Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–14
© Experimental Psychology Society 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1747021819879416
qjep.sagepub.com
SAGE

Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning

Shana K Carpenter[1] and Jason Geller[2]

Abstract

Previous research shows that participants are overconfident in their ability to learn foreign language vocabulary from pictures compared with English translations. The current study explored whether this tendency is due to processing fluency or beliefs about learning. Using self-paced study of Swahili words paired with either picture cues or English translation cues, picture cues garnered higher confidence judgements but not faster study times, and this was true whether judgements of learning were made after a delay (Experiment 1) or immediately (Experiment 2). In Experiment 3, when participants learned Swahili words with only one type of cue (pictures or English translations) and then estimated which one would be more effective for learning, the majority of participants believed pictures would be more effective regardless of whether they had experienced those cues during learning. Experiment 4 showed the same results when participants had experienced neither type of cue during a learning phase. These results suggest that metacognitive judgements in foreign language vocabulary learning are driven more by students' beliefs about learning than by processing fluency as reflected in self-paced study times.

Keywords

Overconfidence; metacognition; processing fluency; analytic processing; foreign language learning

Received: 20 December 2018; revised: 15 May 2019; accepted: 19 July 2019

People are typically overconfident in what they know. Decades of experimental research on metacognition show that people often over-predict their own performance on simple tasks measuring memory and skill acquisition (e.g., Dunlosky & Tauber, 2016). In academic settings, it is also common for students to overestimate how well they will perform on exams (Foster, Was, Dunlosky, & Isaacson, 2017; Hacker, Bol, Horgan, & Rakow, 2000; Hartwig & Dunlosky, 2017; Miller & Geraci, 2011). As accurate metacognitive judgements underlie effective regulation of one's own learning, understanding the factors influencing these judgements—and in particular, the factors that *mislead* these judgements—is critical for optimising learning.

Judgements about one's own learning can be misled by factors believed to reflect fluency, or the perceived ease with which information is learned. Things that make something *appear* easier to learn often have the effect of increasing confidence in learning, without increasing actual learning (for reviews, see Alter & Oppenheimer, 2009; Finn & Tauber, 2015; Reber & Greifeneder, 2017; Unkelbach & Greifeneder, 2013). This "metacognitive illusion" (Rhodes & Castel, 2009) has been shown for visual stimuli that are presented in larger font compared with smaller font (Kornell, Rhodes, Castel, & Tauber, 2011; Rhodes & Castel, 2008), in clearer font compared with blurred font (Yue, Castel, & Bjork, 2013), or in type-print

[1]Department of Psychology, Iowa State University, Ames, IA, USA
[2]The University of Iowa, Iowa City, IA, USA

Corresponding author:
Shana K Carpenter, Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011, USA.
Email: shacarp@iastate.edu

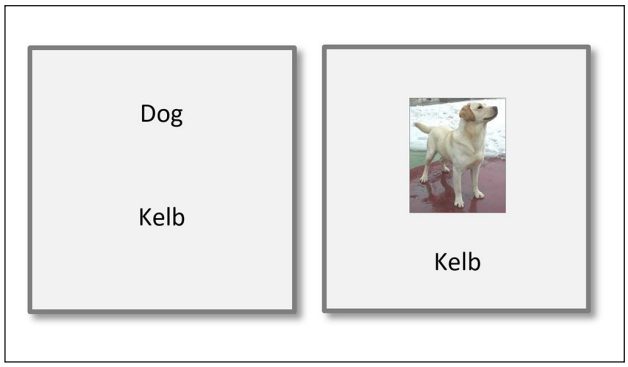

Figure 1. Example stimuli from the study by Carpenter and Olson (2012). Participants learned Swahili words either paired with their English translations (left) or paired with simple pictures to convey the English translations (right).

font compared with handwritten cursive (Geller, Still, Dark, & Carpenter, 2018). In other domains, overconfidence is greater for auditory stimuli that are presented in louder compared with softer volume (Rhodes & Castel, 2009), and for lectures that are delivered in a manner that is smooth and well-polished compared with fumbling and awkward (Carpenter, Mickes, Rahman, & Fernandez, 2016; Carpenter, Northern, Tauber, & Toftness, in press; Carpenter, Wilford, Kornell, & Mullaney, 2013; Toftness et al., 2018).

