on the left side of the screen in a random order, with a blank box appearing on the right side of the screen. Participants were asked to input their answer into the blank box. After that, participants clicked on "Continue" button to submit their answer and to trigger the next test trial. There were no time pressure and no feedback during the test.

Results

All data have been made publicly available on the OSF and can be accessed at https://osf.io/5a9cz. All data were analyzed using JASP 0.17.3 (https://jasp-stats.org).

Test performance

Frequentist and Bayesian paired *t*-tests showed no significant difference in test performance between rhyming and non-rhyming pairs, difference = 0.006 [−0.03, 0.04], $t(63) = 0.34$, $p = .734$, $d = 0.04$, $BF_{10} = 0.15$ (see Fig. 1A), indicating no detectable influence of rhyme on recall performance.

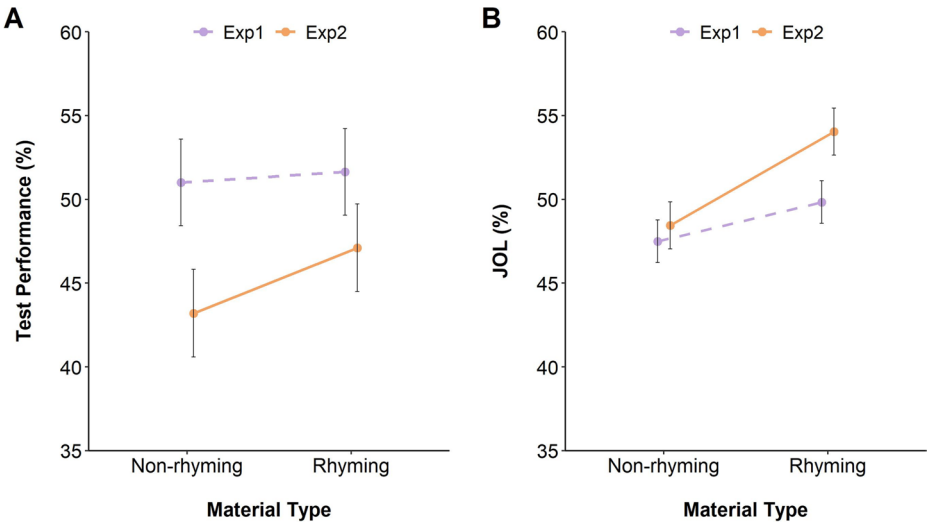

Fig. 1 **A**: Test performance as a function of material type and experiment. **B**: JOLs as a function of material type and experiment. Error bars represent 95% within-subjects confidence interval

JOLs

Mean JOLs for rhyming and non-rhyming pairs are depicted in Fig. 1B. Frequentist and Bayesian paired *t*-tests showed that rhyming pairs received significantly higher JOLs than non-rhyming ones, difference = 2.34 [0.54, 4.14], $t(63) = 2.60$, $p = .012$, $d = 0.33$, $BF_{10} = 3.01$, reflecting a rhyme effect on JOLs. Among these 64 participants, 41 showed a rhyme effect on JOLs, and the other 23 showed the reverse pattern.

Monitoring accuracy

To examine whether participants' monitoring accuracy differed between rhyming and non-rhyming pairs, we assessed both absolute accuracy and relative accuracy for each participant. Relative accuracy was first indexed by the Goodman-Kruskal gamma correlation between item-wise JOLs and recall (Nelson, 1984). Because gamma is undefined when a participant shows no variance in JOLs or outcomes, we excluded 4 participants for this reason and excluded one additional participant due to missing item-level data. The final data analysis included 59 participants. Frequentist and Bayesian paired *t*-tests showed that gamma for rhyming pairs ($M = 0.13$, $SD = 0.43$) was marginally lower than that for non-rhyming ones ($M = 0.26$, $SD = 0.27$), difference = −0.12 [-0.25, 4.69×10^{-4}], $t(58) = -1.99$, $p = .051$, $d = -0.26$, $BF_{10} = 0.90$.

As a robustness check, we also computed the area under the Type-2 ROC curve (AUROC2), which ranges from 0 (chance with reversed ordering) to 1 (perfect discrimination), with 0.5 indicating chance-level discrimination and provides a bias-resistant, non-parametric measure of metacognitive sensitivity (Fleming & Lau, 2014). Results converged with the gamma analysis: AUROC2 for rhyming pairs ($M = 0.56$, $SD = 0.21$) was lower than for non-rhyming pairs ($M = 0.63$, $SD = 0.13$), difference = −0.06 [−0.12, −0.002], $t(58) =$

−2.06, p=.044, d = −0.27, BF_{10}=1.02. Taken together, these results showed that rhyming cues may have slightly impaired participants' ability to distinguish between remembered and forgotten items, leading to reduced relative accuracy compared to non-rhyming pairs.

For the absolute accuracy scores, Frequentist and Bayesian paired t-tests showed no significant difference between rhyming pairs (M = −1.81, SD=27.58) and non-rhyming pairs (M=−3.53, SD=27.03), difference=1.72 [−2.24, 5.68], $t(63)$=0.87, p=.390, d=0.11, BF_{10}=0.20, indicating that participants exhibited similar levels of calibration across both conditions.

Discussion

Experiment 1 clearly demonstrated a rhyme effect on JOLs. However, it also demonstrated that the rhyme effect on JOLs tends to be a metamemory illusion. That is, rhyme has dissociative effects on JOLs and memory (Kleider & Goldinger, 2004). Although we predicted that rhyme should facilitate recall performance, Experiment 1 showed no statistically detectable difference in recall performance between rhyming and non-rhyming pairs. Below, we provide detailed discussion about why Experiment 1 showed no effect of rhyme on recall performance. Furthermore, analyses of relative accuracy revealed that participants exhibited marginally lower gamma correlations for rhyming pairs compared to non-rhyming pairs, suggesting that rhyming cues may have impaired their ability to accurately monitor their memory performance. A plausible account is that a salient phonological cue elevates judgments broadly while offering limited item-specific diagnostic, which compresses JOL variability and hampers differentiation between subsequently remembered and forgotten items.

Experiment 2

The results of Experiment 1 preliminarily demonstrated a rhyme effect on JOLs. However, the observed effect size (d=0.33) was relatively small. To further investigate this effect, we made slight modifications to the experimental procedure in Experiment 2. Given that rhyme is inherently a phonetic feature, we asked participants to read both rhyming and non-rhyming word pairs aloud during the study phase. This manipulation was expected to enhance participants' explicit awareness of the rhyme cue, thereby amplifying its effect on JOLs. This expectation aligns with previous findings suggesting that explicit awareness plays a key role in shaping the mnemonic effectiveness of rhyme (Hoorn, 1996; Nelson et al., 1992).

Additionally, the lack of a rhyme effect on recall performance in Experiment 1 may have resulted from participants not reading the word pairs aloud during learning. In other words, rhyme may exert little influence on memory retention unless learners explicitly engage with its phonological properties through reading aloud. To test this hypothesis, Experiment 2 required participants to read all word pairs aloud during the study phase. We expected that this manipulation would not only reinforce the salience of the rhyme cue but also produce a detectable rhyme effect on recall performance.

Method

Participants

Following Experiment 1, 64 participants (four males; *M* of age = 19.78 years, *SD* = 1.44) were recruited from BNU participant pool. They were native Chinese speakers, reported normal or corrected-to-normal vision, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design and procedure

The materials, experimental design, and procedure were the same as those in Experiment 1, except that participants were asked to read each word pair aloud during the study phase.

Results

All data have been made publicly available on the OSF and can be accessed at https://osf.io/5a9cz.

Test performance

Frequentist and Bayesian paired *t*-tests showed that participants remembered more rhyming pairs than non-rhyming ones, difference = 0.04 [0.002, 0.076], $t(63) = 2.11$, $p = .039$, $d = 0.26$, $BF_{10} = 1.09$ (see Fig. 1A), reflecting a modest rhyme effect on memory. Among the 64 participants, 33 showed superior recall of rhyming pairs, with 25 demonstrating a reverse pattern, and there were 6 ties.

JOLs

Frequentist and Bayesian paired *t*-tests showed that rhyming pairs received significantly higher JOLs than non-rhyming ones, difference = 5.60 [3.62, 7.58], $t(63) = 5.65$, $p < .001$, $d = 0.71$, $BF_{10} > 1{,}000$ (see Fig. 1B), reflecting a substantial rhyme effect on JOLs. Among the 64 participants, 51 showed a rhyme effect on JOLs, with only 12 demonstrating a reverse pattern, and there was one tie.

