

## The Effect of Encoding Strength on the Font Size Effect

**Authors:** Wenbo Zhao, Jiang Yingjie, Wang Zhiwei, Hu Jingyuan, Jiang Yingjie

**Date:** 2020-06-12T00:00:00+00:00

### Abstract

This study employed three experiments to investigate the effect of encoding strength on the font size effect, exploring the underlying generation mechanism of metacognitive illusions induced by perceptual features (Experiment 1) and effective corrective measures (Experiments 2 and 3). The results revealed that: (1) The perceptual fluency of words in large font was significantly better than that of words in small font, and Bayesian multilevel mediation analysis results indicated that perceptual fluency partially mediated the font size effect (Experiment 1); (2) As encoding strength increased, the judgment of learning illusion induced by font size gradually disappeared (Experiments 2 and 3). These results indicate that the influence of stimulus perceptual features (font size) on individuals' judgment of learning gradually weakened with the increase in cues activated by encoding strength. This finding provides a scientific basis for enhancing learners' encoding strength in real-world teaching contexts, thereby reducing the reliance of judgment of learning on perceptual feature cues and accurately monitoring one's own learning progress.

### Full Text

## The Influence of Encoding Strength on the Font Size Effect

**ZHAO Wenbo, JIANG Yingjie, WANG Zhiwei, HU Jingyuan**

School of Psychology, Northeast Normal University, Changchun 130024, China

### Abstract

This study employed three experiments to investigate how encoding strength influences the font size effect, examining both the underlying mechanism of metacognitive illusions triggered by perceptual features (Experiment 1) and effective corrective measures (Experiments 2 and 3). The results revealed that:

(1) large-font words exhibited significantly better perceptual fluency than small-font words, and Bayesian multilevel mediation analysis indicated that perceptual fluency partially mediated the font size effect (Experiment 1); (2) as encoding strength increased, the judgment of learning (JOL) illusion induced by font size gradually disappeared (Experiments 2 and 3). These findings demonstrate that the influence of perceptual features (font size) on individuals' learning judgments diminishes as encoding strength activates additional cues. This provides a scientific basis for enhancing learners' encoding strength in real educational contexts, thereby reducing reliance on perceptual feature cues and enabling more accurate monitoring of one's own learning progress.

**Keywords:** judgments of learning; font size effect; encoding strength; metacognitive illusions

**Classification:** B842

**Received Date:** 2019-12-04

**Funding:** National Social Science Foundation of China (19BSH113)

**Corresponding Author:** JIANG Yingjie, E-mail: jiangyj993@nenu.edu.cn

---

## 1 Introduction

Judgments of learning (JOLs) refer to metacognitive evaluations in which learners assess the likelihood of successfully retrieving studied items in future tests (Tauber & Dunlosky, 2016). Learners typically base their subsequent study plans and allocation of cognitive resources on their current JOLs (Flavell, 1979), such as adjusting study time allocation, deciding whether to restudy, or determining when to stop learning (Nelson, Dunlosky, Graf, & Narens, 1994). Consequently, the accuracy of JOLs directly influences individuals' learning strategy selection and, in turn, affects the efficiency of self-regulated learning (Rawson, O'Neil, & Dunlosky, 2011).

Extensive research indicates that JOLs are essentially inferential in nature, meaning that individuals utilize various cues to predict their future recall performance when making JOLs (Chen & Fu, 2004). However, not all cues effectively facilitate accurate JOLs (Koriat, 1997). For instance, perceptual feature cues often lead to substantial biases in individuals' JOLs. Rhodes and Castel (2008) first revealed the influence of perceptual feature cues on JOLs. In their study, participants learned words presented in different font sizes (18pt vs. 48pt) and immediately made JOLs by assessing the likelihood of recalling each word in a subsequent test (0% ~ 100%). After studying all words, participants completed a free recall test. The results showed no significant difference in actual recall performance between large-font and small-font words, yet participants assigned higher JOLs to large-font words. In other words, although the true memory performance for large-font and small-font items was equivalent, participants erroneously believed they were more likely to successfully memorize large-font

items. This phenomenon is known as the font size effect.

The font size effect reflects a dissociation between metamemory and memory itself, representing a metacognitive illusion triggered by item perceptual features (Rhodes & Castel, 2008; Yang, Huang, & Shanks, 2018). How to effectively reduce such metacognitive illusions has been a topic of widespread concern among researchers (McCabe & Soderstrom, 2011; Soderstrom & Rhodes, 2014; Yan, Bjork, & Bjork, 2016). Therefore, this study uses the font size effect as an entry point to examine whether increasing the depth of information processing can correct JOL illusions induced by perceptual features.

To effectively correct JOL illusions based on perceptual features, the prerequisite is to clarify the internal mechanism through which perceptual features influence JOLs. Currently, the mechanism underlying the font size effect remains unclear and controversial. Some studies suggest that perceptual fluency is the primary cause of the font size effect (Bjork, Dunlosky, & Kornell, 2013; Rhodes & Castel, 2008). Perceptual fluency originates from low-level perceptual processing of stimuli and represents individuals' subjective experience of the ease or difficulty of processing perceptual information (Zhang, 2010). In previous font size effect studies, researchers have employed both subjective self-report methods and objective measures such as task response time and judgment accuracy to assess perceptual fluency. For example, Yang et al. (2018) used a continuous identification (CID) task to investigate the role of perceptual fluency in the font size effect. The study found that large-font words had shorter identification response times than small-font words, indicating higher perceptual fluency for large-font words. Bayesian mediation analysis revealed that perceptual fluency partially mediated the relationship between font size and JOLs. In contrast, Mueller, Dunlosky, Tauber, and Rhodes (2014) employed a lexical decision task and found no difference in processing fluency between large and small words, suggesting that the font size effect is unrelated to perceptual fluency.

To further examine the role of perceptual fluency in the font size effect, Yang et al. (2018) directly compared the effectiveness of the CID task and the lexical decision task in measuring item perceptual fluency. The CID task is a visual perceptual identification task in which participants press a key immediately upon identifying a presented word (Ward, Berry, & Shanks, 2013). The lexical decision task requires participants to discriminate as quickly and accurately as possible whether a string presented on the screen is a real word or a pseudoword, and to make the corresponding key response. The results showed that in the CID task, all participants identified large-font words significantly faster than small-font words. However, in the lexical decision task, there was no difference in judgment response times between large-font and small-font words. Yang et al. (2018) argued that the main reason for this measurement difference is that the CID task only involves individuals' perceptual processing of items, and the obtained continuous identification response times are not subject to potential interference from additional factors (such as semantic processing or key selection) (Chumbley & Balota, 1984; Yap, Sibley, Balota, Ratcliff, & Rueckl,

2015). These findings indicate that the CID task has higher sensitivity for measuring perceptual fluency.

