

The Effect of Difficulty on the Font Size Effect: The Role of Deep Semantic Encoding

Authors: Ma Xiaoxiao, Yingjie Jiang, Guo Yanlin, Zhang Xiaojing, Jiang Yingjie

Date: 2024-05-09T00:00:00+00:00

Abstract

Using event-related potential (ERP) technology, we investigated the influence of difficulty on the font size effect. The results revealed: (1) Compared with simple word pairs, participants assigned lower judgments of learning (JOLs) to difficult word pairs; however, there was no difference in JOLs between large and small fonts. (2) Font size affected the early encoding stage (200~300 ms), eliciting shallow processing (non-semantic encoding) of learning items; difficulty affected the middle-to-late encoding stage (450~600 ms), with difficult word pairs eliciting deep processing (semantic encoding) of learning items. (3) When making JOLs, participants elicited the LPN component in the frontal lobe, which is representative of attempted retrieval. These results indicate that difficulty cues during the encoding stage eliminate the font size effect by eliciting semantic encoding of items; during the JOLs stage, individuals engage in attempted retrieval, and in this process, they make accurate JOLs based on semantically encoded information. The findings of this study not only elucidate the electrophysiological mechanisms underlying the generation and elimination of the font size effect, but also provide neural evidence for how to improve the accuracy of JOLs, a metamemory monitoring process.

Full Text

Abstract

This study investigated the influence of difficulty on the font size effect using event-related potentials (ERPs). The results revealed: (1) Participants gave lower judgments of learning (JOLs) for difficult word pairs compared to easy word pairs, while JOLs did not differ between large and small font sizes. (2) Font size affected early encoding stages (200-300 ms), eliciting shallow processing (non-semantic encoding) of learning items, whereas difficulty influenced middle-to-late encoding stages (450-600 ms), with difficult word pairs eliciting

deep processing (semantic encoding). (3) During JOLs, a late positive component (LPN) representing attempted retrieval was elicited in the frontal cortex. These findings demonstrate that difficulty cues eliminate the font size effect by inducing semantic encoding during the encoding phase. During the JOLs phase, individuals engage in attempted retrieval based on semantic information to make accurate JOLs. This study not only elucidates the electrophysiological mechanisms underlying the generation and elimination of the font size effect but also provides neural evidence for improving the accuracy of JOLs as a metamemory monitoring process.

Keywords: font size; JOLs; metamemory; memory; ERPs

Introduction

In daily life, people often improve learning efficiency through reflective evaluation of their own learning and memory. This self-assessment of memory accuracy is termed metamemory (McDonough et al., 2021; Undorf et al., 2022). Metamemory comprises two components: metamemory monitoring and metamemory control (Dunlosky et al., 2016; Dunlosky & Metcalfe, 2008; Nelson & Narens, 1990; Rivers et al., 2021). Among existing research, the most commonly used metamemory monitoring indicator is judgments of learning (JOLs), which refer to predictive judgments individuals make about their memory performance after studying items but before being tested (Nelson & Narens, 1994). Accurate JOLs form the foundation for effective self-regulation in learning. However, a disconnect often emerges between individuals' judgments of their memory performance and their actual memory performance, resulting in metamemory illusions that can lead to misallocation of cognitive resources and reduced learning efficiency (Undorf et al., 2022).

The font size effect represents a typical metamemory illusion. It describes the phenomenon where, despite equivalent actual memory performance for large and small font items, individuals erroneously believe they will remember large-font items better. Using a classic JOLs paradigm, Rhodes and Castel (2008) presented experimental materials in 18pt small font and 48pt large font, employing a study-JOLs-test procedure. They found that participants gave higher JOLs to large-font items, yet recall performance showed no significant difference between large and small fonts—a dissociation between memory and metamemory they termed the font size effect.

While the font size effect produces metamemory illusions, accurate metamemory is crucial for optimal cognitive resource utilization and memory improvement. To reduce the font size effect, Rhodes and Castel introduced word-pair relatedness by presenting related and unrelated word pairs in both 18pt small font and 48pt large font. The results showed that the font size effect diminished, with participants primarily using relatedness as a cue for JOLs, thereby improving JOLs accuracy. When both relatedness and font size cues were available, the effect size of font size was significantly smaller than when font size was the sole

cue, though the font size effect persisted. Mendes and Undorf (2022) examined the influence of word frequency and font size on JOLs, finding that participants primarily relied on word frequency as the main cue for their judgments.