Moderating effect of reading aloud on the rhyme effect on JOLs

Experiments 1 and 2 showed that the rhyme effect on JOLs existed regardless of whether participants read aloud the word pairs or not. However, the effect size tended to be larger in Experiment 2 ($d = 0.71$) than in Experiment 1 ($d = 0.33$). Further analyses were conducted to compare the results of these two experiments. Frequentist and Bayesian mixed-measures ANOVAs were employed to explore whether reading aloud experience can enlarge the rhyme effect on JOLs. Experiment (1 vs. 2) was taken as a between-subjects variable, with material type (rhyming vs. non-rhyming) as a within-subjects variable, and JOLs as the dependent variable. The results showed a main effect of material type, $F(1, 126) = 35.15$, $p < .001$, $\eta_p^2 = 0.22$, $BF_{10} > 1{,}000$. As shown in Fig. 1B, JOLs for rhyming pairs were higher than those

for non-rhyming pairs, reconfirming the rhyme effect on JOLs. The effect of experiment was not statistically significant, $F(1, 126) = 1.07$, $p = .304$, $\eta_p^2 = 0.008$, $BF_{10} = 0.53$.

Of critical importance, there was a significant interaction between experiment and material type, $F(1, 126) = 5.92$, $p = .016$, $\eta_p^2 = 0.045$, $BF_{10} = 2.63$. As shown in Fig. 1B, the main driver of this interaction is that the rhyme effect on JOLs (represented as the difference in JOLs between rhyming and non-rhyming pairs) was larger (almost doubled) in Experiment 2 ($M = 5.60$, $SD = 7.94$) than in Experiment 1 ($M = 2.34$, $SD = 7.20$), difference = 3.26 [0.61, 5.91], $t(126) = 2.43$, $p = .016$, $d = 0.43$, $BF_{10} = 2.68$. Overall, these results reflect that reading aloud experience can exaggerate the rhyme effect on JOLs.

Similar Frequentist and Bayesian mixed ANOVAs were performed, with recall accuracy as the dependent variable (Fig. 1A). The results showed no main effect of experiment, $F(1, 126) = 2.30$, $p = .130$, $\eta_p^2 = 0.018$, $BF_{10} = 0.89$, no main effect of material type, $F(1, 126) = 3.04$, $p = .084$, $\eta_p^2 = 0.02$, $BF_{10} = 0.56$, and no detectable interaction, $F(1, 126) = 1.59$, $p = .209$, $\eta_p^2 = 0.01$, $BF_{10} = 0.44$.

Monitoring accuracy

For relative-accuracy analyses, one participant was excluded due to no variance in JOLs or outcomes. Frequentist and Bayesian paired *t*-tests showed that gamma for rhyming pairs ($M = 0.11$, $SD = 0.32$) was marginally lower than that for non-rhyming ones ($M = 0.22$, $SD = 0.36$), difference = −0.11 [−0.23, 0.01], $t(62) = -1.78$, $p = .080$, $d = -0.22$, $BF_{10} = 0.61$. AUROC2 analysis yielded the same directional pattern, with rhyming pairs ($M = 0.55$, $SD = 0.15$) lower than that for non-rhyming ones ($M = 0.61$, $SD = 0.17$), difference = −0.06 [−0.12, 0.005], $t(62) = -1.83$, $p = .072$, $d = -0.23$, $BF_{10} = 0.66$. These results indicate a small reduction in item-level discrimination for rhymed items, consistent with uniformly elevated JOLs not matched by item-specific diagnostic.

For the absolute accuracy scores, Frequentist and Bayesian paired *t*-tests showed no significant difference between rhyming pairs ($M = 6.93$, $SD = 23.06$) and non-rhyming pairs ($M = 5.24$, $SD = 21.65$), difference = 1.70 [−2.61, 6.00], $t(63) = 0.79$, $p = .434$, $d = 0.10$, $BF_{10} = 0.18$, indicating that participants exhibited similar levels of calibration across both conditions.

Discussion

The results of Experiment 2 verified that rhyme substantially affects JOLs when learners read the study materials aloud. Experiments 1 and 2 jointly showed that reading aloud exaggerates the influence of rhyme on JOLs. A possible explanation is that reading aloud experience enhances participants' realization of the difference in the levels of rhyme among word pairs, which then activates beliefs about the effect of rhyme on memory to guide JOL formation. Consistent with Experiment 1, analyses of relative accuracy revealed that participants continued to exhibit lower gamma correlations for rhyming pairs compared to non-rhyming pairs, suggesting that monitoring accuracy remained compromised when learning rhyming items.

Experiment 3

Results of Experiments 1 and 2 jointly showed that JOLs are sensitive to the rhyme cue regardless of whether participants read the word pairs aloud or not. We hypothesize that people may hold a priori beliefs that rhyme acts as a catalyst for boosting memory, which is a possible mechanism underlying the rhyme effect on JOLs. Experiment 3 employed a standard questionnaire-based approach for assessing metacognitive beliefs, a method frequently used in prior research (e.g., Mueller et al., 2016). This approach systematically simulates a memory experiment to elicit participants' beliefs about the mnemonic benefits of rhyme.

Method

Participants

Following Experiments 1 and 2, 64 students (18 male; *M* of age = 21.39 years, *SD* = 4.12) were recruited from BNU participant pool. They were native Chinese speakers, reported normal or corrected-to-normal vision, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design and procedure

This experiment employed a within-subjects design (material type: rhyming vs. non-rhyming). The Qualtrics Online Questionnaire System was used to implement the questionnaire. After signing the informed consent, participants read the following instructions:

Imagine a memory experiment, *in which a student is presented with a list of 40 word pairs*, *in which the word pairs are studied one-by-one. Each word pair includes a cue word (presented at the left side) and a target word (presented at the right side). Among the word pairs*, *half of them (20 pairs) do not rhyme with each other*, *and the other half (20 pairs) do. The cue and target words in a non-rhyming pair have different final vowels*, such as 酒瓶 (jiuping) - 滑翔 (huaxiang), *but those in a rhyming word pair have the same final vowel*, such as 灯泡 (dengpao) - 碉堡 (diaobao). Each word pair is presented for five seconds, and the task for this student is to endeavor to memorize as many word pairs as he or she could. After learning these 40 word pairs, *he or she needs to spend 90s to answer some mathematics questions. Next*, *the student attends to a memory test*, in which he or she is required to recall the target (right) word when prompted with the cue (left) word (e.g., 酒瓶 - ___?, 灯泡 - ___?).

1. *How many of the 20 rhyming pairs do you think the student will be able to correctly recall in the final memory test?*
2. *How many of the 20 non-rhyming pairs do you think the student will be able to correctly recall in the final memory test?*

The order of Questions 1 and 2 was counterbalanced across participants. At the end of the questionnaire, participants were asked to provide specific explanations for their predictions. The explanation question was framed according to their predictions ("*Why do you think that the student will remember more rhyming word pairs than non-rhyming ones?*", "*Why do*

you think that he or she will remember more non-rhyming word pairs than rhyming ones?", or "*Why do you think that he or she will remember the same number of rhyming and non-rhyming word pairs?*").

Results

All data have been made publicly available on the OSF and can be accessed at https://osf.io/5a9cz.

Participants predicted that the student would remember more rhyming pairs (M=0.58, SD=0.18) than non-rhyming ones (M=0.45, SD=0.21), difference=0.13 [0.08, 0.18], $t(63)$=4.91, p<.001, d=0.61, BF_{10}>1,000, reflecting that people do hold a priori beliefs that rhyming word pairs are easier to remember than non-rhyming ones. Among the 64 participants, 93.75% (60 out of 64) of them predicted that the student would remember more rhyming pairs, with only 4.69% (3) giving predictions in the opposite direction, and the other participant's predictions (1.56%) were tied.

The reasons, provided by the 60 participants who predicted superior recall of rhyming pairs, could be divided into 5 categories (see Table 1). Over half of the participants (53.3%) thought that rhyme affects memory by facilitating the encoding process, such as through promoting associative memory, providing same vowels as additional mnemonic cues, and heightening attention. 30.0% held the perspective that rhyme improves memory by enhancing reading fluency, saying that rhyming pairs are easier to read in contrast to non-rhyming ones. 10.0% responded with circulative reasons, for example, simply stating "rhyming word pairs are easier to remember". Other reasons, such as analogical explanations, are provided in Table 1 as well.