The presence of colourful images and graphics, which might increase the appearance of fluency, can also mislead metacognitive judgements. Participants judge their own learning of a scientific text to be higher when that text is accompanied by photos, even though the presence of these photos does not affect actual learning of the text (Serra & Dunlosky, 2010). Similarly, Lenzner, Schnotz, and Mueller (2013) found that decorative pictures inserted within a scientific text led to students' perceptions that the text was easier to learn, but made no difference in actual learning of the text. Thus, pictures can produce illusions of understanding that may not coincide with actual learning of text materials (see also Wiley, Sarmento, Griffin, & Hinze, 2017).

A similar type of illusion has been demonstrated with the use of pictures in foreign language vocabulary learning. Although pictures are commonly used in foreign language instruction to denote the meanings of words, experimental research shows that these pictures may not promote learning and could even mislead students' judgements of learning (JOLs). Carpenter and Olson (2012, Experiment 2) had participants learn new Swahili words in the context of either picture cues or English translation cues (see Figure 1). After encoding a number of pairs in this way, participants completed a cued recall test in which they were given the original cue that was paired with the Swahili word—either the picture or the English translation—and had to produce the Swahili word from that cue. Performance on this test did not differ according to the type of cue (for similar results, see Chen, 1990; Lotto & de Groot, 1998) indicating, somewhat surprisingly, that pictures do not provide more effective cues for remembering new words in a foreign language. However, Carpenter and Olson also collected JOLs during initial encoding, and found that participants were more confident in their ability to remember Swahili words from pictures compared with English translations. The effects of pictures on confidence, but not on actual learning, suggest that they may provide an intuitive but misleading cue for metacognitive judgements.

Presently, it is unknown what kind of mechanism underlies the misleading effects of pictures on foreign language vocabulary learning. Carpenter and Olson (2012) reasoned that this effect was due to processing fluency, in that Swahili words are perceived as easier to process in the context of picture cues compared with English translation cues. The intuitive assumption that pictures would make learning easier, the authors reasoned, had the effect of undermining effective processing of the Swahili words, leading to no observable benefit of picture cues on an initial test. These findings appear to be consistent with others in the literature showing that stimuli that are presumed to be more fluent or easier to process—for example, by appearing in larger font (Kornell et al., 2011; Rhodes & Castel, 2008), clearer font (Yue et al., 2013), or louder volume (Rhodes & Castel, 2009)—are processed more fluently during learning and as a consequence receive higher JOLs.

Processing fluency is a highly intuitive theory. However, it has seldom been directly tested. Doing so requires obtaining online measurements of the ease with which participants process information during study, and whether those measurements map onto metacognitive judgements. Fortunately, recent studies have begun incorporating objective measures to examine online processing fluency. One of them entails allowing participants to pace themselves through the learning materials, with the assumption that faster self-paced study times reflect more fluent processing. Using this approach, some studies have found no influence of self-paced study times on metacognitive effects that have traditionally been attributed to processing fluency, such as font size (Mueller, Dunlosky, Tauber, & Rhodes, 2014), word concreteness (Witherby & Tauber, 2017), and word pair relatedness (Mueller, Dunlosky, & Tauber, 2016; Mueller, Tauber, & Dunlosky, 2013). Others, however, have found that self-paced study times do underlie metacognitive judgements associated with word pair relatedness (Undorf & Erdfelder, 2015) and the bolding of text materials (Ball, Klein, & Brewer, 2014).

These findings indicate that manipulations traditionally believed to influence processing fluency may not actually reflect processing fluency. Thus, an alternative theory recently proposed is that beliefs about learning are more likely to underlie metacognitive judgements associated with factors such as font size (Mueller et al., 2014; Undorf & Zimdahl, 2019), concreteness (Witherby & Tauber,

2017), and word pair relatedness (Mueller et al., 2016; Mueller et al., 2013). This view proposes that people develop naïve beliefs about their own learning that influence their metacognitive judgements—for example, the idea that information is better learned if it is presented in a way that is more salient through increasing the size, volume, and so on of the presented information—independently of whether or not these materials are processed more fluently during study. In support of this account, some studies have found that participants' JOLs do not depend on a direct experience with the stimuli, and as such could not be influenced by experienced fluency of processing. For example, giving participants a hypothetical description of a manipulation (e.g., font size) and asking them to predict its effects on memory produces the usual font size effect on JOLs (Kornell et al., 2011; Mueller et al., 2014), indicating that these judgements are likely more attributable to beliefs about learning than to experienced fluency of processing.