The introduction of relatedness and word frequency weakened but did not eliminate the font size effect, contradicting researchers' hypotheses. Rhodes and Castel (2008) argued that relatedness, being more diagnostic than font size, should eliminate the font size effect entirely. Some studies have used relatedness and word frequency as bases for controlling difficulty (Soderstrom & McCabe, 2011), suggesting that their influence on the font size effect might actually reflect the role of difficulty. Psychometrics uses pass rates as the criterion for classifying difficulty, yet no previous research has controlled material difficulty using pass rates. This may explain why difficulty has not completely eliminated the font size effect—JOLs may be more sensitive to difficulty controlled by pass rates than to relatedness or word frequency.

Second, while reducing metamemory illusions has attracted considerable research attention, few studies have explored why font size leads to biased JOLs or why introducing difficulty can reduce metamemory illusions and improve JOLs accuracy. ERPs technology offers strong temporal resolution, enabling millisecond-level analysis of cognitive processes such as encoding and JOLs. By examining relevant ERP components, we can reveal dynamic changes in attentional resources and cognitive control when both difficulty and font size cues are present, thereby addressing the neural mechanisms through which difficulty eliminates the font size effect.

Bayer et al. (2012) investigated ERP effects of font size for emotional and neutral words. Although P1 and N1 amplitudes were not affected by emotion, large-font words elicited earlier onset and longer duration of early posterior negativity compared to small-font words. The P1 component is associated with early visual processing of attentional work (Kalsi et al., 2019) and is considered an exogenous component influenced solely by physical stimulus characteristics. Zhou (2022) found different P2 amplitudes for different fonts in ERP waveforms during word-pair learning. P2 is an endogenous component appearing in frontal and parieto-occipital regions associated with orthographic processing of characters (Misra & Holcomb, 2003). These studies suggest that font size influences early encoding stages.

Undorf et al. (2020) examined the neural mechanisms of difficulty's influence on JOLs using related and unrelated word pairs. They found that related word pairs received higher JOLs but elicited smaller N400 amplitudes, indicating higher retrieval fluency. Seren (2022) used ERPs to explore categorically related word pairs in long-term memory, finding differential effects of related, unrelated, and pseudo-word conditions on LPC amplitude. Both LPC and N400 appear in middle-to-late encoding stages; LPC represents a cognitive processing and evaluation phase (Fan & Han, 2008) involving elaborative processing. Research on word-pair relatedness suggests that relatedness influences middle-to-late encoding components, during which individuals engage in deep processing of learning

items.

Moreover, previous studies lacked direct ERP recordings during the JOLs phase, potentially overlooking how cues operate. Therefore, using ERPs to examine the temporal characteristics of difficulty and font size during both encoding and JOLs phases can reveal the electrophysiological mechanisms through which difficulty influences the font size effect.

In summary, Experiment 1 of this study used pass rates to control difficulty and examined whether difficulty could influence the font size effect when both served as JOLs cues, thereby improving JOLs accuracy. We hypothesized that when both font size and difficulty cues were available, participants would ignore font size and base JOLs on difficulty, with difficulty weakening the font size effect by influencing both memory and metamemory. Experiment 2 employed ERPs technology to analyze ERP amplitudes during encoding and JOLs phases, investigating why difficulty attenuates the font size effect when both cues are present.

Experiment 1a

Experiment 1a used a single-factor (font size: 18pt, 48pt) within-subjects design to replicate the font size effect and verify its existence in the present study. The procedure followed the classic JOLs paradigm.

2.1.1 Participants

The same participants completed Experiment 1b one week after Experiment 1a; therefore, sample size was determined based on the combined experiments. Using G*Power 3 with a small effect size of $d = 0.25$ (Grogan et al., 2021), a minimum of 24 participants was required for power $(1 - \beta) = 0.8$ and significance level $\alpha = 0.05$. Twenty-eight participants were recruited from a university in Northeast China. Based on final recall performance, participants with excessively low recall accuracy (below 10%) were excluded, retaining 26 valid participants (10 males, 16 females, $M = 20.35$ years, $SD = 2.46$). All participants volunteered for the experiment, had no prior experience with JOLs experiments, and had normal or corrected-to-normal vision. The experiment followed ethical requirements, and participants signed informed consent forms before the experiment began.