Interestingly, three participants, who predicted better recall of non-rhyming pairs, regarded rhyme as a distractor, which may divide their attention from processing of semantic relations between the cue and target. The other participant, who predicted no difference

Table 1 Reasons for beliefs about why rhyme enhances memory

Type		Sample Responses	Percentage
Facilitating encoding	Improving associative memory	"More repetitions of the vowels are helpful for associative memory"	21.67%
	Additional cues provided by rhyme	"Same vowels can be a hint"	31.67%
	Attention enhancement	"Rhyming word pairs are catchy, thus easier to remember"	30.00%
Circulation		"Rhyming word pairs are easier to remember"	10.00%
Analogy		"Just as lyrics and poems are certainly easier to remember than proses, and rhyme can strengthen phonetic memory"	1.66%
Others		"There is an internal logic between rhyming pairs, and the formed neurons are more closely connected"	5.00%

between the two types of word pairs, explained that memory performance is dependent on how many times a given item is repeatedly studied, but not other factors (e.g., rhyme).

Discussion

The above results suggest that most people hold a priori beliefs that rhyme is an important characteristic of study materials that can facilitate memory, and there are a few common explanations for such beliefs.

Experiment 4

Experiment 3 provided primary findings regarding people's pre-existing beliefs about the effect of rhyme on memory. Experiment 4 was conducted to further explore whether beliefs play an important role in the rhyme effect on JOLs. Specifically, Experiment 4 measured two types of beliefs, including pre-existing beliefs (measured by a questionnaire implemented before the study task) and post-study beliefs (measured by a questionnaire implemented after the study task), to explore the mediating role of beliefs in the rhyme effect on JOLs.

Method

Participants

Based on the results of Experiment 2 ($d = 0.71$ for the rhyme effect on JOLs), a power analysis showed that 28 participants were required to observe a significant (2-tailed, $\alpha = 0.05$) rhyme effect on JOLs at 0.95 power. Because Experiment 4 also aimed to explore the potential mediating role of beliefs in the rhyme effect on JOLs, we hence decided to conservatively increase the sample to 80. Accordingly, 81 participants from BNU participant pool (6 male; *M* of age = 20.80 years, *SD* = 2.03) were recruited. They were native Chinese speakers, reported normal or corrected-to-normal vision, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design and procedure

The materials, experimental design and procedure were similar to those in Experiment 2, but with several exceptions. Before the formal experiment, participants needed to complete a questionnaire to measure their pre-existing beliefs about the effect of rhyme on memory. They were asked to imagine that they were going to study 20 rhyming and 20 non-rhyming pairs in a learning task, and needed to predict what proportions of rhyming and non-rhyming pairs they would remember in a later memory test. The order for measuring pre-existing beliefs for rhyming and non-rhyming pairs were counterbalanced across participants.

The procedure of the study task was identical to those in Experiment 2 but with one difference. The change was that, after studying all pairs, participants were asked to predict what proportions of rhyming and non-rhyming word pairs they would remember in the later test. The order of questions for measuring post-study beliefs for rhyming and non-rhyming

pairs were counterbalanced across participants. After answering the second questionnaire, they completed in the same distractor task and final test as those in Experiments 1 and 2.

Results

All data have been made publicly available on the OSF and can be accessed at https://osf.io/5a9cz.

Test performance

Frequentist and Bayesian paired *t*-tests showed a significant difference in test performance between rhyming and non-rhyming pairs, difference = 0.06 [0.02, 0.09], $t(80) = 3.46$, $p < .001$, $d = 0.39$, $BF_{10} = 27.40$, replicating the rhyme effect on memory observed in Experiment 2 (see Fig. 2A).

JOLs and beliefs

The results of pre-existing beliefs, JOLs, and post-study beliefs are depicted in Fig. 2B C. First, Frequentist and Bayesian paired *t*-tests showed that there was a significant difference in JOLs between rhyming and non-rhyming pairs, difference = 7.03 [5.38, 8.68], $t(80) = 8.47$, $p < .001$, $d = 0.94$, $BF_{10} > 1{,}000$, replicating the rhyme effect on JOLs observed in Experiments 1 and 2 (see Fig. 2B).

Second, the results showed that participants held pre-existing beliefs that rhyming pairs were easier to remember than non-rhyming pairs, difference = 15.12 [12.39, 17.85], $t(80) = 11.03$, $p < .001$, $d = 1.23$, $BF_{10} > 1{,}000$ (see Fig. 2C), replicating the main findings of Experiment 3.

Finally, participants held post-study beliefs that rhyming pairs were easier to remember than non-rhyming pairs, difference = 3.47 [0.10, 6.84], $t(80) = 2.05$, $p = .044$, $d = 0.23$, $BF_{10} = 0.89$ (see Fig. 2C). Such a finding reflects that, after the study phase, participants still believed that rhyming pairs were easier to remember than non-rhyming ones.

Monitoring accuracy

For relative-accuracy analyses, four participants were excluded due to no variance in JOLs or outcomes. Frequentist and Bayesian paired *t*-tests showed that gamma for rhyming pairs ($M = 0.15$, $SD = 0.35$) was significantly lower than that for non-rhyming ones ($M = 0.28$, $SD = 0.37$), difference = −0.12 [−0.22, −0.03], $t(76) = -2.58$, $p = .012$, $d = -0.29$, $BF_{10} = 2.75$. AUROC2 analysis yielded the same directional pattern, with rhyming pairs ($M = 0.57$, $SD = 0.17$) significantly lower than that for non-rhyming ones ($M = 0.63$, $SD = 0.18$), difference = −0.06 [−0.11, −0.01], $t(76) = -2.60$, $p = .011$, $d = -0.30$, $BF_{10} = 2.89$. These results showed that with the increased sample size in Experiment 4, the difference in relative accuracy reached statistical significance, confirming that rhyming cues indeed disrupted participants' ability to accurately monitor their memory performance.

Given that Experiments 1 and 2 produced marginal effects and Experiment 4 yielded a significant decrement, we conducted a random-effects meta-analysis using the R *metafor* package to estimate the overall effect of rhyme on relative accuracy (gamma) across studies.