Although Yang et al.'s (2018) study demonstrated that the CID task can more effectively measure differences in perceptual fluency between items and provided direct evidence for the influence of perceptual fluency on JOLs, their study still had significant design flaws. For example, the experimental design might have substantially amplified the influence of perceptual fluency on JOLs, thereby exaggerating the mediating role of perceptual fluency in the font size effect. Specifically, the study required participants to make JOLs immediately after successfully identifying the target word. Since there were significant differences in identification response times between large-font and small-font words, this immediate JOL task made the fluency differences more salient and more likely to capture individuals' attention, thereby prompting participants to rely more heavily on perceptual fluency when making JOLs (Castel, 2008). Therefore, whether perceptual fluency is the internal mechanism through which font size influences JOLs remains highly controversial. To address these limitations, this study separated the CID task from the JOL task to examine whether perceptual fluency plays a mediating role in the font size effect.

Recent research has found that people can integrate multiple cues for JOLs (Undorf & Bröder, 2019; Undorf, Söllner, & Bröder, 2018). Could the emergence of the font size effect be related to insufficient activation of deep cues other than perceptual fluency? Analysis of previous font size effect studies suggests that this metacognitive illusion may also be related to insufficient encoding strength during learning and inadequate activation of diagnostic cues. In previous studies, participants' free recall accuracy rates were relatively low (0.12 ~ 0.25). For example, Rhodes and Castel (2008) required participants to study each word for 5s and make JOLs, followed by a free recall test. In the test, individuals' free recall accuracy was only 0.16, and similar results were replicated in studies by Mueller et al. (2014) and Yang et al. (2018). It should be noted that this encoding strength issue was not caused by insufficient study time. In Rhodes and Castel's (2008) experiment, each word's presentation time was as long as 5s, but participants may not have fully utilized all the study time for continuous encoding of the words. In other words, participants' encoding time did not always equal the presentation time of the studied items. Participants might have terminated encoding processing before the studied item disappeared, or experienced attention lapses or mind wandering during the long study period (Cooper & Pantle, 1967; Mueller et al., 2014; Seli, Risko, Smilek, & Schacter, 2016).

According to the cue-utilization model, the main reason for biases in JOLs is that individuals utilize cues unrelated to true memory strength when making JOLs (Benjamin, Bjork, & Schwartz, 1998; Koriat, 1997). When encoding strength is weak, perceptual feature cues of learning materials can easily mislead JOLs, leading to the font size effect illusion. Enhancing information encoding strength can activate more and more effective cues (such as semantic cues) that

are related to individuals' true memory strength and have higher diagnosticity, thereby weakening the influence of perceptual feature cues on JOLs. To test this hypothesis, this study examined whether increasing effective study time (and requiring participants to rehearse during learning to ensure they fully utilize study time for continuous encoding of information) and increasing semantic processing depth (i.e., requiring participants to generate sentences for studied words) can effectively reduce or eliminate the font size effect.

In summary, the main objectives of this study are: (1) to examine whether perceptual fluency plays a mediating role in the font size effect; (2) to investigate whether increasing encoding strength can effectively correct JOL illusions caused by font size. To achieve these objectives, three experiments were conducted. Experiment 1 separated the CID task from the JOL task to reduce the salience of perceptual fluency, thereby examining whether perceptual fluency is one of the internal mechanisms underlying the font size effect. Experiment 2 promoted encoding strength by increasing effective study time to investigate whether this method can weaken or eliminate the font size effect illusion. Experiment 3 promoted encoding strength by enhancing semantic processing depth to examine the influence of deep cue (semantic cue) activation on the font size effect (Lupker, Harbluk, & Patrick, 1991). Exploring these issues will help reveal the internal mechanism of the font size effect and provide scientific guidance for correcting metacognitive illusions induced by perceptual features in real educational contexts.

---

## 2 Experiment 1: The Role of Perceptual Fluency in the Font Size Effect

As mentioned above, given the limitations in Yang et al.'s (2018) study, Experiment 1 implemented the CID task and JOL task separately to reduce the salience of perceptual fluency and examine whether perceptual fluency is an internal mechanism underlying the font size effect.

### 2.1 Method

**2.1.1 Participants** Twenty-six undergraduate students (11 male) with a mean age of 21.06 years ( $SD = 2.44$ ) participated in the experiment. All participants reported normal or corrected-to-normal vision and received monetary compensation after the experiment.

**2.1.2 Apparatus and Materials** The experimental program ran on a Pentium IV IBM microcomputer, and study words were presented on a 22-inch CRT color monitor with a resolution of  $1280 \times 1024$  pixels and a refresh rate of 85Hz.

**2.1.3 Materials** Sixty-four two-character words were selected from the *Modern Chinese Frequency Dictionary*, with word frequency controlled between

0.00033 and 0.00897, emotionally neutral valence, and stroke counts between 10 and 21. Eight words were used for practice trials (practice trial data were not included in the analysis), and the remaining words were used for the formal experiment. The 48 words in the formal experiment were randomly divided into two sets (a and b), with 24 study words in each set. One set was presented in 9pt (small font), and the other set in 70pt (large font). Font size and presentation order were counterbalanced across participants (Hu et al., 2015). Independent samples t-tests on the experimental materials revealed no significant differences between the two word sets in word frequency or stroke count ( $p > 0.20$ ). In addition to these words, eight other words were used as primacy and recency buffers, with buffer words presented equally often in large and small fonts, and buffer data were not included in the final analysis.

**2.1.4 Experimental Design** Experiment 1 employed a single-factor within-subjects design (font size: small vs. large). The dependent variables were identification response time, JOL values, and free recall accuracy.

**2.1.5 Procedure** The experimental program was written in E-prime 2.0. Participants sat approximately 70 cm from the computer screen with their eyes level with the center of the screen. Participants first completed eight practice trials to understand the experimental tasks.