2.1.2 Apparatus

The experimental program was developed using E-Prime 3.0 software. The experiment was conducted on a Lenovo multimedia computer with a 21-inch CRT color monitor (resolution = 1920×1080 pixels; refresh rate = 75 Hz). Data processing and analysis were performed using SPSS 26.0.

2.1.3 Materials

Word-pair difficulty was indexed by pass rates in cued recall tests. Low pass rates (0.03–0.13) indicated difficult word pairs, high pass rates (0.63–0.80) indicated easy word pairs, and pass rates around 0.5 indicated medium-difficulty word pairs (姜英杰 et al., 2023). Stroke and character form were balanced across conditions. Experiment 1a aimed to verify the existence of the font size effect; therefore, difficulty was controlled at medium levels (pass rates: 0.43–0.60). Four word pairs were selected for practice, and 30 word pairs were used for the formal experiment, totaling 34 word pairs. Materials were randomly divided into two groups: one presented in 18pt small font and the other in 48pt large font, with random presentation order during testing.

2.1.4 Procedure

This experiment used a two-factor within-subjects design following the classic JOLs paradigm: study-JOLs-test, with testing after all word pairs were studied. A fixation point appeared at the center of the screen for 500 ms, followed by a word pair presented in black font for 4 s. Participants were required to memorize the pair within 4 s. After each word pair, participants made JOLs ratings from 0 to 100 based on their memory confidence (0 = would definitely not remember, 100 = would definitely remember). Following the study phase, participants completed a distractor task to prevent rehearsal: orally reporting the results of continuously subtracting 4 from 3000 for 3 minutes. Finally, a cued recall test was administered, presenting the left cue word on screen for participants to fill in the target word (Figure 1 [Figure 1: see original paper]).

Figure 1. Experimental procedure for Experiment 1

2.2 Results

A paired-samples t-test examined the effect of font size on JOLs, revealing a significant main effect: $t(1, 25) = 2.366$, $p = 0.026$, $d = 0.18$. Large-font JOLs ($M = 52.65$) were significantly higher than small-font JOLs ($M = 49.28$), indicating that when font size was the only available cue, participants noticed and used this cue for JOLs.

Another paired-samples t-test examined font size effects on recall accuracy, showing no significant main effect: $t(1, 25) = 0.113$, $p = 0.911$. Recall accuracy did not differ significantly between large-font ($M = 22.31$) and small-font ($M = 22.05$) word pairs (Figure 2 [Figure 2: see original paper]).

Figure 2. Effects of font size on JOLs and recall accuracy. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Error bars represent standard errors.

Experiment 1a controlled difficulty at medium levels with only font size as a cue. The results demonstrated that when font size was the sole cue, participants used it for JOLs, producing a metamemory illusion. These findings replicate previous research, confirming the font size effect.

Experiment 1b

Experiment 1b employed a 2 (font size: 18pt, 48pt) \times 2 (difficulty: easy, difficult) within-subjects design, alternating large- and small-font word pairs to examine how difficulty cues influence the font size effect. The procedure followed the classic JOLs paradigm.

3.1.1 Participants

The same participants from Experiment 1a completed this experiment.

3.1.3 Materials

Sixty-six word pairs were selected. Based on cued recall pass rates, 30 difficult and 30 easy word pairs were chosen for the formal experiment, plus four medium-difficulty (0.43–0.60) pairs for practice and two additional pairs to control for primacy and recency effects. Formal experiment materials were randomly divided into two groups of 30 pairs each, with one group presented in 18pt small font and the other in 48pt large font. Presentation order was randomized.

3.2 Results

Table 1 presents means and standard deviations for recall accuracy and JOLs across difficulty and font size conditions.

3.2.1 Effects of Difficulty and Font Size on JOLs A 2×2 repeated-measures ANOVA on JOLs revealed no significant main effect of font size, $F(1, 25) = 0.001$, $p = 0.980$, with no difference between large-font ($M = 49.57$) and small-font ($M = 49.60$) JOLs. The main effect of difficulty was significant, $F(1, 25) = 17.306$, $p < 0.001$, $p^2 = 0.409$, with easy word pairs receiving higher JOLs ($M = 53.54$) than difficult pairs ($M = 45.64$). The interaction was not significant, $F(1, 25) = 0.770$, $p = 0.388$, indicating that difficulty eliminated the font size effect. Regardless of difficulty level, JOLs did not differ between large and small fonts, suggesting participants ignored font size cues (Figure 3 [Figure 3: see original paper]).