The misleading effects of pictures in foreign language vocabulary could be due to processing fluency (Carpenter & Olson, 2012). However, the limited research on this effect so far has not directly investigated the contributions of processing fluency and beliefs about learning. Distinguishing between these accounts is important both for theoretical development of the origin of metacognitive biases and for the practical objective of mitigating overconfidence in foreign language vocabulary learning.

The current experiments set out to measure the contributions of processing fluency and beliefs about learning in participants' metacognitive judgements about the effectiveness of pictures in foreign language vocabulary learning. Experiments 1 and 2 used the same basic design as Carpenter and Olson's (2012, Experiment 2) study, but added new features to directly measure the contribution of processing fluency. Using the same learning materials, Experiments 3 and 4 employed new learning tasks designed to measure the degree to which these metacognitive judgements are driven by participants' beliefs about learning.

Experiment 1

In Carpenter and Olson's (2012) study, Swahili words were presented for a fixed rate (6 s) with either pictures or English translations as cues. After the entire list was presented, the list was presented again and participants made a JOL for each item. All items were then tested, with each cue (the picture or the English translation) shown and participants being asked to remember the Swahili word that had been paired with it. Carpenter and Olson observed that performance on the memory test did not differ between the two cue types. However, participants were initially more confident in their learning from picture cues compared with English translation cues. After completing a test with feedback, however, participants' overconfidence was greatly reduced, and subsequent cued recall of Swahili words was in fact better from picture cues than from English translation cues. The conclusion was that the greater processing fluency associated with picture cues during encoding led to a false sense of confidence that picture cues would be more effective, undermining effective processing of those items and obscuring a true advantage for picture cues that was only apparent after the bias was removed via a test opportunity.

Experiment 1 set out to test whether processing fluency during encoding does indeed underlie metacognitive judgements in foreign language vocabulary learning. We employed the same basic design as in Carpenter and Olson's (2012, Experiment 2) study, with one critical modification. Instead of presenting the items for 6 s each during encoding, participants were permitted to study each item for as long as they wished. If picture cues are processed more fluently than English translation cues, then self-paced study times should be faster, on average, for Swahili words presented with picture cues compared with English translation cues.

Participants

Twenty-five undergraduate students participated to fulfil partial course requirements for introductory psychology courses at Iowa State University. The sample size was chosen on the basis of Carpenter and Olson's (2012) Experiment 2, which included 24 participants. Participants completed the experiment individually on personal computers.

Materials

The same 42 English–Swahili word pairs from Carpenter and Olson's (2012) Experiment 2 were used. The English words were one-syllable nouns, ranging between three and six letters, with an average concreteness rating of 4.86 on a 5-point scale ($SD = 0.16$) (Brysbaert, Warriner, & Kuperman, 2014), and an average frequency of 106.52 per million ($SD = 113.40$) (Brysbaert & New, 2009). Their corresponding pictures were simple colour photographs (see Figure 1). The experimental stimuli, along with the raw and aggregate data for each experiment, can be found at: https://osf.io/st4dn/. Each participant learned a unique, randomly determined set of 21 Swahili words with picture cues, and 21 different Swahili words with English translation cues.

Design and procedure

Participants began the experiment with instructions informing them that they would be learning Swahili words paired with either pictures or English translations as cues. To illustrate each type of cue, they were given an example of an item (*Train: Reli*) that was not included among the 42 experimental items. They were informed that each pair of items

(English–Swahili pairs or picture–Swahili pairs) would be presented one at a time, and they would have as much time as they needed to study it. Participants were encouraged to do their best to learn each pair, and to encourage full and meaningful processing of each, they were instructed to press the spacebar once they felt they had fully "digested" it.

For each participant, 21 items were randomly selected to be presented as English–Swahili pairs, and 21 as picture–Swahili pairs. Participants saw each stimulus pair one at a time, in a unique random order with English–Swahili pairs and picture–Swahili pairs intermixed. Each pair was presented in the centre of the computer screen and remained on screen until participants pressed the spacebar to move on to the next pair.

After each of the 42 pairs was presented for self-paced study in this way, the same pairs were presented again for JOLs. During a JOL trial, each cue–target pair was presented on the screen and participants were asked to estimate—using a scale from 0% (definitely will NOT recall) to 100% (definitely will recall)—the likelihood of recalling the Swahili word from its cue (either the picture or English translation) after about 5 min. Participants entered a value between 0 and 100 and pressed the ENTER key to advance to the next item.