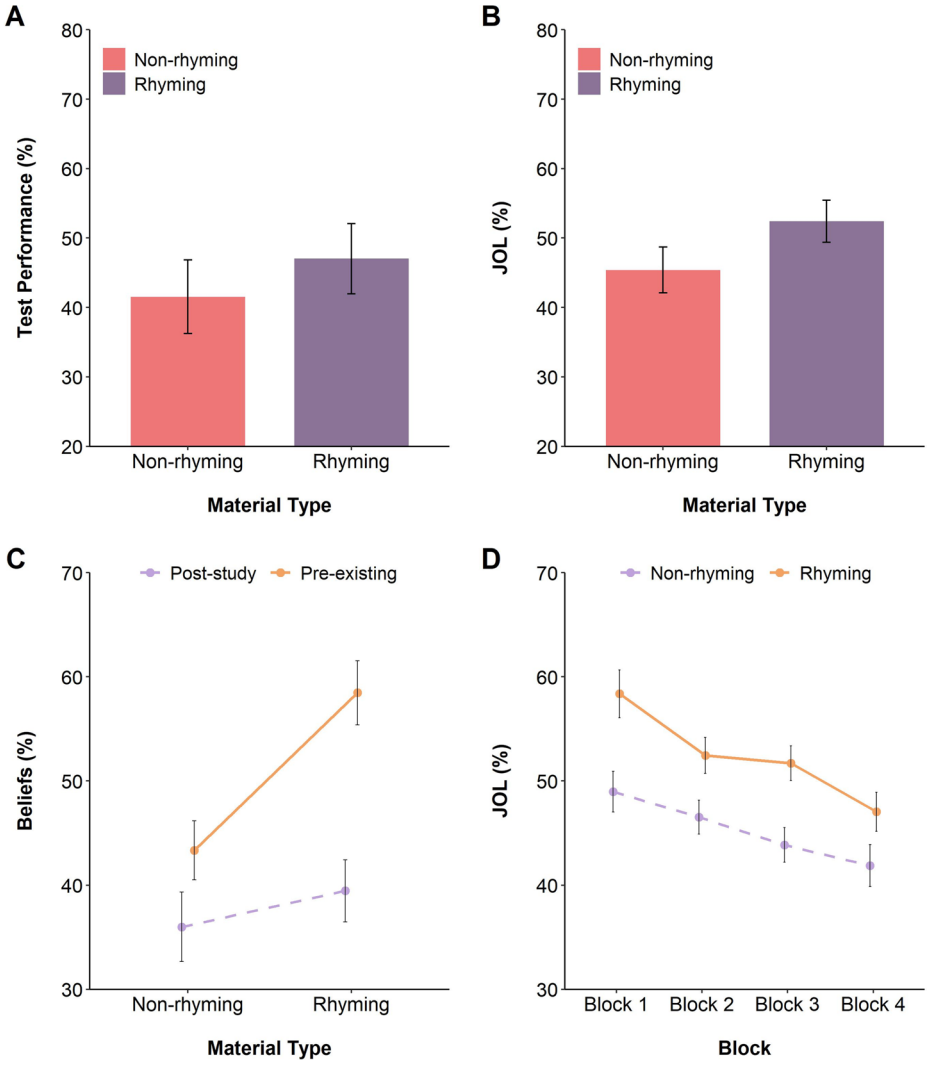

Fig. 2 **A**: Test performance (i.e., recall accuracy) as a function of material type in Experiment 4. **B**: JOLs as a function of material type in Experiment 4. **C**: Beliefs as a function of material type and belief type. **D**: JOLs for both rhyming and non-rhyming pairs in each of Blocks 1 and 4. Error bars represent 95% within-subjects confidence interval

For each experiment, we computed the paired-sample Cohen's *d* for the difference between rhyming and non-rhyming conditions and converted it to Hedges' *g* to correct for small-sample bias. As shown in Fig. 3, the random-effects model (REML) yielded a significant negative overall effect, Hedges' $g = -0.26$, 95% CI = [−0.45, −0.08], $p = .004$. This result indicates that, on average, rhyme reduces relative accuracy across experiments. A test of heterogeneity was not significant, $Q(2) = 0.10$, $p = .953$, $I^2 < 0.1\%$, suggesting little between-study variability in the magnitude of this effect. As a robustness check, we also performed a parallel random-effects meta-analysis of AUROC2. The pooled effect was Hedges' *g* =

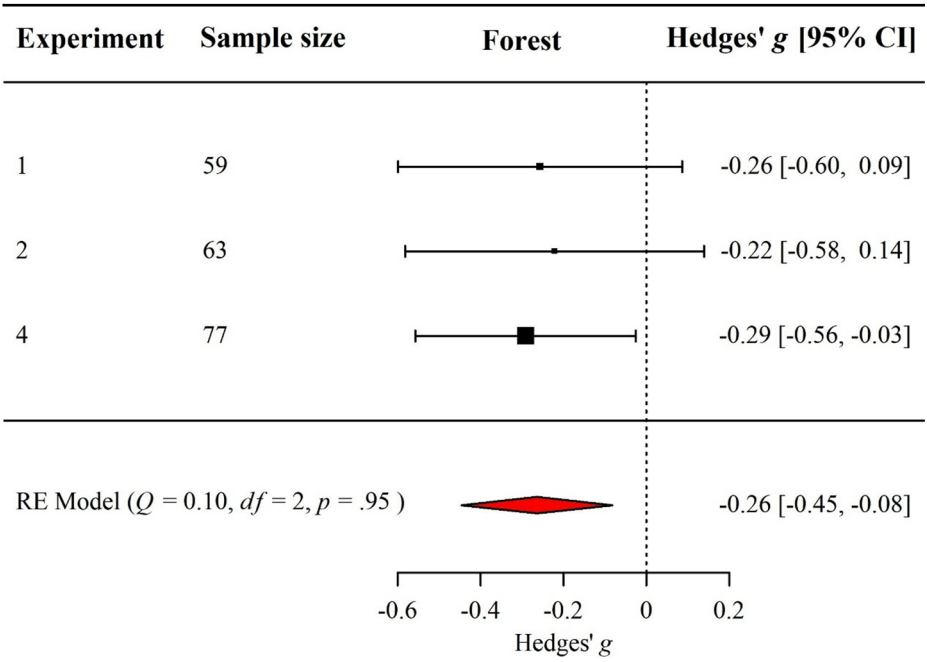

Fig. 3 Forest plot of the random-effects meta-analysis on gamma-based relative accuracy (rhyming minus non-rhyming) across Experiments 1, 2, and 4

−0.27, 95% CI = [−0.45, −0.09], p=.003, again indicating lower discrimination for rhymed items. Heterogeneity was $Q(2)=0.09$, p=.958, $I^2 < 0.1\%$.

For the absolute accuracy scores, Frequentist and Bayesian paired t-tests still showed no significant difference between rhyming pairs (M=5.39, SD=22.07) and non-rhyming pairs (M=3.85, SD=22.97), difference=1.53 [−1.72, 4.79], $t(80)=0.94$, p=.351, d=0.10, $BF_{10}=0.19$, indicating that participants exhibited similar levels of calibration across both conditions.

Change of beliefs across the learning task

If beliefs are recruited during judgment, they should also be sensitive to task feedback and experience. As shown above, pre-existing beliefs (d=1.23) about the rhyme effect on memory tended to be stronger than post-study beliefs (d=0.23), suggesting that processing experience obtained from the study task might change participants' beliefs about the effect of rhyme on memory. Frequentist and Bayesian repeated-measures ANOVAs were performed, with belief type (pre-existing beliefs vs. post-study beliefs) and material type (rhyming vs. non-rhyming) as the independent variables and beliefs as the dependent variable. The results showed a main effect of material type, $F(1, 80)=57.25$, p<.001, $\eta_p^2 = 0.42$, $BF_{10}>1{,}000$, re-confirming that participants believed that rhyming pairs are easier to remember than non-rhyming ones. There was also a main effect of belief type, $F(1, 80) = 36.82$, p <.001, $\eta_p^2 = 0.32$, $BF_{10}>1{,}000$. As shown in Fig. 2C, participants' beliefs about their overall memory performance were lower after than before the study task. Such a decline in JOLs might

result from the fact that, across the study task, participants gradually realized the difficulty of the study task or they gradually became fatigued across the task.

Of critical importance, there was a significant interaction between material type and belief type, $F(1, 80) = 39.39$, $p < .001$, $\eta_p^2 = 0.33$, $BF_{10} > 1{,}000$. As shown in Fig. 2C, the main driver of this interaction is that pre-existing beliefs about the rhyme effect on memory was stronger ($M = 15.12$, $SD = 12.34$) than post-study beliefs ($M = 3.47$, $SD = 15.23$), difference = 11.65 [7.96, 15.35], $t(80) = 6.28$, $p < .001$, $d = 0.70$, $BF_{10} > 1{,}000$. This interaction suggests that the experience obtained from the study task reduced participants' beliefs about the rhyme effect on memory.

Furthermore, we divided the 40 word pairs into 4 blocks, with each block containing 10 word pairs. Then, we calculated the average JOLs for both rhyming and non-rhyming pairs in each block. Frequentist and Bayesian repeated-measures ANOVAs were performed, with block (block 1 vs. block 4) and material type (rhyming vs. non-rhyming) as the independent variables and JOLs as the dependent variable. As shown in Fig. 2D, the results showed a main effect of material type, $F(1, 80) = 52.89$, $p < .001$, $\eta_p^2 = 0.40$, $BF_{10} > 1{,}000$, re-confirming the rhyme effect on JOLs. There was also a main effect of block, $F(1, 80) = 47.43$, $p < .001$, $\eta_p^2 = 0.37$, $BF_{10} > 1{,}000$, reflecting that participants' JOLs gradually declined throughout the study task. The decline of JOLs across the learning task was consistent with the decline of JOLs (i.e., the difference in memory predictions measured before and after the study task) reported above.

Of critical importance, there was a significant interaction between material type and block, $F(1, 80) = 6.87$, $p = .010$, $\eta_p^2 = 0.08$, $BF_{10} = 4.04$. As shown in Fig. 2D, the main driver of this interaction is that the difference in JOLs between rhyming and non-rhyming pairs declined across blocks, corresponding to the decline of beliefs about the rhyme effect on memory from the beginning to the end of the study task. These results clearly support the role of beliefs in the rhyme effect on JOLs.