In the formal experiment, participants completed two separate tasks: the CID task and the JOL task. As shown in Figure 1 [Figure 1: see original paper], in the CID task, a 30pt fixation point was first presented at the center of the screen for 0.5s, followed by alternating presentations of the target word (e.g., 字典) and a mask (鱗鱗) in the same font size (9pt or 70pt). Each target word had 14 identification cycles. In the first cycle, the target word was presented for 17ms and the mask for 233ms; in the second cycle, the target word was presented for 34ms and the mask for 216ms; and so on, with the target word presentation time increasing by 17ms and the mask presentation time decreasing by 17ms each cycle. This cycle continued until participants made a key response or until 14 cycles were completed. If participants identified the target word, they immediately pressed the spacebar, causing the target word and mask to disappear. Participants then verbally reported the identification result, and the computer recorded the identification response time. Target words were presented in a pseudorandom order to ensure that words of the same font size did not appear consecutively more than twice.

The JOL task was administered after participants completed the CID task and consisted of three phases: a study phase, a distractor phase, and a free recall test. Based on participants' prior completion of the CID task and pilot study results, the presentation time for study words in Experiment 1 was set to 2s. In the study phase, a single 9pt or 70pt study word was presented at the center of the screen, and participants were instructed to study the word and try to memorize it. After each study word disappeared, participants made a JOL by clicking on

a progress bar on the screen to assess the likelihood of recalling that word in a subsequent test (0 ~ 100). Study words were presented in a pseudorandom order to ensure that words of the same font size did not appear consecutively more than twice.

Following the study phase, the distractor phase began immediately. A three-digit number was randomly presented at the center of the screen, and participants performed continuous subtraction by 3, verbally reporting the results. The distractor phase lasted for 3 minutes. After the distractor phase, participants began the free recall test and had 5 minutes to recall as many studied words as possible. Participants wrote their answers on a response sheet, with no requirement for recall order.

**Footnote:** In the CID task, pilot studies initially used meaningless symbols (e.g., ##) as masking stimuli, but these symbols could not form a complete mask for Chinese words. Therefore, this study used orthographically similar characters as masking stimuli (e.g., 𩇛𩇛), where the masked study words and masking stimuli did not form semantic relationships (Yan, Zhang, Zhang, & Bai, 2013).

## 2.2 Results

In the CID task, two indices were calculated: identification accuracy and response time. Identification accuracy was calculated as: (number of correctly identified items / total number of items)  $\times$  100%. Due to the low difficulty of the CID task, participants' identification accuracy was 97.84%, showing a ceiling effect. Therefore, Experiment 1 used identification response time as the measure of perceptual fluency. Trials with incorrect identification or no key response within the time limit (i.e., during the 14 identification cycles) were first excluded, and then the average response times for correctly identified large-font and small-font words were calculated separately. The identification response times, free recall accuracy, and JOL values for different font sizes are shown in Table 1.

**Table 1** Free Recall Accuracy, JOL Values, and Identification Response Times (M  $\pm$  SD) for Different Font Sizes

Font Size	Free Recall Accuracy	JOL Value	Identification Response Time (ms)
Small	0.26 $\pm$ 0.07	48.63 $\pm$ 12.69	1864.77 $\pm$ 310.97
Large	0.27 $\pm$ 0.07	61.68 $\pm$ 13.01	1519.95 $\pm$ 342.17

Consistent with previous research results, there was no significant difference in free recall accuracy between large-font and small-font words,  $t(25) = 0.77$ ,  $p$

= 0.45 (see right side of Figure 2 [Figure 2: see original paper]), but large-font words received higher JOL values than small-font words,  $t(25) = 5.27$ ,  $p < 0.001$ , Cohen' s  $d = 1.01$  (see left side of Figure 2). These results replicate the classic font size effect.

**Figure 2** JOL Values and Free Recall Accuracy for Large-Font and Small-Font Words

**2.2.2 Identification Response Time** Participants identified large-font words significantly faster than small-font words (see Table 1),  $t(25) = 7.63$ ,  $p < 0.001$ , Cohen' s  $d = 1.05$ . To further examine the role of perceptual fluency in the font size effect, Bayesian multilevel mediation analysis was conducted using the R `bmlm` package (Vuorre, 2017), following the statistical method of Yang et al. (2018). As shown in Figure 3 [Figure 3: see original paper], the mediation analysis used font size (small font = 0, large font = 1) as the independent variable, identification response time as the mediator variable, and JOL values as the dependent variable.

**Figure 3** Mediation Model of Perceptual Fluency Between Font Size and JOL Values

The results showed that the total effect of font size on JOLs was 13.05, 95% CI = [7.89, 18.23]. The indirect effect of font size on JOLs through identification response time (perceptual fluency) was 0.79, 95% CI = [0.07, 1.55], indicating that large-font words increased JOL values by enhancing perceptual fluency. Perceptual fluency explained 6% (95% CI = [1%, 14%]) of the font size effect. The direct effect of font size on JOLs was 12.27, 95% CI = [7.12, 17.43], suggesting that perceptual fluency cannot explain the entire font size effect. In other words, after controlling for identification response time, the direct influence of font size on JOLs remained significant. Detailed mediation analysis results are shown in Table 2 .

**Table 2** Mediation Analysis Results of Perceptual Fluency Between Font Size and JOL Values

Path	Regression Coefficient ( $\beta$ )	Standard Error (SE)	95% CI
Font Size → Identi- fica- tion Re- sponse Time (a)	-344.82	45.20	[-0.41, -0.27]

Path	Regression Coefficient ( $\beta$ )	Standard Error (SE)	95% CI
Identification Time → JOL Values (b)	10.023	0.0010	[-4.32, -0.31]
Font Size → JOL Values: Total Effect (c)	13.05	2.56	[7.89, 18.23]
Font Size → JOL Values: Direct Effect (c' )	12.27	2.56	[7.12, 17.43]
Mediation Effect of Identi- fica- tion Re- sponse Time (me)	0.79	0.38	[0.07, 1.55]
Percentage of Media- tion Effect (pme)	6%	3%	[1%, 14%]

### 2.3 Discussion

Experiment 1 found that although font size had no effect on free recall performance, participants assigned higher JOL values to large-font words, demonstrating the font size effect (e.g., Rhodes & Castel, 2008; Mueller et al., 2014). Identification response times revealed that large-font words had significantly better perceptual fluency than small-font words. To reduce the salience of perceptual fluency in immediate JOLs, Experiment 1 separated the CID task from the JOL task. The results showed that the indirect effect of perceptual fluency was 0.79, 95% CI = [0.07, 1.55], which was slightly lower than the indirect effect observed by Yang et al. (2018) (0.84, 95% CI = [0.31, 1.50]). This difference may be due to sampling error, or alternatively, it may be because this study separated the CID task from the JOL task, reducing the salience of perceptual fluency. Despite this minor difference, both studies found that perceptual fluency partially mediates the font size effect, indicating that perceptual-level cues play a key role between font size and JOLs.