3.2.2 Effects of Difficulty and Font Size on Recall Accuracy A 2×2 repeated-measures ANOVA on recall accuracy showed no significant main effect of font size, $F(1, 25) = 1.741$, $p = 0.065$, with no difference between large-font ($M = 29.42$) and small-font ($M = 27.76$) recall. The main effect of difficulty was significant, $F(1, 25) = 120.832$, $p < 0.001$, $p^2 = 0.829$, with easy word pairs showing higher recall accuracy ($M = 41.41$) than difficult pairs ($M = 15.77$). The interaction was not significant, $F(1, 25) = 0.165$, $p = 0.688$. Both large- and small-font easy word pairs demonstrated higher recall accuracy than difficult pairs (Figure 3 [Figure 3: see original paper]).

Figure 3. Effects of font size and difficulty on JOLs and recall accuracy

3.3 Discussion

When both difficulty and font size cues were available in Experiment 1b, participants based JOLs solely on difficulty, with font size affecting neither JOLs nor recall performance. Difficulty thus eliminated the font size effect by influencing cue selection during JOLs. This likely occurred because difficulty receives deeper processing than font size, becoming the primary basis for JOLs when both cues are present.

Experiment 2

While Experiment 1 demonstrated behaviorally that difficulty can eliminate the font size effect, the underlying mechanism remains unclear. Some research suggests that enhanced encoding strength improves JOLs accuracy and eliminates the font size effect (赵文博 et al., 2020). Difficulty may thus reduce the font size effect by promoting deep semantic processing of learning items. However, previous studies have only inferred this from behavioral data through experimental manipulations. Whether difficulty eliminates the font size effect by increasing processing depth requires further investigation. Experiment 2 employed ERPs to examine the neural basis of difficulty's effect on the font size effect.

4.1.1 Participants

Using G*Power 3 with a small effect size of $d = 0.25$ (Grogan et al., 2021), a minimum of 20 participants was required for power $(1 - \beta) = 0.8$ and $\alpha = 0.05$. Twenty-seven participants were recruited from a university in Northeast China. Participants with rejection rates exceeding 35% during preprocessing were excluded, retaining 25 valid participants (10 males, 15 females, $M = 21.48$ years, $SD = 3.34$). Requirements were identical to Experiment 1a.

4.1.2 Materials

Difficulty was classified using cued recall pass rates. Two hundred six word pairs were selected: 100 difficult pairs (pass rates: 0.03-0.13) and 100 easy pairs (pass rates: 0.63-0.80) for the formal experiment, plus six medium-difficulty (0.43-0.60) pairs for practice. Stroke and character form were balanced. Materials were randomly divided into two groups: one presented in 14pt small font and the other in 50pt large font, with random presentation order. The experiment consisted of four blocks, each studying 50 word pairs.

4.1.3 Apparatus and Data Analysis

The apparatus and data analysis software were identical to Experiment 1a. EEG data recording was programmed using E-Prime 3.0 and presented on the same Lenovo multimedia computer running Windows 10. Participants sat approximately 80 cm from the monitor and were instructed to remain as still as possible.

4.1.4 Procedure

This two-factor within-subjects experiment followed the classic JOLs paradigm: study-JOLs-test, with testing after all word pairs were studied. The study phase procedure is illustrated in Figure 3.1. Instructions informed participants that materials would appear in large and small fonts to ensure they noticed the font size cue. A fixation point appeared for 500 ms, followed by a word pair in black font (50pt or 14pt) for 4 s. Participants memorized each pair within 4 s and then made high or low JOLs (1 = low confidence, 2 = high confidence). After each block of 50 word pairs, participants could choose to rest before completing a recognition test. Four blocks were administered, each comprising a complete study-JOLs-test cycle. The procedure was identical to Experiment 1 except that the test phase used recognition rather than recall.