After making a JOL for each item, participants were tested over each Swahili word by being given the original cue and asked to type in the Swahili word that had been paired with it. Participants had as much time as they needed to enter their responses, and as soon as they typed their response and pressed ENTER, the correct Swahili word was displayed. With the correct answer displayed, participants were asked to again make a JOL about the likelihood of recalling that Swahili word from its cue again after about 5 min. After entering the JOL, the next item was tested, followed by the JOL for that item, and so on, until all 42 items had been tested and presented for JOLs in this way.

The same items were then presented again for the final test. The picture or English translation was again shown for each item, and participants were asked to type in the Swahili word that had been paired with it. This time, however, no feedback was provided and no JOL was made. Each time the list of items was presented (for original encoding, the initial test, and the final test), it was presented in a unique random order for each participant. The final test phase marked the end of the experiment. Upon completing the experiment, participants were debriefed and asked whether they were familiar with any of the Swahili words prior to participating in the study.

Results and discussion

Pre-analyses and scoring. Data from one participant were discarded due to failure to follow instructions (i.e., no responses were entered during the JOL or test phases), and from one additional participant whose computer malfunctioned during the experiment. Data from the remaining 23 participants were analysed.

Responses were hand-scored by two independent coders who awarded full credit for participants' responses that were exact matches to the correct Swahili words, and half credit for any responses that contained minor spelling errors. Half credit was assigned in cases where both coders agreed that half credit should be assigned.

JOLs. Because we were primarily interested in the effects of cue type on metacognitive judgements, we first report participants' JOLs associated with picture cues versus English translation cues across the initial test (referred to here as Test 1) and final test (referred to as Test 2). Examination of individual data revealed that participants' JOLs were not normally distributed. Significant departures from normality, according to Shapiro–Wilk tests (Shapiro & Wilk, 1965), occurred for the majority of participants' JOLs associated with picture cues (78% of participants) and English translation cues (87% of participants) during both encoding and at Test 1. We thus compared participants' median JOLs for picture versus English translation cues.

Table 1 reports average median JOLs for each cue type across encoding and Test 1. These data were analysed using a 2×2 repeated-measures analysis of variance (ANOVA), with cue type (picture vs. English translation) and time of JOL (encoding vs. Test 1) as factors. Results revealed a significant main effect of cue, $F(1, 22) = 7.33$, $p = .013$, $\eta_p^2 = .25$, but no main effect of time, $p = .84$, and no interaction, $p = .99$. Thus, overall participants gave higher JOLs to Swahili words associated with pictures compared with English translations.[1]

Test accuracy. Figure 2 shows cued recall performance for Swahili words from pictures versus English translations at Test 1 and Test 2. A 2×2 repeated-measures ANOVA with cue type (picture vs. English translation) and test (Test 1 vs. Test 2) as factors revealed a significant main effect for cue type, $F(1, 22) = 6.218$, $p = .021$, $\eta_p^2 = .220$, indicating that picture cues were overall more effective than English translation cues. A significant main effect for test also emerged, $F(1, 22) = 91.784$, $p < .001$, $\eta_p^2 = .807$, revealing that performance improved from Test 1 to Test 2. Finally, a significant interaction occurred, $F(1, 22) = 4.764$, $p = .040$, $\eta_p^2 = .178$, indicating that an advantage of picture cues over English translation cues occurred at Test 2, $t(22) = 2.765$, $p = .011$, $d = 0.53$, but not at Test 1, $t(22) = 1.781$, $p = .089$.

Self-pace study during encoding. Thus far, the data demonstrate the same pattern of findings from Carpenter and Olson's (2012) Experiment 2, showing that picture cues (compared with English translation cues) inflate confidence during encoding but do not significantly enhance accuracy on an initial test. Do the higher JOLs during encoding for picture cues reflect greater processing fluency?

Table 1. Average median JOLs according to cue type and time of JOL in Experiments 1 and 2.

	Encoding	Test 1	Total
Experiment 1			
Picture	40.00 (4.07)	40.87 (4.02)	40.44 (3.15)
English Translation	30.70 (4.72)	31.52 (4.22)	31.11 (3.95)
Experiment 2			
Picture	38.35 (2.62)	27.00 (3.35)	32.67 (2.48)
English Translation	33.87 (3.19)	24.91 (3.44)	29.39 (2.90)

JOL: judgment of learning. Standard errors are given in parentheses.