---

## 3 Experiment 2: The Influence of Effective Study Time on the Font Size Effect

In previous font size effect studies, participants only needed to study words within a limited time, but the research paradigm could not ensure that participants fully utilized the allotted study time to encode the learning materials. For example, when study materials were presented for longer durations, participants might experience attention lapses or prematurely terminate their study, resulting in failure to use the entire presentation time for learning (Mueller et al., 2014; Seli et al., 2016). Therefore, Experiment 2, building upon limited study time, required participants to rehearse the words presented on the screen to ensure they could fully utilize all study time for encoding processing (Lupker et al., 1991; Mulligan, Buchin, & West, 2019). Additionally, the rehearsal operation prevented participants from using extra study time for elaborate semantic processing, avoiding the influence of processing depth on experimental results. According to the cue-utilization model, individuals compare multiple cues when making JOLs and select the more diagnostic cues to utilize, thereby forming JOLs (Koriat, 1997). Therefore, Experiment 2 investigated whether increasing encoding strength through increased effective study time could weaken the misleading effect of perceptual feature cues on JOLs.

### 3.1 Method

**3.1.1 Participants** Seventy-five undergraduate students (33 male) with a mean age of 22.01 years (SD = 2.70) participated in the experiment. All participants reported normal or corrected-to-normal vision and received monetary compensation after the experiment. Participants were randomly assigned to three groups, with 25 participants in each group.

**3.1.2 Apparatus and Materials** Same as Experiment 1.

**3.1.3 Experimental Design** Experiment 2 employed a 2 (font size: small vs. large)  $\times$  3 (study time: 2s vs. 4s vs. 8s) mixed design. Font size was a within-subjects variable, and study time was a between-subjects variable. The dependent variables were free recall accuracy and JOL values.

**3.1.4 Procedure** The procedure of Experiment 2 was essentially the same as Experiment 1, but Experiment 2 did not include the CID task. Participants only completed the JOL task, which included a study phase, a distractor phase, and a free recall test. In the study phase, participants were randomly assigned to one of three study time conditions: 2s, 4s, and 8s. Participants made JOLs (0 ~ 100) after studying each word. Notably, as mentioned above, to ensure participants fully utilized the study time, Experiment 2 required participants to rehearse the words presented on the screen at a rate of once per second (a flashing asterisk of equal size appeared below the word, and participants were required to verbally rehearse according to the flashing frequency). The experimenter recorded participants' rehearsal reports to ensure compliance with experimental requirements (Lupker et al., 1991). After the study phase, participants began the distractor task (subtraction by 3) and free recall test, which were identical to Experiment 1.

## 3.2 Results and Analysis

Descriptive statistics for free recall performance and JOL values for large-font and small-font words under the three study time conditions are shown in Table 3.

**Table 3** Recall Performance and JOL Values ( $M \pm SD$ ) for Different Font Sizes Under Three Study Time Conditions

Study Time	Font Size	Free Recall Accuracy	JOL Value
2s (n = 25)	Small	0.16 $\pm$ 0.06	50.83 $\pm$ 11.93
	Large	0.15 $\pm$ 0.05	64.91 $\pm$ 13.53
4s (n = 25)	Small	0.26 $\pm$ 0.05	45.12 $\pm$ 13.02
	Large	0.27 $\pm$ 0.07	57.80 $\pm$ 17.60
8s (n = 25)	Small	0.34 $\pm$ 0.17	50.97 $\pm$ 18.97
	Large	0.34 $\pm$ 0.18	54.61 $\pm$ 17.89

**3.2.1 Free Recall Performance** A 2 (font size: small vs. large)  $\times$  3 (study time: 2s vs. 4s vs. 8s) mixed ANOVA was conducted with free recall accuracy as the dependent variable. The results showed a significant main effect of study time,  $F(2, 72) = 9.88, p < 0.001, p^2 = 0.36$ . Free recall accuracy in the 2s condition was significantly lower than in the 4s and 8s conditions ( $p < 0.001$ ), but there was no significant difference between the 4s and 8s conditions ( $p =$

0.11). Neither the main effect of font size nor the interaction between font size and study time was significant ( $p_s > 0.5$ ).

**3.2.2 JOL Values** A 2 (font size: small vs. large)  $\times$  3 (study time: 2s vs. 4s vs. 8s) mixed ANOVA was conducted with JOL values as the dependent variable. The results showed that JOL values for large-font words were significantly higher than for small-font words,  $F(1, 72) = 62.07$ ,  $p < 0.001$ ,  $p^2 = 0.46$ . The main effect of study time was not significant,  $F(2, 72) = 1.32$ ,  $p = 0.27$ . Importantly, the interaction between study time and font size was significant,  $F(2, 72) = 6.46$ ,  $p = 0.003$ ,  $p^2 = 0.15$  (see Figure 4 [Figure 4: see original paper]).

Simple effects analysis revealed that in both the 2s and 4s study time conditions, JOL values for large-font words were significantly higher than for small-font words ( $p_s < 0.001$ ). To examine the influence of increased study time on the font size effect, the difference between each participant's JOLs for large-font and small-font words was calculated as an index of the font size effect, and between-group comparisons were conducted. The results showed that the font size effect in the 2s and 4s conditions was significantly larger than in the 8s condition ( $p_1 = 0.001$ ;  $p_2 = 0.02$ ), but there was no significant difference between the 2s and 4s conditions ( $p = 0.97$ ). Notably, although increasing study time could reduce the font size effect illusion, the font size effect remained significant in the 8s condition,  $t(24) = 2.87$ ,  $p = 0.01$ , Cohen's  $d = 0.20$ .

**Figure 4** JOL Values for Different Study Times and Font Sizes

### 3.3 Discussion

The results of Experiment 2 indicate that when study time is short, perceptual features (i.e., font size) have a strong misleading effect on individuals' JOLs. However, as effective study time increases, the influence of font size on JOLs gradually weakens. Increasing effective study time may activate more cues related to memory strength for individuals to utilize in forming JOLs, thereby weakening the influence of perceptual feature cues on JOLs and resulting in a reduced font size effect. These results are consistent with the theoretical assumptions of the cue-utilization model, which posits that JOLs involve a comparative processing of multiple cues (Koriat, 1997). Notably, although increasing study time can reduce the influence of perceptual feature cues, the font size effect illusion remained significant in the 8s study time condition. Therefore, Experiment 3 will further clarify the type of activated cues by enhancing learners' encoding strength through increased processing depth, examining the influence of deep cue (semantic cue) activation on the font size effect.