Mediation analyses

All mediation analyses presented below were performed via MLmed macro in SPSS (Level 1: items; Level 2: participants; Hayes & Rockwood, 2020) which was the standard 1-1-1 design. The results from the multilevel mediation model are shown in Fig. 4.

First, we conducted a mediation analysis to estimate the indirect effect of material type on JOLs through pre-existing beliefs. As shown in Fig. 4A, the indirect effect was significant, $a*b = 3.56$, $SE = 0.69$, $Z = 5.14$, $p < .001$, 95% CI [2.20, 4.94], which means that rhyme increased JOLs indirectly through pre-existing beliefs about the rhyme effect on memory. The direct effect of rhyme on JOLs was also significant, $c' = 3.46$, $SE = 0.89$, $t(3157) = 3.89$, $p < .001$, 95% CI [1.72, 5.21], suggesting that pre-existing beliefs tend to partially (but not fully) mediate the rhyme effect on JOLs.

The second mediation analysis was performed to explore the mediating role of post-study beliefs in the rhyme effect on JOLs. As shown in Fig. 4B, the indirect effect of post-study beliefs was significant, $a*b = 1.33$, $SE = 0.34$, $Z = 3.91$, $p < .001$, 95% CI [0.67, 2.01], and the direct effect was also statistically significant, $c' = 5.69$, $SE = 0.66$, $t(3157) = 8.68$, $p < .001$, 95% CI [4.41, 6.98]. Therefore, these results reflect that, after obtaining learning experience, beliefs about the rhyme effect on memory still partially mediated the rhyme effect on JOLs.

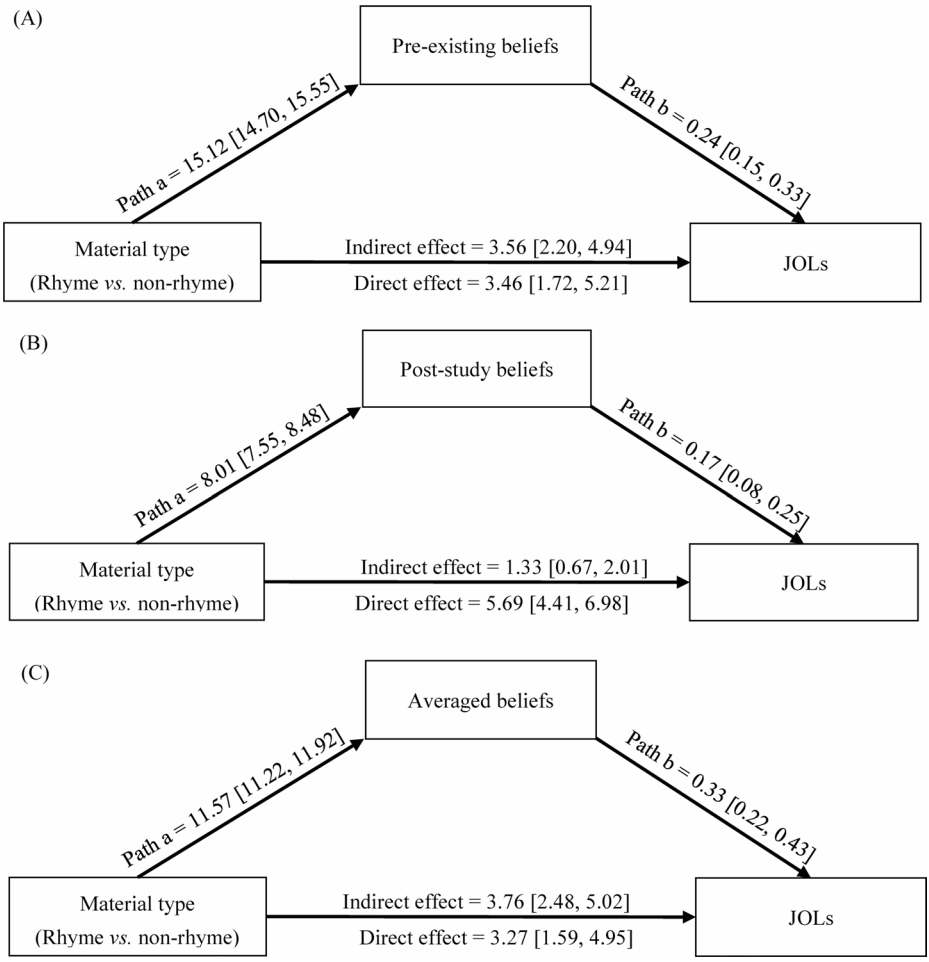

Fig. 4 Results from the multilevel mediation analyses. **A**: Pre-existing beliefs as the mediator between material type and JOLs. **B**: Post-study beliefs as the mediator between material type and JOLs. **C**: Averaged beliefs (i.e., average of pre-existing beliefs and post-study beliefs) as the mediator between material type and JOLs

Considering that beliefs about the rhyme effect on memory changed across the study task, we averaged pre-existing beliefs and post-study beliefs to form averaged beliefs. Such belief scores should better represent beliefs participants used to form JOLs during the study task. As showed in Fig. 4C, the indirect effect of averaged beliefs was significant, $a*b=3.76$, $SE=0.65$, $Z=5.77$, $p<.001$, 95% CI [2.48, 5.02], and the direct effect was also statistically significant, $c' = 3.27$, $SE=0.86$, $t(3157)=3.81$, $p<.001$, 95% CI [1.59, 4.95].

Discussion

Experiment 4 demonstrated that the rhyme effect on JOLs is stable and replicable, yet it also revealed that rhyming cues impair metacognitive monitoring, as evidenced by sig-

nificantly lower relative accuracy for rhyming pairs compared to non-rhyming pairs. More importantly, participants' beliefs about the effect of rhyme on memory changed as the task went on. At the same time, pre-existing, post-study and averaged beliefs could successfully mediate the rhyme effect on JOLs, consistent with analytic-processing theory (Koriat, 1997; Mueller et al., 2016). Furthermore, learning experience reduced participants' beliefs about the effect of rhyme on memory. This reduction seems to be beneficial, as recall performance results showed that participants' memory for rhyming pairs only increased by 5.5% [0.02, 0.09] compared to non-rhyming pairs. In contrast, their pre-existing beliefs predicted a 15.1% [0.12, 0.18] improvement in memory, while post-study beliefs predicted a 3.5% [0.001, 0.06] enhancement. Hence, post-study beliefs, by comparison with pre-existing beliefs, are more in line with the actual effect of rhyme on memory.

General discussion

Metacognitive judgments, such as JOLs, do not fully reflect memory performance but are instead influenced by various cues and heuristics (Dunlosky & Metcalfe, 2009). Understanding how different cues affect both JOLs and actual memory performance is crucial for improving self-regulated learning. Across four experiments, we examined the influence of rhyme cues on metacognition and memory, extending cue-utilization accounts to phonological features (Koriat, 1997). Our results revealed a rhyme effect on JOLs, where participants consistently judged rhyming pairs as more memorable than non-rhyming pairs, despite rhyme not always enhancing recall. This pattern aligns with the analytic-processing theory (Mueller et al., 2016), as our mediation analyses showed that pre-existing, post-study, and averaged beliefs about the mnemonic benefits of rhyme made reliable contributions to JOL formation. However, the actual memory benefits of rhyme were inconsistent, emerging only when participants read the word pairs aloud, suggesting that rhyme alone may not reliably enhance memory unless combined with additional encoding strategies. Furthermore, we found that rhyme cues impaired monitoring accuracy, with convergent evidence from both gamma and AUROC2, indicating that while learners relied on rhyme as a basis for their JOLs, they had more difficulty accurately distinguishing between well-learned and poorly-learned rhyming items. This finding suggests a potential misalignment between subjective memory predictions and actual recall performance, reinforcing the idea that salient but weakly diagnostic surface-level cues can sometimes mislead metacognitive monitoring (Hertzog et al., 2013). The following sections will further discuss these findings, their theoretical implications, and their relevance to learning and memory research.