4.1.5 EEG Recording

EEG data were recorded using an ESI-64 system (NeuroScan) with a 64-channel electrode cap based on the international 10-20 system. The left mastoid served as the online reference, with the right mastoid as the recording electrode. Offline analysis re-referenced data to the average of left and right mastoids. Signals were digitized at 1000 Hz with a 0.05–100 Hz bandpass filter, maintaining electrode impedance below 10 K Ω . Preprocessing used EEGLAB in Matlab: downsampling to 500 Hz, 0.1–40 Hz bandpass filtering, and baseline correction using the 200 ms pre-stimulus interval. Epochs from -200 ms to 800 ms relative to stimulus onset were extracted for each font size and difficulty condition. Eye blink and horizontal eye movement artifacts were corrected, and trials with amplitudes exceeding ± 100 μ V were rejected before averaging. Further analysis was conducted using ERPlab.

4.2 Results

Table 2 presents means and standard deviations for JOLs and recognition accuracy across conditions.

4.2.1 Effects of Difficulty and Font Size on JOLs A 2×2 repeated-measures ANOVA on JOLs showed no significant main effect of font size, $F(1, 24) = 1.104$, $p = 0.304$, with no difference between large-font ($M = 0.57$) and small-font ($M = 0.59$) JOLs. The main effect of difficulty was significant, $F(1, 24) = 61.981$, $p < 0.001$, $\eta^2 = 0.721$, with easy word pairs receiving higher JOLs ($M = 0.68$) than difficult pairs ($M = 0.48$). The interaction was not significant, $F(1, 24) = 0.081$, $p = 0.778$. Easy word pairs received higher JOLs than difficult pairs regardless of font size, consistent with Experiment 1b (Figure 4 [Figure 4: see original paper]).

4.2.2 Effects of Difficulty and Font Size on Recognition Accuracy A 2×2 repeated-measures ANOVA on recognition accuracy revealed no significant

main effect of font size, $F(1, 24) = 1.915$, $p = 0.179$, with no difference between large-font ($M = 0.71$) and small-font ($M = 0.72$) accuracy. The main effect of difficulty was significant, $F(1, 24) = 83.231$, $p < 0.001$, $p^2 = 0.776$, with easy word pairs showing higher recognition accuracy ($M = 0.78$) than difficult pairs ($M = 0.65$). This pattern matched participants' JOLs. The interaction was not significant, $F(1, 24) = 0.144$, $p = 0.708$. Easy word pairs showed higher accuracy than difficult pairs for both font sizes (Figure 4 [Figure 4: see original paper]).

Figure 4. Effects of difficulty and font size on JOLs and recognition accuracy

4.2.3 Absolute Accuracy of JOLs A 2 (font size: 14pt, 50pt) \times 2 (difficulty: easy, difficult) repeated-measures ANOVA on JOLs accuracy revealed no significant main effect of font size, $F(1, 24) = 0.068$, $p = 0.796$, with no difference between large-font ($M = -0.13$) and small-font ($M = -0.12$) accuracy. The main effect of difficulty was significant, $F(1, 24) = 13.238$, $p = 0.001$, $p^2 = 0.355$, with easy word pairs showing higher accuracy ($M = -0.08$) than difficult pairs ($M = -0.17$). The interaction was not significant, $F(1, 24) = 0.356$, $p = 0.556$. Participants underestimated their memory performance across all conditions, with higher metamemory accuracy for easy word pairs (Figure 5 [Figure 5: see original paper]).

Figure 5. Absolute accuracy

4.3 ERP Results

4.3.1 Encoding Phase P2 is an endogenous component appearing in frontal and parieto-occipital regions associated with orthographic processing (Misra & Holcomb, 2003). We analyzed amplitudes in the 200–300 ms time window in left frontal and frontocentral regions of interest (F3, F5, FC3, FC5) using a 2 (font size: 14pt, 50pt) \times 2 (difficulty: easy, difficult) repeated-measures ANOVA. Results showed a significant main effect of font size in left frontal and frontocentral regions, $F(1, 24) = 7.933$, $p = 0.01$, $p^2 = 0.248$, with large-font word pairs ($M = 2.61 \mu\text{V}$) eliciting more positive P2 amplitudes than small-font pairs ($M = 1.82 \mu\text{V}$). The main effect of difficulty was not significant, $F(1, 24) = 0.003$, $p = 0.958$, with no difference between difficult ($M = -2.79 \mu\text{V}$) and easy ($M = -1.60 \mu\text{V}$) pairs. The interaction was not significant, $F(1, 24) = 1.058$, $p = 0.314$ (Figure 6 [Figure 6: see original paper]).