## 4 Experiment 3: The Influence of Processing Depth on the Font Size Effect

Previous research has found that activating item-related contextual information can improve the accuracy of metacognitive monitoring (Soderstrom & Rhodes, 2014). Moreover, compared to manipulating study time, increasing processing depth is a more direct and effective way to manipulate encoding strength. Deep processing learning methods activate more cues, further weakening the misleading effect of perceptual feature cues on JOLs (Hertzog, Dunlosky, & Sinclair, 2010; Lupker et al., 1991). Experiment 3 clarified the type of activated cues (i.e., semantic cues) and examined their influence on the font size effect. In the deep processing condition, participants' learning method was sentence generation, which activates semantic information of items and increases encoding strength; in the shallow processing condition, no specific requirements were made regarding learning strategies (Luo & Lin, 2000).

### 4.1 Method

**4.1.1 Participants** Fifty undergraduate students (20 male) with a mean age of 22.49 years ( $SD = 2.60$ ) participated in the experiment. All participants reported normal or corrected-to-normal vision and received monetary compensation after the experiment.

**4.1.2 Apparatus and Materials** Same as Experiment 1.

**4.1.3 Experimental Design** Experiment 3 employed a  $2$  (font size: small vs. large)  $\times 2$  (processing depth: deep vs. shallow) mixed design. Font size was a within-subjects variable, and processing depth was a between-subjects variable. The dependent variables were free recall accuracy and JOL values.

**4.1.4 Procedure** In the deep processing condition, participants used sentence generation as the learning method; in the shallow processing condition, no specific requirements were made regarding learning strategies. Before the study phase, participants in the sentence generation group were instructed that a two-character word would be presented at the center of the screen and that they should generate a sentence for the word. Each word was presented for 4 seconds, and participants needed to complete the sentence generation within the presentation time and try their best to memorize the studied words. Participants in the control group were instructed that a two-character word would be presented at the center of the screen for 4 seconds and that they should try their best to memorize the studied words.

The formal experiment included a study phase, a distractor phase, and a free recall test. In the study phase, each word was presented for 4s in pseudorandom order. After studying each word, participants made JOLs on a  $0 \sim 100$  scale.

Following the study phase, participants completed the distractor task and free recall test, which were identical to Experiment 1.

## 4.2 Results and Analysis

Descriptive statistics for free recall performance and JOL values for large-font and small-font words under the two processing depth conditions are shown in Table 4.

**Table 4** Recall Performance and JOL Values ( $M \pm SD$ ) for Different Font Sizes Under Two Processing Depth Conditions

Processing Depth	Font Size	Free Recall Accuracy	JOL Value
Deep (n = 25)	Small	$0.42 \pm 0.09$	$56.42 \pm 14.53$
	Large	$0.42 \pm 0.14$	$56.96 \pm 15.16$
Shallow (n = 25)	Small	$0.19 \pm 0.08$	$45.77 \pm 15.18$
	Large	$0.20 \pm 0.08$	$60.16 \pm 15.08$

**4.2.1 Free Recall Performance** A  $2$  (font size: small vs. large)  $\times 2$  (processing depth: deep vs. shallow) mixed ANOVA was conducted with free recall accuracy as the dependent variable. The results showed a significant main effect of processing depth,  $F(1, 48) = 67.66$ ,  $p < 0.001$ ,  $\eta^2 = 0.59$ , with significantly higher free recall accuracy in the deep processing condition than in the shallow processing condition. Neither the main effect of font size nor the interaction between font size and processing depth was significant ( $ps > 0.5$ ).

An independent samples t-test was conducted to compare free recall performance between the 4s study time condition in Experiment 2 (with rehearsal) and the shallow processing condition in Experiment 3 (4s presentation time with no learning strategy requirements). The results showed that free recall performance in the shallow processing condition of Experiment 3 ( $M = 0.19$ ,  $SD = 0.07$ ) was significantly lower than in the 4s study time condition of Experiment 2 ( $M = 0.26$ ,  $SD = 0.05$ ),  $t(48) = 3.73$ ,  $p = 0.001$ , Cohen's  $d = 0.87$ . These results indicate that under conditions of long presentation time, if participants are not required to rehearse, encoding strength is insufficient, and learning is accompanied by distraction, attention lapses, or premature termination of encoding processing. This finding demonstrates the effectiveness of the rehearsal manipulation in Experiment 2.

**4.2.2 JOL Values** A  $2$  (font size: small vs. large)  $\times 2$  (processing depth: deep vs. shallow) mixed ANOVA was conducted with JOL values as the dependent

variable. The results showed a significant main effect of font size,  $F(1, 48) = 29.20$ ,  $p < 0.001$ ,  $p^2 = 0.38$ , with JOL values for large-font words significantly higher than for small-font words. The main effect of processing depth was not significant,  $F(1, 48) = 0.87$ ,  $p = 0.36$ . The interaction between processing depth and font size was significant,  $F(1, 48) = 25.14$ ,  $p < 0.001$ ,  $p^2 = 0.34$  (see Figure 5 [Figure 5: see original paper]). Simple effects analysis revealed that in the shallow processing condition, JOL values for large-font words were significantly higher than for small-font words ( $p < 0.001$ ), demonstrating the font size effect. However, in the deep processing (sentence generation) condition, there was no significant difference in JOL values between large-font and small-font words ( $p = 0.61$ ), indicating that the font size effect disappeared.

**Figure 5** JOL Values for Different Processing Depths and Font Sizes

### 4.3 Discussion

The results of Experiment 3 are consistent with the research hypothesis: as processing depth increased, the font size effect induced by perceptual features disappeared. In shallow processing, individuals' JOLs relied on perceptual feature cues, but these cues were misleading and detrimental to accurate metacognitive monitoring. In deep processing, sentence generation activated semantic information cues of words and weakened the misleading effect of perceptual features on JOLs. Previous research found that activating information beyond the item itself could reduce metacognitive monitoring illusions but not eliminate them completely (McCabe & Soderstrom, 2011; Soderstrom & Rhodes, 2014). The inconsistency with previous research may be due to the fact that this study explicitly activated semantic information, whereas previous studies did not distinguish the specific type of activated information. The type and specificity of activated information may be key factors in reducing or eliminating metacognitive illusions.