Rhyme effect on JOLs

The present study provides robust evidence for the rhyme effect on JOLs. Specifically, Experiments 1, 2, and 4 consistently demonstrated that participants assigned significantly higher JOLs to rhyming word pairs compared to non-rhyming ones. Across studies, the proportions of participants showing higher JOLs for rhymes were 64.06%, 79.69%, and 82.72%, further confirming the reliability of this effect. However, in Experiment 1, the observed effect size was relatively small ($d = 0.33$), suggesting that the influence of rhyme on JOLs was modest under standard learning conditions. By contrast, instructing participants to read pairs aloud

in Experiments 2 and 4 markedly increased the magnitude of the effect ($ds \geq 0.71$), indicating that task conditions that heighten phonological engagement strengthen the impact of rhyme on judgments. As shown in Fig. 1B, the aloud-reading manipulation amplified the mean difference in JOLs between rhymed and non-rhymed items. A plausible account is that aloud reading increased attention to phonological regularities and enhanced the subjective ease of processing, which in turn elevated JOLs. These results indicate that, in addition to beliefs, experience-based fluency can magnify the rhyme effect on JOLs under conditions that emphasize phonology.

To further investigate the mechanisms underlying this effect, Experiments 3 and 4 examined participants' beliefs about the mnemonic benefits of rhyme. The results showed that participants generally perceived rhyming word pairs as easier to remember than non-rhyming ones. Crucially, multilevel mediation in Experiment 4 showed that both pre-study and post-study beliefs significantly mediated the relationship between rhyme and JOLs, indicating that beliefs are a key factor shaping JOL formation. One possible explanation is that individuals acquire and apply a belief that rhyme enhances memory, likely influenced by cultural and linguistic exposure to rhyming content in poetry, music, advertising slogans, and traditional sayings (Menninghaus et al., 2014). Repeated exposure may normalize the assumption that rhyme facilitates learning and retention, leading learners to import this belief into item-level JOLs.

While our findings support the central role of beliefs in JOL formation, they do not entirely rule out the contribution of processing fluency. First, the partial mediation indicates that additional mechanisms, including fluency, also contribute. Second, the between-experiment contrast (silent study vs. aloud reading) suggests that enhancing phonological engagement increases the subjective fluency difference between rhymed and non-rhymed items, thereby increasing JOLs. This pattern aligns with the dual-basis model (Koriat, 1997), which posits that both beliefs and fluency contribute to JOL construction. At the same time, the consistent mediation by beliefs accords with analytic-processing theory (Mueller et al., 2016) that assign beliefs a primary role, with fluency acting as a secondary input contingent on task demands. Future work might manipulate phonological engagement and belief salience orthogonally to test how these sources interact in shaping JOLs and to identify conditions under which fluency-based processing exerts a larger influence.

Rhyme effect on memory

Consistent with previous research, Experiments 2 and 4 indicated that rhyme can function as a mnemonic cue that facilitates learning and memory under conditions that heighten phonological engagement (Király et al., 2017; Plastikwala, 2017). However, Experiment 1 showed no reliable mnemonic advantage when participants were not allowed to read the word pairs aloud. One likely source of this discrepancy is the stimulus format: the present studies used paired associates, whereas much of the prior work employed poems or rhymed lists (e.g., Gupta et al., 2005; Király et al., 2017). Without vocalization, participants in Experiment 1 may have failed to detect the phonological overlap between cues and targets, reducing the salience of rhyme and limiting its contribution at encoding. By contrast, aloud reading in Experiments 2 and 4 likely increased attention to rhythm and prosody and made the phonological regularity more explicit, thereby strengthening the cue's effectiveness and improving recall. This interpretation is consistent with accounts in which task demands and

encoding orientation determine whether phonological structure is utilized during learning. At the same time, the present design did not orthogonally manipulate reading mode, so we cannot make a definitive causal claim that vocalization per se produced the memory benefit. Future work should systematically vary overt articulation (e.g., silent vs. mouthed vs. spoken study) and measure perceived rhyme at the item level to establish the mechanisms by which phonological engagement amplifies rhyme's mnemonic impact.

The alignment between monitoring and memory

In Experiments 1, 2, and 4, the rhyme effect on JOLs appeared numerically larger than the rhyme effect on memory, suggesting a potential misalignment between monitoring and learning outcomes. To adjudicate this, we conducted a random-effects meta-analysis pooling the three studies and directly contrasted the two effects. The results showed no reliable difference, Hedges' $g = 0.10$, 95% CI = [−0.08, 0.28], $p = .268$. Consistent with this aggregate alignment, prior analyses also indicated no difference in absolute accuracy between rhyming and non-rhyming pairs, implying no systematic over- or underestimation at the group level. Together, these findings suggest that, on average, learners' global assessments keep pace with rhyme's modest mnemonic benefit.

However, while the current findings indicate that rhyme serves as a reliable mnemonic cue at the group level, an important question is whether individuals who experience greater mnemonic benefits from rhyme also provide correspondingly higher JOLs. To better explore this relationship, we calculated the correlation between two participant-level difference scores: (1) the rhyme effect on memory (recall for rhyming minus non-rhyming pairs) and (2) the rhyme effect on metamemory (mean JOLs for rhyming minus non-rhyming pairs). Across the three experiments, these correlations were small and non-significant (Experiment 1: $r = .07$, 95% CI [−0.18, 0.31], $p = .59$; Experiment 2: $r = -.06$, 95% CI [−0.30, 0.19], $p = .62$; Experiment 4: $r = .20$, 95% CI [−0.02, 0.40], $p = .07$). Thus, even when calibration is preserved on average, individuals who benefit more from rhyme in recall are not necessarily those who increase their JOLs the most. This pattern reinforces the view that metacognitive judgments often rely on heuristic cues (e.g., phonological fluency or pre-existing beliefs) rather than continuous item- or person-specific updating from experience, helping to explain why relative accuracy can decline even when mean-level calibration remains intact (Dunlosky & Metcalfe, 2009; Koriat, 2006; Rhodes & Castel, 2008).

Belief updating

Another important finding documented here is that both metamemory beliefs and item-level JOLs changed dynamically throughout the study task. In Experiment 4, we assessed not only pre-existing beliefs but also post-study beliefs, and we observed a significant attenuation of the presumed mnemonic benefit of rhyme after study. Specifically, the difference in memory predictions between rhyming and non-rhyming pairs was significantly smaller when measured after the study task than it was before the study task (Frank & Kuhlmann, 2017). Despite this decline, participants still endorsed a residual post-study belief that rhyming pairs were easier to remember than non-rhyming ones. Paired t-tests further confirmed that pre-existing beliefs about the rhyme effect ($M = 15.12$, $SD = 12.34$) were significantly larger than the actual effect ($M = 5.49$, $SD = 14.29$), difference = 9.63 [5.27, 13.99], $t(80) = 4.40$,

$p < .001$, $d = 0.49$, $BF_{10} = 541.95$. However, no detectable difference was found between post-study beliefs about the rhyme effect ($M = 3.47$, $SD = 15.23$) and the actual effect ($M = 5.49$, $SD = 14.29$), difference = −2.03 [−6.66, 2.61], $t(80) = -0.87$, $p = .387$, $d = -0.10$, $BF_{10} = 0.09$. Taken together, these results indicate belief updating toward the empirical signal of the task, with pre-existing overestimation giving way to post-study expectations that approximate the observed memory benefit.

Beyond this shift in general beliefs, we also observed a progressive adjustment in item-level JOLs. Specifically, the rhyme effect on JOLs was significantly larger for the first 10 items compared to the last 10 items, consistent with within-task recalibration of judgments. Notably, these adjustments occurred without explicit external feedback, suggesting that participants modified their JOLs based on learning experience alone. A similar analysis conducted on Experiments 1 and 2, reported in the *Belief Change Analyses* file on OSF (https://osf.io/5a9cz), did not reveal the same pattern. A sensitivity power analysis indicated that these null results were not simply due to a lack of statistical power, as the analyses had 80% power to detect a small-to-moderate interaction effect of $\eta_p^2 \geq 0.02$. One plausible reason for the different outcomes is procedural: in Experiment 4, the pre-study belief questionnaire may have increased the salience of rhyme as a cue, thereby encouraging participants to monitor and update its perceived diagnostic value during study. This finding underscores that eliciting beliefs can itself shape metacognitive monitoring, and that measuring judgments at multiple points can reveal dynamic shifts that might otherwise go undetected.