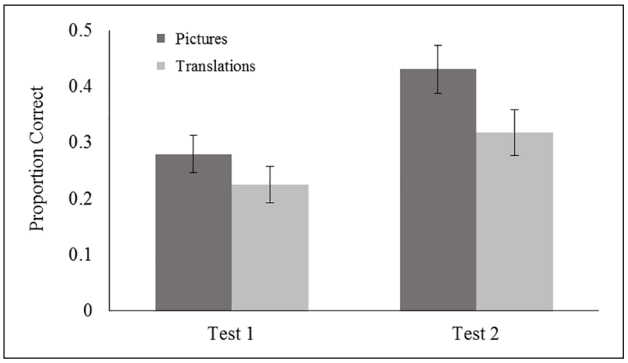

Figure 2. Proportion of Swahili words correctly recalled from picture cues versus English translation cues at Test 1 and Test 2 in Experiment 1. Error bars represent 95% within-subjects confidence intervals (Cousineau, 2005).

We examined this question by comparing self-paced study times during encoding for Swahili words paired with picture cues versus English translation cues. Examination of individual data revealed that many participants' self-paced study times were not normally distributed. Shapiro–Wilk tests revealed significant departures from normality for approximately half of the participants' self-paced study times during encoding associated with picture cues (44% of participants) and English translation cues (57% of participants). Thus, we compared each participant's median self-paced study time for picture cues versus English translation cues.

Self-paced study times during encoding were not faster for Swahili words associated with picture cues compared with English translation cues. In fact, a paired-samples *t*-test revealed that average median self-paced study times (in milliseconds) were significantly *slower* for picture cues (M=7,583, SD=5,312) than for English translation cues (M=7,027, SD=4,994), $t(22)=2.103$, $p=.047$, $d=0.44$.[2]

These findings run counter to the idea that metacognitive judgements for foreign language vocabulary learning are driven by processing fluency, at least to the extent that processing fluency is reflected in self-paced study times. However, it is worth noting that JOLs in Experiment 1 (and in Carpenter & Olson's, 2012, Experiment 2) were always given after the entire list of items had been encoded. It is well-established that JOLs given after a delay—compared with immediately after each item—result in better calibration (Dunlosky & Nelson, 1992; T. O. Nelson & Dunlosky, 1991; Weaver & Kelemen, 1997), presumably because delayed JOLs are more likely to rely on diagnostic cues from long-term memory than fleeting perceptual cues that can lead to miscalibrated confidence, such as processing fluency (Rhodes & Tauber, 2011).

Indeed, studies showing effects of perceptually salient cues (e.g., font size, relatedness, stimulus degradation, etc.) on confidence have relied exclusively on immediate JOLs (Mueller et al., 2016; Rhodes & Castel, 2008, 2009; Yue et al., 2013), and one study found that the contribution of processing fluency (reflected in self-paced study times) on metacognitive judgements for word pairs was indeed stronger when JOLs were made immediately versus after a delay (Koriat & Ma'ayan, 2005). It is possible, therefore, that effects of processing fluency in the current paradigm are more apparent under conditions of immediate JOLs than delayed JOLs. Experiment 2 set out to test this possibility.

Experiment 2

Participants

A new group of undergraduate students was recruited from the same participant pool as in Experiment 1. A total of 48 participants completed Experiment 2. The sample size was determined from previous laboratory-based studies using self-paced study times and immediate JOLs to examine contributions of processing fluency to metacognitive judgements (Mueller et al., 2016; Mueller et al., 2014; Undorf & Erdfelder, 2015; Witherby & Tauber, 2017), where sample sizes have typically consisted of about 40 participants. Participants completed the experiment individually on personal computers.

Materials, design, and procedure

Experiment 2 was in all respects identical to Experiment 1, except for the time during which the JOLs were made. Immediately after each (self-paced) encoding trial, participants were shown the same item again and asked to make a JOL for that item. Thus, the two presentations of each item (first self-paced encoding, then JOL) occurred back-to-back. The procedures for the initial test with feedback (Test 1) and the final test (Test 2) were identical to Experiment 1.

Results and discussion

Pre-analyses and scoring. Data from one participant were discarded due to familiarity with the Swahili language, and from one additional participant who started the experiment

but did not complete it. Data from the remaining 46 participants were analysed. Data were scored in the same way, by the same two coders, as in Experiment 1.

JOLs. Shapiro–Wilk tests again revealed significant departures from normality for the majority of participants' JOLs associated with picture cues (72% of participants) and English translation cues (67% of participants) during both encoding and at Test 1. Thus, we again compared JOLs for picture cues versus English translation cues using each participant's median JOL.