---

## 5 General Discussion

This study examined the influence of encoding strength on the font size effect, exploring both the internal mechanism underlying the effect and effective corrective measures. Experiment 1 used a CID task to investigate the role of perceptual fluency in the font size effect. The results revealed that large-font words were identified significantly faster than small-font words. Further Bayesian multilevel mediation analysis found that perceptual fluency (identification response time) partially mediated the relationship between font size and JOL values. This result reveals that perceptual fluency forms a partial basis for the font size effect and demonstrates the role of perceptual-level cues between font size and JOLs, providing direct supporting evidence for the fluency hypothesis.

Experiment 2 found that the influence of font size on JOLs weakened as effective study time increased, indicating that longer effective study time can reduce

the font size effect illusion. The main reason is that as effective study time increases, individuals seek and construct more cues beyond perceptual features and utilize these cues, thereby reducing reliance on perceptual feature cues for JOLs. However, it is worth emphasizing that although the font size effect was reduced in the 8s study time condition, it remained significant. Experiment 3 further clarified the type of activated cues (i.e., semantic cues) to correct the font size effect. The results showed that in the shallow processing condition, JOLs remained susceptible to misleading by font size, demonstrating the font size effect. However, in the deep processing (sentence generation) condition, this metacognitive illusion was corrected, and the font size effect disappeared.

### 5.1 Processing Fluency and the Font Size Effect

The results of this study indicate that perceptual fluency forms a partial basis for the font size effect. Experiment 1 replicated the findings of Yang et al. (2018), showing that individuals' JOL values increase with increased perceptual fluency. This result is consistent with previous research demonstrating that perceptual fluency has a universal influence on JOLs in both visual and auditory domains (Besken & Mulligan, 2014; Rhodes & Castel, 2009). Moreover, researchers have found that the influence of perceptual fluency on JOLs is not limited to laboratory learning materials (e.g., words, Yang et al., 2018; word pairs, Rhodes & Castel, 2008; pictures, Undorf, Zimdahl, & Bernstein, 2017) but also significantly affects JOLs for real educational materials (Ball, Klein, & Brewer, 2014). Thus, using different font sizes to present learning content, while serving to emphasize, highlight, or attract attention to the content, simultaneously creates different processing fluency experiences for students, leading to metacognitive illusions (Carpenter, Mickes, Rahman, & Fernandez, 2016; Carpenter, Wilford, Kornell, & Mullaney, 2013).

Compared to other types of perceptual cues (e.g., text clarity and resolution, Strukelj, Scheiter, Nyström, & Holmqvist, 2016; font type and presentation format, Rummer, Schweppe, & Schwede, 2016), font size cues appear more frequently in real educational contexts (Magreehan, Serra, Schwartz, & Narciss, 2016). Therefore, the potential impact of metacognitive monitoring illusions triggered by font size on self-regulated learning deserves greater attention. For example, in educational material design, educators should consider not only using different font sizes to emphasize key knowledge points but also how to avoid JOL illusions caused by font size. The results of this study indicate that promoting deeper semantic processing is an effective measure to eliminate the font size effect illusion. Therefore, educators could consider adding test questions (such as sentence generation or example nomination) after knowledge content presented in different font sizes to promote deeper semantic processing and activate more semantic cues, thereby reducing or eliminating JOL errors caused by perceptual features.

## 5.2 The Influence of Encoding Strength on the Font Size Effect

Experiment 2 found that by increasing effective study time, individuals could seek or construct more cues beyond perceptual features, weakening the influence of perceptual feature cues on JOLs and thereby reducing the font size effect. This finding is similar to the results of Magreehan et al. (2016). In Experiment 3 of Magreehan et al. (2016), when participants were required to learn mixed lists composed of related and unrelated word pairs, semantic relatedness of word pairs had a significant influence on JOLs (related word pairs received higher JOLs), but the contrast between word pairs and background had no effect on JOLs. However, when participants learned pure lists composed only of unrelated word pairs, contrast significantly influenced JOLs (high-contrast word pairs received higher JOLs). These results indicate that when more valuable cues (e.g., semantic relatedness of word pairs) are available, people's JOLs rely less on perceptual feature cues (e.g., contrast level).

Experiment 3 found that the font size effect disappeared under deep processing (sentence generation) conditions, demonstrating that increasing processing depth is an effective measure for correcting metacognitive illusions induced by perceptual features. The main reason for this phenomenon is that requiring participants to generate sentences in Experiment 3 likely activated more semantic cues of words (Belmore, 1981; Hertzog, Fulton, Sinclair, & Dunlosky, 2014). Since there is no real difference in semantic information between large-font and small-font words, participants in the sentence generation condition may have realized that the memorability of words in different font sizes did not actually differ, thereby weakening their reliance on font size cues during JOLs and correcting the metacognitive illusion caused by perceptual features.

The results of this study are consistent with Soderstrom and Rhodes (2014), who found that having individuals make judgments of remembering and knowing (JORKs) could effectively reduce metacognitive illusions. In the study phase of their research, participants needed to evaluate whether they would remember detailed information about an item in a subsequent test and make a 3-point rating (1 = remember; 2 = familiar; 3 = forget). Compared to JOLs, JORKs monitor whether individuals remember detailed information internally constructed about the material, rather than simply monitoring the memory level of the learning material itself. Soderstrom argued that the JORKs monitoring process activates more feature information related to the item, and this activation is the reason for reducing metacognitive monitoring illusions (McCabe & Soderstrom, 2011). It is worth emphasizing that although both of Soderstrom's studies found that JORKs can reduce metacognitive illusions, JORKs far from eliminate them. This may be because those studies did not further clarify the type of activated information (perceptual feature information or semantic information) or the activation was insufficient. Experiment 3 supplemented this by finding that sentence generation processing can successfully activate semantic cues of words and effectively eliminate metacognitive illusions.

Undorf et al. (2018) found that JOLs do not rely on a single cue. When multiple cues are presented simultaneously (e.g., font size, study frequency, emotional attributes, and concreteness of studied words), these cues undergo cue integration and competition in individuals' JOLs. Specifically, the study found that under concrete word conditions, emotional words received significantly higher JOLs than neutral words, demonstrating the emotionality effect. Under abstract word conditions, although this emotionality effect still existed, it was significantly reduced. In other words, the concreteness cue of studied words weakened the influence of emotional attribute cues on JOLs. This study provides supporting evidence for the view that multiple cues undergo integration and competition in JOLs. Both Experiment 2 and Experiment 3 found that when participants engaged in deeper encoding processing (long-time rehearsal and sentence generation), more cues were activated (e.g., semantic cues), and participants relied less on perceptual feature cues (i.e., font size) when making JOLs, thereby weakening the font size effect.