The adaptive nature of these adjustments aligns with the knowledge updating process, a concept originally explored by Brigham and Pressley (1988). This process describes how learners refine their understanding of memory cues by monitoring performance across study-test experiences. Typically, when learners complete a full study-test cycle, they integrate their experiences into more accurate JOLs in subsequent study phases. For example, Dunlosky and Hertzog (2000) examined knowledge updating by asking participants to study word pairs encoded with either imagery or repetition strategies. Participants initially failed to distinguish the effectiveness of these strategies in their JOLs, assigning similar ratings to both conditions. However, after experiencing the memory test, they updated their JOLs in the second study phase, assigning higher ratings to imagery-encoded pairs, reflecting their improved awareness of the effectiveness of different encoding strategies.

The present findings extend this framework in two important ways. First, they demonstrate that belief revision and JOL updating can occur even in the absence of explicit test feedback. This implies that learners can rely on internally generated evidence, such as covert retrieval or subjective ease during re-encoding, to spontaneously recalibrate their expectations. Second, our results suggest a mechanism for this process: making beliefs explicit before study appears to facilitate such recalibration by focusing attention on the cue's diagnosticity. In this way, even without external correction, participants gradually refined their initial overestimation of rhyme's mnemonic advantage. Future research should specify the conditions under which explicit belief elicitation enhances or impedes optimal updating, and how these belief dynamics translate into the downstream regulation of study time and strategy choice over extended learning cycles.

Impact of rhyme on relative accuracy

Our study found that although rhyme cues provided a mnemonic benefit and participants assigned higher JOLs to rhyming items, these cues also impaired their ability to accurately assess memory performance at the item level. Specifically, participants showed lower relative accuracy when judging their memory for rhyming pairs, suggesting that rhyme hindered their ability to differentiate well-learned from poorly-learned items. Converging analyses using gamma and AUROC2, as well as the meta-analyses across studies, indicated a small but reliable decrement in discrimination for rhymed items. By contrast, absolute accuracy did not differ between conditions, underscoring that the impairment was selective to item-level monitoring rather than global calibration.

This decline in relative accuracy may be due to participants' over-reliance on the perceived ease of learning induced by rhyme, leading them to neglect more diagnostic memory cues and ultimately lose precision in predicting item-level memory performance (Koriat, 1997). Prior research has shown that processing fluency often serves as a metacognitive cue influencing learners' judgments (Koriat, 2006). As a highly fluent linguistic feature, rhyme may have enhanced the subjective ease of encoding, making the materials feel inherently easier to learn and thereby elevating JOLs broadly across rhymed items (Lindstromberg & Boers, 2008). Such broad elevation can compress the effective range of JOLs and divert attention from item-specific, diagnostic cues (for example, successful mediator formation or semantic linkage), which in turn reduces discrimination. Beliefs that rhyme aids memory, documented in Experiments 3 and 4 and shown to mediate JOLs, likely amplify this cue weighting and further bias judgments toward rhyme even when item-wise diagnosticity is limited. Compared to non-rhyming materials, rhyme may have promoted a more intuition-based monitoring process, reducing metacognitive discrimination and thereby impairing relative accuracy (Schwartz & Efklides, 2012).

Additionally, the mnemonic benefits of rhyme may be more context-dependent than learners assume. While some studies have reported that rhyme enhances memory performance (Plastikwala, 2017), others suggest that its effects depend on task demands and encoding conditions (Gupta et al., 2005). Our findings align with this: memory gains were modest and most evident when phonological engagement was heightened by aloud reading, whereas JOL increases were robust across settings. This mismatch, in which cue salience is strong but its diagnosticity is context-dependent, offers a principled explanation for why relative accuracy declined for rhyming items. Language-specific interference remains a possible explanation, but the present evidence does not isolate such mechanisms; future work could manipulate rhyme density and phonological similarity to adjudicate among these accounts.

Practical implications

Our findings provide valuable insights into how learners can effectively use or critically assess rhyme as a learning strategy. While rhyme-based encoding can facilitate learning, its mnemonic benefits are not always reliable, and learners may overestimate its effectiveness due to inflated JOLs. Given that rhyme increased perceived ease without consistently improving recall, learners should treat rhyme as a supplementary cue rather than a stand-alone strategy. A practical approach is to pair phonological regularities with more diagnostic

operations, such as elaboration or semantic mediation, to promote deeper processing and more durable retention (Horton & Pavlick, 1993).

Furthermore, our study found that learners struggled to distinguish well-learned from poorly learned rhyming items, indicating that rhyme may hinder metacognitive discrimination at the item level. To counter this loss of discrimination, instructors can emphasize retrieval practice with immediate correctness feedback and brief confidence ratings, which helps students evaluate their actual memory performance more accurately. Structured reflection prompts, for example asking "*What cues did you use for high-JOL items that were later forgotten?*", can help learners reassess the diagnosticity of rhyme and reweight cues accordingly, thereby reducing fluency-based illusions in judgment.

Beyond individual study strategies, our findings also suggest ways to enhance self-regulated learning tools and interventions. Since Experiment 4 demonstrated that learners can update their beliefs about mnemonic cues, low-cost belief-updating routines, such as comparing block-wise JOLs with subsequent recall or displaying simple calibration and discrimination dashboards, may accelerate recalibration. Digital platforms can surface item-level discrepancies, including high JOL with wrong answer and the converse, and then prompt targeted review that highlights gaps between perceived and actual learning outcomes. By encouraging ongoing checks of cue reliability and by combining rhyme with diagnostic study techniques, such as spaced retrieval and elaborative interrogation, these interventions can mitigate fluency-driven biases and strengthen self-regulated learning over time.

Limitations and future research direction

The findings presented here support the analytic processing framework to some extent, resonating with the observations by Mueller and colleagues (2016) that beliefs play a crucial role in JOL construction. However, the present work did not obtain a direct process measure of fluency (for example, articulation latency, identification time, pupil dilation, or ease-of-processing ratings), which limits a fuller test of analytic-processing theory and of dual-basis models. Future research should concurrently measure both metamemory beliefs and processing fluency within the same participants and trials to more deeply investigate their roles in the rhyme effect on JOLs and, critically, to estimate their unique and shared variance in explaining judgments.

Additionally, while using word pairs as learning materials enhances experimental control, this approach may not fully capture the application of rhyme in real-world learning scenarios, such as the complex textual structures prevalent in poetry and lyrics. Generalizability may also be constrained by the linguistic context and sample composition because all studies used Chinese two-character words and student participants. Future studies should utilize more authentic materials and more diverse populations to uncover a broader spectrum of evidence regarding the mnemonic benefits of rhyme and its impact on JOLs. It would also be valuable to separate the effects of reading aloud from attention or rehearsal, to compare additional phonological patterns such as alliteration and assonance, to examine boundary conditions created by spacing and retrieval practice, and to strengthen causal interpretation of mediation by preregistered designs and observer-learner manipulations that dissociate fluency from belief.

Concluding remarks

Incorporating rhyme in learning materials can enhance memory performance, particularly when learners read rhyming content aloud. This mnemonic benefit is metacognitively acknowledged by learners, as they provide higher JOLs for rhyming than for non-rhyming materials. Metamemory beliefs that rhyming pairs are easier to learn contribute importantly to the rhyme effect on JOLs. Importantly, these beliefs are malleable; our findings show that individuals can adjust their beliefs based on learning experience, decreasing their reliance on rhyme when it proves less diagnostic than initially predicted. However, our findings also suggest that rhyme can impair relative accuracy, as learners sometimes struggled to differentiate well-learned from poorly learned rhyming items. This work therefore moves beyond simple metacognitive illusions to highlight a more nuanced paradox, demonstrating that even diagnostic cues can compromise monitoring when learners rely on them too heavily. This finding demonstrates that the key to effective self-regulated learning lies not just in finding useful mnemonic strategies, but in understanding and navigating their hidden metacognitive costs.

Funding This research was supported by Beijing Philosophy and Social Science Foundation (24DTR067).

Data availability All data and experimental stimuli have been made publicly available via the Open Science Framework (OSF) at https://osf.io/5a9cz/.

Declaration

Ethical approval Not applicable.

Conflict of interest We have no known conflict of interest to disclose.