P3 represents attention and orientation processes, while LPC similarly reflects cognitive processing and evaluation (Fan & Han, 2008). This stage involves elaborative processing, with semantic and orthographic encoding eliciting different LPC amplitudes. Larger LPC amplitudes are observed when items receive more thorough processing (Olofsson et al., 2008; Song et al., 2019). LPC appears in parietal regions within a 400–700 ms time window (Meng et al., 2012; 程家萍 et al., 2017; 杨东 et al., 2015). We analyzed amplitudes in the 450–600 ms time window in right parietal regions of interest (CP2, CP4, P2, P4) using a 2×2

repeated-measures ANOVA. Results showed no significant main effect of font size, $F(1, 24) = 0.305$, $p = 0.586$, with no difference between large-font ($M = 0.37 \mu\text{V}$) and small-font ($M = 0.55 \mu\text{V}$) amplitudes. The main effect of difficulty was marginally significant, $F(1, 24) = 4.156$, $p = 0.053$, $p^2 = 0.148$, with difficult word pairs ($M = 0.96 \mu\text{V}$) eliciting more negative amplitudes than easy pairs ($M = -0.04 \mu\text{V}$). The interaction was not significant, $F(1, 24) = 0.2751$, $p = 0.11$ (Figure 6).

Figure 6. Topographic maps and waveforms for P2 and LPC during encoding phase. Note: DL = 48pt difficult, DS = 18pt difficult, EL = 48pt easy, ES = 18pt easy

4.3.2 JOLs Phase Based on previous research and our results, we analyzed LPN amplitudes in the 850-1000 ms time window at right frontal sites (F2, F4, F6). A 2 (font size: 14pt, 50pt) \times 2 (difficulty: easy, difficult) repeated-measures ANOVA revealed no significant main effect of font size, $F(1, 24) = 0.389$, $p = 0.539$, with no difference between large-font ($M = -2.00 \mu\text{V}$) and small-font ($M = -2.00 \mu\text{V}$) amplitudes. The main effect of difficulty was significant, $F(1, 24) = 4.257$, $p = 0.05$, $p^2 = 0.151$, with difficult word pairs ($M = -2.79 \mu\text{V}$) eliciting more negative amplitudes than easy pairs ($M = -1.60 \mu\text{V}$). The interaction was not significant, $F(1, 24) = 0.24$, $p = 0.628$ (Figure 7 [Figure 7: see original paper]).

Figure 7. Topographic maps and waveforms for LPN during JOLs phase

The behavioral results replicated Experiment 1b, demonstrating consistency and stability in showing that difficulty can eliminate the font size effect by influencing JOLs. At the ERP level, font size affected the early encoding component P2, where large-font words elicited larger P2 amplitudes, indicating that orthographic processing of large-font words required more cognitive resources. Difficulty affected LPC during middle-to-late encoding stages, where participants engaged in semantic encoding—deep processing of learning items. Difficult word pairs elicited larger LPC amplitudes, indicating deeper encoding compared to easy pairs.

During the JOLs phase, the LPN component associated with memory retrieval was observed. These ERP results provide neural-level explanations for the behavioral findings. Perceptual features of font size are processed during early encoding, when individuals engage in shallow encoding based on visual characteristics. Perception of difficulty occurs during middle-to-late encoding, when individuals encode semantic information about word pairs, engaging in deeper processing that facilitates effective JOLs and corrects metamemory illusions (赵文博 et al., 2020). The LPN component indicates that during JOLs, individuals make judgments based on analytic cues like difficulty, engaging in attempted retrieval and cognitive control based on conceptual information to make more accurate JOLs and reduce metamemory illusions.

General Discussion

This study examined whether difficulty cues can eliminate the font size effect—a metamemory illusion—and investigated the underlying mechanism. Experiment 1 used the classic JOLs paradigm to test whether difficulty could eliminate the font size effect. Results showed that participants primarily used difficulty cues for JOLs, which improved accuracy and eliminated the font size effect. Experiment 2 employed ERPs to further explore why difficulty eliminates the font size effect. Behavioral results replicated Experiment 1, confirming that difficulty cues eliminate the font size effect and improve JOLs accuracy, enabling effective metamemory monitoring.