Neither Experiment 2 nor Experiment 3 of this study found an effect of encoding strength on JOLs. The reason for this phenomenon may be that this study manipulated encoding strength as a between-subjects variable. According to Koriat's (1997) cue-utilization model, the relative weight of different cues on JOLs is influenced by the format of the study list. When external cues (e.g., study time) are manipulated within-subjects, these cues not only influence learners' learning outcomes but also their metacognitive judgments. However, when these cues are manipulated between-subjects, they only influence learning outcomes but not metacognitive judgments. Within-subjects manipulation creates comparisons between different external cues, thereby amplifying the potential role of external cues and prompting individuals to utilize these cues for JOLs. Conversely, JOLs are not sensitive to external cues manipulated between-subjects. This phenomenon has been repeatedly confirmed in numerous studies (Susser, Mulligan, & Besken, 2013).

In summary, the results of this study indicate that as encoding strength increases, the font size effect illusion gradually weakens. During shallow encoding, font size has a strong misleading effect on JOLs. During deep encoding, more cues are activated, which weakens the influence of font size on JOLs and corrects metacognitive illusions induced by perceptual features.

### 5.3 Limitations and Future Directions

This study examined the influence of encoding strength on the font size effect, enriching previous explanations of the font size effect, providing corrective measures for metacognitive illusions induced by perceptual features, and deepening understanding of metacognitive illusions. However, this study still has some limitations, and future research could explore the following directions:

First, this study revealed that increasing processing depth is an effective measure for correcting the font size effect, but it did not further reveal the underlying

correction mechanism. To date, there is considerable controversy regarding the mechanism underlying the font size effect, with two core perspectives: processing fluency and beliefs. For example, the research by Yang et al. (2018) and this study both found that perceptual fluency is one of the internal mechanisms of the font size effect. However, Mueller et al. (2014) argued that the font size effect arises from people's beliefs (i.e., the belief that large-font words are easier to memorize than small-font words), and the process through which beliefs influence JOLs is often analytical and consciously mediated (Ball et al., 2014). Future research could explore research paradigms better suited to disentangling perceptual fluency and beliefs, isolating their respective roles in the font size effect and examining how processing depth influences each.

Second, future research could use eye-tracking technology to explore the patterns of how encoding strength corrects metacognitive illusions. Although this study confirmed that the font size effect gradually weakens with increased encoding strength and preliminarily found that the font size effect decreases correspondingly with increased effective study time, it did not systematically examine the changing trajectory of JOL values for large-font and small-font words using a time-course analysis. Eye-tracking technology can serve as an important tool for analyzing learning processes, and changes in pupil diameter can reflect encoding strength during learning (Bradley & Lang, 2015; Gog & Scheiter, 2010). Future research could further investigate the dynamic changes in the correction process using eye-tracking technology.

Finally, this study only discussed the corrective effect of increased encoding strength on the font size effect illusion under single-cue (i.e., font size) conditions. Although, compared to other perceptual cues (e.g., font blur), font size cues appear more frequently in learning materials, this single-cue condition still lacks ecological validity for real classrooms or learning processes. In real classroom contexts, students' JOLs may be driven by multiple cues (Toftness et al., 2018). Therefore, future research should examine the interactive effects of multiple cues on metacognitive monitoring and corrective measures for metacognitive illusions in real classroom contexts.

---

## 6 Conclusion

This study investigated the influence of encoding strength on the font size effect through three experiments. The conclusions are as follows:

1. The font size effect weakens or disappears as encoding strength increases. During shallow encoding, font size has a strong misleading influence on JOLs, and perceptual fluency is one of the internal mechanisms underlying the font size effect.
2. As encoding strength increases, more cues are activated, and JOLs involve comparative weighting of multiple cues. When other diagnostic cues related to memory become available, the misleading effect of perceptual

feature cues on JOLs is weakened. Therefore, increasing processing depth is an effective measure for correcting metacognitive illusions induced by perceptual features.

---

## References

- Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied*, 20(4), 336-348.
- Belmore, S. M. (1981). Imagery and semantic elaboration in hypermnnesia for words. *Journal of Experimental Psychology: Human Learning & Memory*, 7(3), 191-203.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, 127(1), 55-68.
- Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: Analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 429-440.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, 64(1), 417-444.
- Bradley, M. M., & Lang, P. J. (2015). Memory, emotion, and pupil diameter: Repetition of natural scenes. *Psychophysiology*, 52(9), 1186-1193.
- Carpenter, S. K., Mickes, L., Rahman, S., & Fernandez, C. (2016). The effect of instructor fluency on students' perceptions of instructors, confidence in learning, and actual learning. *Journal of Experimental Psychology: Applied*, 22(2), 161-172.
- Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, 20(6), 1350-1356.
- Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, 36(2), 429-437.
- Chen, G. X., & Fu, X. L. (2004). Judgment of learning and its accuracy. *Advances in Psychological Science*, 12(2), 176-184. [陈功香, 傅小兰. (2004). 学习判断及其准确性. 心理科学进展, 12(2), 176-184.]

- Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, 12(6), 590-606.
- Cooper, E. C., & Pantle, A. J. (1967). The total time hypothesis in verbal learning. *Psychological Bulletin*, 68(4), 221-234.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Gog, T. V., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction*, 20(2), 95-99.
- Hertzog, C., Dunlosky, J., & Sinclair, S. M. (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, 38(6), 771-784.
- Hertzog, C., Fulton, E. K., Sinclair, S. M., & Dunlosky, J. (2014). Recalled aspects of original encoding strategies influence episodic feelings of knowing. *Memory & Cognition*, 42(1), 126-140.
- 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.
- 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.
- Luo, J., & Lin, Z. X. (2000). Monitoring of multiple memory systems: The influence of LOP and metamemory training. *Acta Psychologica Sinica*, 32(1), 25-29. [罗劲, 林仲贤. (2000). 加工深度和元记忆训练对多重记忆系统监测的影响. *心理学报*, 32(1), 25-29.]
- Lupker, S. J., Harbluk, J. L., & Patrick, A. S. (1991). Memory for things forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 897-907.
- Magreehan, D. A., Serra, M. J., Schwartz, N. H., & Narciss, S. (2016). Further boundary conditions for the effects of perceptual disfluency on judgments of learning. *Metacognition and Learning*, 11(1), 35-56.
- McCabe, D. P., & Soderstrom, N. C. (2011). Recollection-based prospective metamemory judgments are more accurate than those based on confidence: Judgments of remembering and knowing (JORKs). *Journal of Experimental Psychology: General*, 140(4), 605-621.
- 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.
- Mulligan, N. W., Buchin, Z. L., & West, J. T. (2019). Assessing why the testing effect is moderated by experimental design. *Journal of Experimental Psychology:*

*Learning, Memory, and Cognition*. Advance online publication.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5(4), 207-213.