References

Avons, S., Wragg, C. A., Cupples, W. L., & Lovegrove, W. J. (1998). Measures of phonological short-term memory and their relationship to vocabulary development. *Applied Psycholinguistics, 19*(4), 583–601. https://doi.org/10.1017/S0142716400010377

Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging*, *9*(2), 287. https://doi.org/10.1037/0882-7974.9.2.287

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging*, *3*(3), 249. https://doi.org/10.1037/0882-7974.3.3.249

Bryant, P., MacLean, M., & Bradley, L. (1990). Rhyme, language, and children's reading. *Applied Psycholinguistics*, *11*(3), 237–252. https://doi.org/10.1017/S0142716400008870

Carr, D., & Miles, C. (1997). Rhyme attenuates the auditory suffix effect: Alliteration does not. *Quarterly Journal of Experimental Psychology Section A*, *50*(3), 518–527. https://doi.org/10.1080/713755722

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*, 107–111. https://doi.org/10.3758/BF03194036

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*(2), 443–476. https://doi.org/10.1007/s11409-021-09287-3

Chang, M., & Brainerd, C. J. (2023). The font size effect depends on inter-item relation. *Memory & Cognition*, *51*(7), 1702–1713. https://doi.org/10.3758/s13421-023-01419-1

de Leeuw, J. R. (2015). JsPsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, *15*(3), 462–474. https://doi.org/10.1037/0882-7974.15.3.462

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1180–1191. https://doi.org/10.1037/0278-7393.27.5.1180

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*(4), 545–565.

Fabb, N. (1997). *Linguistics and literature: Language in the verbal arts of the world*. Blackwell.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Frank, D. J., & Kuhlmann, B. G. (2017). More than just beliefs: Experience and beliefs jointly contribute to volume effects on metacognitive judgments. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 43*(5), 680–693. https://doi.org/10.1037/xlm0000332

Fu, C. H., McIntosh, A. R., Kim, J., Chau, W., Bullmore, E. T., Williams, S. C., Honey, G. D., & McGuire, P. K. (2006). Modulation of effective connectivity by cognitive demand in phonological verbal fluency. *NeuroImage, 30*(1), 266–271. https://doi.org/10.1016/j.neuroimage.2005.09.035

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: The roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, *33*(6), 1001–1016. https://doi.org/10.3758/bf03193208

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*(1), 19–54. https://doi.org/10.1177/0002764219859633

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging*, *17*(2), 209–225. https://doi.org/10.1037/0882-7974.17.2.209

Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of learning are influenced by multiple cues in addition to memory for past test accuracy. *Archives of Scientific Psychology*, *1*(1), 23. https://doi.org/10.1037/arc0000003

Hoorn, J. (1996). Psychophysiology and literary processing: ERPs to semantic and phonological deviations in reading small verses. In R. J. K. M. S. MacNealy (Ed.), *Empirical approaches to literature and aesthetics* (Vol. 52, pp. 339–360). Ablex Publishing.

Horton, D. L., & Pavlick, T. J. (1993). Semantic cues, rhyme cues, and two varieties of recognition memory. *Bulletin of the Psychonomic Society*, *31*(1), 16–18. https://doi.org/10.3758/BF03334127

Hu, X., Li, T., Zheng, J., Su, N., Liu, Z., & Luo, L. (2015). How much do metamemory beliefs contribute to the font-size effect in judgments of learning? *PLoS One, 10*(11), e0142351. https://doi.org/10.1371/journal.pone.0142351

Johnson, J. L., & Hayes, D. S. (1987). Preschool children's retention of rhyming and nonrhyming text: Paraphrase and rote recitation measures. *Journal of Applied Developmental Psychology, 8*(3), 317–327. https://doi.org/10.1016/0193-3973(87)90007-4

Király, I., Takacs, S., Kaldy, Z., & Blaser, E. (2017). Preschoolers have better long-term memory for rhyming text than adults. *Developmental Science, 20*(3), Article e12398. https://doi.org/10.1111/desc.12398

Kleider, H. M., & Goldinger, S. D. (2004). Illusions of face memory: Clarity breeds familiarity. *Journal of Memory and Language*, *50*(2), 196–211. https://doi.org/10.1016/j.jml.2003.09.001

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A. (2006). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge University Press.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*(6), 787–794. https://doi.org/10.1177/0956797611407929

Lindstromberg, S., & Boers, F. (2008). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, *29*(2), 200–222. https://doi.org/10.1093/applin/amn007

Macaruso, P., Locke, J. L., Smith, S. T., & Powers, S. (1996). Short-term memory and phonological coding in developmental dyslexia. *Journal of Neurolinguistics*, *9*(2), 135–146. https://doi.org/10.1016/0911-6044(95)00008-9

Matvey, G., Dunlosky, J., Shaw, R. J., Parks, C., & Hertzog, C. (2002). Age-related equivalence and deficit in knowledge updating of cue effectiveness. *Psychology and Aging*, *17*(4), 589. https://doi.org/10.1037/0882-7974.17.4.589

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*(5), 424–428. https://doi.org/10.1111/1467-9280.00282

Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O., & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts, 8*(1), 71–76. https://doi.org/10.1037/a0035309

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. https://doi.org/10.3758/pbr.15.1.174

Molinari, M., & Leggio, M. (2016). Cerebellum and verbal fluency (phonological and semantic). *The linguistic cerebellum* (pp. 63–80). Elsevier. https://doi.org/10.1016/B978-0-12-801608-4.00004-9

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*(2), 378–384. https://doi.org/10.3758/s13423-012-0343-6

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Mueller, M. L., Dunlosky, J., & Tauber, S. K. (2016). The effect of identical word pairs on people's metamemory judgments: What are the contributions of processing fluency and beliefs about memory? *Quarterly Journal of Experimental Psychology, 69*(4), 781–799. https://doi.org/10.1080/17470218.2015.1058404

Murphy, D. H., & Castel, A. D. (2022). Selective remembering and directed forgetting are influenced by similar stimulus properties. *Memory, 30*(9), 1130–1147. https://doi.org/10.1080/09658211.2022.2092152

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2*(4), 267–271. https://doi.org/10.1080/09658211.2017.1406523

Nelson, D. L., Bajo, M. T., & Canas, J. J. (1987). Prior knowledge and memory: The episodic encoding of implicitly activated associates and rhymes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 54–63. https://doi.org/10.1037/0278-7393.13.1.54

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review*, *99*(2), 322–348. https://doi.org/10.1037//0033-295X.99.2.322

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., & Kotz, S. A. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, *4*, 10. https://doi.org/10.3389/fpsyg.2013.00010

Obermeier, C., Kotz, S. A., Jessen, S., Raettig, T., von Koppenfels, M., & Menninghaus, W. (2016). Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cognitive, Affective & Behavioral Neuroscience, 16*(2), 362–373. https://doi.org/10.3758/s13415-015-0396-x

Plastikwala, F. (2017, April 20). *The Use of Rhyme, Rhythm, and Melody as a Form of Repetition Priming to Aid in Encoding, Storage, and Retrieval of Semantic Memories in Alzheimer's Patients* [Oral presentation]. Virginia's Collegiate Honors Council 2017, Norfolk, VA.

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. https://doi.org/10.1037/a0013684

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition*, *1*(3), 145–151. https://doi.org/10.1016/j.jarmac.2012.06.002

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, *22*(4), 1014–1019. https://doi.org/10.3758/s13423-014-0768-1

Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' judgments of learning in multitrial learning. *Psychology and Aging, 27*(2), 474.

Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning? *Psychology and Aging*, *34*(6), 836–847. https://doi.org/10.1037/pag0000376

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304. https://doi.org/10.1016/j.jml.2016.07.003

Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(7), 1073. https://doi.org/10.1037/xlm0000363

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, *99*, 99–110. https://doi.org/10.1016/j.jml.2017.11.005

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S. T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: Methods and pitfalls. *Metacognition and Learning*, *16*(2), 319–343. https://doi.org/10.1007/s11409-020-09254-4

Authors and Affiliations

Mengqi Hu[1] · Anran Li[1] · Chenyuqi Xu[2,3] · Hongyu Chen[4] · Aike Shi[5] · Wenbo Zhao[6] · Xiao Hu[7] · Liang Luo[1,8] · Chunliang Yang[1,7]

✉ Chunliang Yang
chunliang.yang@bnu.edu.cn

[1] Institute of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China

[2] Faculty of Psychology and Educational Sciences, Centre for Instructional Psychology and Technology, KU Leuven, Leuven, Belgium

[3] Imec Research Group Itec, KU Leuven, Kortrijk, Belgium

[4] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Science, Beijing, China

[5] Department of Psychology, The Chinese University of Hong Kong, Hong Kong SAR, China

[6] School of Social Development and Public Policy, Beijing Normal University, Beijing, China

[7] Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Beijing Normal University, Beijing, China

[8] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China