5.1 Difficulty Eliminates the Font Size Effect

Previous research indicates that relying solely on font size as a perceptual feature cue produces metamemory illusions. Researchers have introduced word-pair relatedness and word frequency to reduce these illusions and improve JOLs accuracy. Rhodes and Castel (2008) found that font size effects diminished when relatedness was introduced, and Mendes and Undorf (2022) observed similar results with word frequency. However, contrary to hypotheses, the font size effect was not completely eliminated.

Previous studies have used relatedness or word frequency as proxies for difficulty, assuming high-relatedness and high-frequency words are easier to remember (Soderstrom & McCabe, 2011). However, these approaches do not reflect psychometric criteria for difficulty control. According to Koriat’s definition of internal cues and psychometric standards, pass rates provide the optimal basis for difficulty control, with high-pass-rate items being easy and low-pass-rate items being difficult. The present study used pass rates to control difficulty and examined whether difficulty could eliminate the font size effect and improve JOLs accuracy. Results demonstrated that participants selected difficulty as the primary JOLs cue, causing font size cues to become irrelevant and eliminating the font size effect. Participants noticed font size differences but still based JOLs on difficulty, ignoring font size cues. Difficulty thus eliminated the font size effect and improved metamemory accuracy by influencing cue selection during JOLs. Experiment 2 altered the test format and increased the font size difference, yet results remained consistent with Experiment 1, demonstrating that difficulty’s elimination of the font size effect is stable and unaffected by test format.

This study found no significant main effect of font size or interaction between difficulty and font size, contrasting with some previous research. Earlier studies showed that JOLs are highly sensitive to semantic relatedness (Sheffer & Bar, 2004), suggesting that relatedness might be a strong enough cue to reduce font size effects. The introduction of more diagnostic variables should theoretically eliminate font size effects on JOLs (Rhodes & Castel, 2008). However, previous research using relatedness did not eliminate the font size effect, contrary to

hypotheses. The present study, using pass-rate-controlled difficulty, successfully eliminated the font size effect, aligning with original hypotheses. This suggests that difficulty controlled by pass rates may be a more diagnostic variable than relatedness, with individuals being more sensitive to item difficulty than to relatedness during encoding.

5.2 Semantic Encoding Improves JOLs Accuracy

Using ERPs, this study revealed the temporal characteristics underlying how difficulty cues eliminate the font size effect during both encoding and JOLs phases. During learning, P2 and LPC components were observed. Large fonts elicited more positive P2 amplitudes than small fonts in the 200–300 ms window over left frontal regions. P2 is associated with orthographic processing (Misra & Holcomb, 2003), and although both font sizes elicited P2, this indicates that font size encoding involves perceptual processing of character form—non-semantic encoding and shallow processing. The study also found LPC amplitudes in the 450–600 ms window over right parietal regions. Previous research indicates LPC represents attention and orientation processes elicited during semantic encoding, representing a stage of clear item identification and elaborative processing (Fan & Han, 2008). Difficult word pairs elicited larger LPC amplitudes, indicating that difficulty prompts deep encoding to improve memory performance. Deep encoding during learning can activate retrieval-related neural processes.

In levels-of-processing research, perceptual tasks typically serve as non-semantic tasks while relatedness judgments serve as semantic tasks, a manipulation proven effective for controlling processing depth (Nieznanski, 2020). This suggests that difficulty's effect on the font size effect may stem from its ability to induce deep processing. Font size processing occurs at the perceptual level, similar to shallow processing tasks in levels-of-processing research, where participants first process physical features like font size. Difficulty processing resembles semantic tasks in levels-of-processing research, where semantic information is encoded. Compared to font size, difficulty cues elicit more important semantic processing. Research also indicates that deep processing produces stronger memory traces than shallow processing, thereby improving memory performance (Walla et al., 2001). Font size as a perceptual cue triggers shallow encoding, while difficulty cues promote semantic encoding associated with late deep processing. Therefore, difficulty can eliminate the font size effect and improve JOLs accuracy by increasing processing depth.