Rawson, K. A., O'Neil, R., & Dunlosky, J. (2011). Accurate monitoring leads to effective control and greater learning of patient education materials. *Journal of Experimental Psychology: Applied*, 17(3), 288-302.

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.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16(3), 550-554.

Rummer, R., Schweppe, J., & Schwede, A. (2016). Fortune is fickle: Null-effects of disfluency on learning outcomes. *Metacognition and Learning*, 11(1), 57-70.

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in Cognitive Sciences*, 20(8), 605-617.

Soderstrom, N. C., & Rhodes, M. G. (2014). Metacognitive illusions can be reduced by monitoring recollection during study. *Journal of Cognitive Psychology*, 26(1), 118-126.

Strukelj, A., Scheiter, K., Nyström, M., & Holmqvist, K. (2016). Exploring the lack of a disfluency effect: Evidence from eye movements. *Metacognition and Learning*, 11(1), 71-88.

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, 41(7), 1000-1011.

Tauber, S., & Dunlosky, J. (2016). A brief history of metamemory research and handbook overview. In S. Tauber & J. Dunlosky (Eds.), *The Oxford handbook of metamemory* (pp. 7-22). New York: Oxford University Press.

Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*, 13(1), 1-14.

Undorf, M., & Bröder, A. (2019). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, 73(4), 629-642.

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory & Cognition*, 46(4), 507-519.

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.

- Vuorre, M. (2017). bmlm: Bayesian multilevel mediation. R package version 1.3.0. Retrieved from <https://cran.r-project.org/package=bmlm>.
- Ward, E. V., Berry, C. J., & Shanks, D. R. (2013). An effect of age on implicit memory that is not due to explicit contamination: Implications for single and multiple-systems theories. *Psychology and Aging*, 28(2), 429–442.
- Yan, G. L., Zhang, Q. M., Zhang, L. L., & Bai, X. J. (2013). The effect of masking materials on perceptual span in Chinese reading. *Journal of Psychological Science*, 36(6), 1317–1322. [闫国利, 张巧明, 张兰兰, 白学军. (2013). 不同掩蔽材料对阅读知觉广度的影响. *心理科学*, 36(6), 1317–1322.]
- Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General*, 145(7), 918–933.
- Yang, C., Huang, S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language*, 99, 99–110.
- Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English lexicon project. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3), 597–613.
- Zhang, X. J. (2010). Effects of perceptual fluency on judgment and decision. *Advances in Psychological Science*, 18(4), 639–645. [张旭锦. (2010). 知觉流畅性对判断和决策的影响. *心理科学进展*, 18(4), 639–645.]

---

### **Influence of encoding strength on the font size effect**

**ZHAO Wenbo; JIANG Yingjie; WANG Zhiwei; HU Jingyuan**

(School of Psychology, Northeast Normal University, Changchun 130024, China)

### **Abstract**

Judgments of learning (JOLs) are characterised as metacognitive judgments regarding the likelihood that studied items can be successfully retrieved in a future memory test. Previous studies found that people employ different types of cues to inform their online JOLs. Some of these cues can guide JOLs to accurately reflect memory status but others cannot (and are even misleading in some situations). A widely studied cue for JOL formation is subjective processing experience (e.g., perceptual fluency) while completing a given task, which often confers metacognitive illusions. It has been found that people give higher JOLs to large than to small words, despite the fact that font size has minimal influence on retention, a phenomenon termed the font size effect on JOLs. A potential mechanism underlying the effect is perceptual fluency: Large words

are perceived more fluently than small ones, and fluent processing experience of large words induces a feeling of knowing, which drives people to offer higher JOLs. The font size effect is important because it spotlights a dissociation between metacognitive judgments and memory itself. The current study aims to explore the influences of encoding strength on the font size effect, and to explore practical techniques to calibrate metacognitive illusions induced by perceptual fluency.

Experiment 1 aimed to delineate the role of perceptual fluency in the font size effect. Twenty-six participants first completed a continuous identification (CID) task to measure the difference in perceptual fluency (indexed by response times; RTs) between large (70-pt) and small (9-pt) words, after which they attended a classic learning task. In the learning task, participants studied large and small words one-by-one, for 2s each, and made item-by-item JOLs. Immediately following the learning task, they completed a distractor task, followed by a free recall test. The results showed that, in the CID task, participants responded much faster to large than to small words, indicating the natural difference in perceptual fluency between large and small words. In addition, perceptual fluency (i.e., RTs in the CID task) significantly correlated with JOLs, reflecting the fluency effect on JOLs. More importantly, perceptual fluency significantly mediated the font size effect on JOLs, supporting the claim that perceptual fluency is responsible for the font size effect.

Experiment 2 manipulated study durations to investigate the influence of enhancing encoding strength (through prolonging study duration) on the font size effect. Specifically, three groups of participants studied each word for 2s, 4s, and 8s, respectively, and made item-by-item JOLs. The results demonstrated that prolonging study duration correspondingly decreased the font size effect on JOLs. It is, however, worth highlighting that expanding study time cannot fully eliminate the font size effect because the results still showed a significant font size effect even when the study time was increased to 8s.

Experiment 3 was conducted to further investigate the effectiveness of enhancing encoding strength for calibration of the font size effect. A sentence-making group was instructed to encode each word by generating a sentence to deepen the level of processing (i.e., encoding strength). By contrast, there were no explicit requirements of encoding strategies in the control group (i.e., participants in the control group could use any strategies they liked). In the control group, the classic font size effect on JOLs was successfully replicated; of critical interest, the effect disappeared in the sentence-making group. Such results reflect the power of improving encoding strength to calibrate metacognitive illusions induced by perceptual features.

In summary, the current study establishes that perceptual fluency is at least one of the mechanisms underlying the font size effect on JOLs; prolonging study duration reduces but fails to eliminate the font size effect on JOLs; more importantly, directly deepening the level of processing through sentence-making is a valid strategy to calibrate metacognitive illusions induced by perceptual

features. Theoretical and practical implications are discussed in the main text.

**Key words:** judgments of learning; the font size effect; encoding strength; metacognitive illusions

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

*Source: ChinaXiv –Machine translation. Verify with original.*