Direct ERP analysis revealed that JOLs are based not solely on analytic, theory-based reasoning or non-analytic, experience-based heuristics. The LPN component was elicited in frontal regions during the 850–1000 ms JOLs phase, with different amplitudes for easy and difficult word pairs. LPN has been associated with cognitive control processes during conceptually driven retrieval and memory evaluation of retrieved feature information. This indicates that during JOLs, individuals use semantic encoding-related cues while ignoring perceptual cues. Deep processing affects neural activation related to memory retrieval, resulting

in differential LPN amplitudes for difficulty cues requiring semantic processing. This potentially suggests that JOLs for verbal materials are based on semantic rather than visual feature retrieval.

5.3 Limitations and Future Directions

This study examined how difficulty and font size operate as cues during JOLs, finding that when both are available, individuals select difficulty as the primary cue, thereby correcting metamemory illusions induced by font size. In Experiment 1a, where font size was the only cue, participants gave higher JOLs to large fonts, but the effect size was small, suggesting low utilization of the font size cue. Future research could modify experimental designs to make the font size cue more salient to test whether difficulty can still modulate it.

During JOLs analysis, we found evidence of attempted retrieval processes, suggesting difficulty influences JOLs not only through encoding fluency but also through attempted retrieval. Future studies could use fMRI to examine whether attempted retrieval occurs during JOLs for pictures and whether pictures and words activate different brain regions during attempted retrieval.

Conclusion

Difficulty cues can eliminate the font size effect and improve JOLs accuracy. (1) During encoding, difficult word pairs prompt semantic encoding, and the semantic encoding induced by difficulty increases processing depth, thereby eliminating the font size effect. (2) During the JOLs phase, individuals engage in attempted retrieval based on semantically encoded information while ignoring non-semantically encoded information, thereby improving JOLs accuracy.

References

- McDonough, I. M., Enam, T., Kraemer, K. R., Eakin, D. K., & Kim, M. (2021). Is there more to metamemory? An argument for two specialized monitoring abilities. *Psychonomic Bulletin & Review*, 28(5), 1657-1667.
- Undorf, M., Navarro-Báez, S., & Zimdahl, M. F. (2022). Metacognitive illusions. *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment, and Memory*, 307
- Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. *Handbook of metamemory and memory*, 11-28.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about knowing*, 13,
- Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review*, 33(3), 823-862.

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.

Mendes, P. S., & Undorf, M. (2022). On the pervasive effect of word frequency in metamemory. *Quarterly Journal of Experimental Psychology*, 75(8), 1411-1427.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1236.

Kalsi, N., Tambelli, R., Altavilla, D., Trentini, C., Panunzi, S., Stanca, M., ...& Lai, C. (2018). Neurophysiological correlate of emotional regulation in cognitive and motor deficits in Tourette' s syndrome. *The World Journal of Biological Psychiatry*.

Aibao, Z. H. O. U., Haiyan, Z. H. A. O., Ruixue, X. I. A., Yue, Y. U. A. N., Tingting, J. I. A., & Pei, X. I. E. The Impact of Reference Styles, Emotional Valence, and Personality-Trait Assessment at Different Semantic Levels on Memory. *Studies of Psychology and Behavior*, 20(2), 145.

Misra, M., & Holcomb, P. (2003). Event-related potential indices of masked repetition priming. *Psychophysiology*, 40(1), 115-130.

Undorf, M., Amaefule, C. O., & Kamp, S. M. (2020). The neurocognitive basis of metamemory: Using the N400 to study the contribution of fluency to judgments of learning. *Neurobiology of Learning and Memory*, 169,

Düzenli-Öztürk, S., Hünlerli-Gündüz, D., Emek-Savaş, D. D., Olichney, J., Yener, G. G., & Ergenç, H. İ. (2022). Taxonomically-related word pairs evoke both N400 and LPC at long SOA in Turkish. *Journal of Psycholinguistic Research*, 51(6), 1431-1451.

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.

Nieznański, M. (2020). Levels-of-processing effects on context and target recollection for words and pictures. *Acta Psychologica*, 209, 103127.

Walla, P., Hufnagl, B., Lindinger, G., Imhof, H., Deecke, L., & Lang, W. (2001). Left temporal and temporoparietal brain activity depends on depth of word encoding: A magnetoencephalographic study in healthy young subjects. *NeuroImage*, 13(3), 402-409.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.