Table 1 reports average median JOLs for each cue type across encoding and Test 1. A 2 × 2 repeated-measures ANOVA with cue type (picture vs. English translation) and time of JOL (encoding vs. Test 1) as factors revealed a significant main effect of cue, $F(1, 45)=4.63$, $p=.037$, $\eta_p^2=.093$, and a significant main effect of time, $F(1, 45)=11.40$, $p=.002$, $\eta_p^2=.202$, but no interaction, $p=.40$. Thus, overall participants gave higher JOLs to Swahili words associated with pictures compared with English translations, and also gave higher JOLs during encoding than at Test 1. The JOL effect for cues was smaller than in Experiment 1. This was accompanied by a sharp decrease in JOLs from encoding to Test 1—an "underconfidence-with-practice" effect (Koriat, Sheffer, & Ma'ayan, 2002)—both types of cues, which coincided with reduced differences in JOLs between the two cue types at Test 1. It is possible that this underconfidence-with-practice effect, which did not occur in Experiment 1, was driven by the immediate JOLs in Experiment 2, as previous studies have shown that the effect occurs with immediate JOLs but not with delayed JOLs (Finn & Metcalfe, 2007).

Test accuracy. Figure 3 shows cued recall performance for Swahili words from pictures versus English translations at Test 1 and Test 2. A 2 × 2 repeated-measures ANOVA with cue type (picture vs. English translation) and test (Test 1 vs. Test 2) as factors revealed a significant main effect of cue type, $F(1, 45)=6.298$, $p=.016$, $\eta_p^2=.123$, a significant main effect of test, $F(1, 45)=166.761$, $p<.001$, $\eta_p^2=.787$, and a significant interaction, $F(1, 45)=27.708$, $p<.001$, $\eta_p^2=.381$, whereby the advantage of picture cues over English translation cues occurred at Test 2, $t(45)=4.328$, $p<.001$, $d=0.638$, but not at Test 1, $t(45)=0.747$, $p=.459$.

Self-paced study during encoding. To examine whether the higher JOLs during encoding for picture cues were driven by processing fluency, we again compared self-paced study times during encoding for picture cues versus English translation cues. Shapiro–Wilk tests again revealed significant departures from normality for participants' self-paced study times during encoding for picture cues (63% of participants) and English translation cues (67% of participants). Thus, we compared each participant's median self-paced study time for picture versus English translation cues.

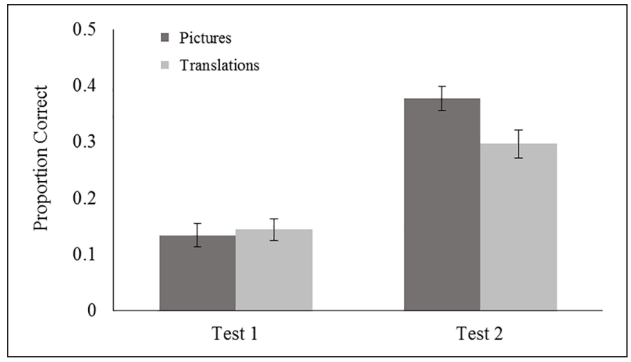

Figure 3. Proportion of Swahili words correctly recalled from picture cues versus English translation cues at Test 1 and Test 2 in Experiment 2. Error bars represent 95% within-subjects confidence intervals (Cousineau, 2005).

A paired-sample *t*-test on median self-paced study times during encoding (in milliseconds) indicated no difference between picture cues ($M=8{,}247$, $SD=7{,}008$) and English translation cues ($M=7{,}743$, $SD=7{,}030$), $t(45)=1.261$, $p=.214$.[3] These findings, like those of Experiment 1, run counter to the idea that metacognitive judgements for foreign language vocabulary learning are driven by processing fluency, at least reflected in self-paced study times.

If not processing fluency, then what? A series of recent studies has revealed that manipulations that commonly inflate metacognitive judgements and have been traditionally attributed to processing fluency—such as font size, cue–target relatedness, and concreteness—may not be due to processing fluency but rather to participants' subjective beliefs about learning. Manipulations designed to directly affect processing fluency—such as perceptual degradation—do not moderate the effects of relatedness on JOLs (Mueller et al., 2013, Experiment 2). Other studies have found that participants' self-paced study times do not differ between concrete and abstract words (Witherby & Tauber, 2017, Experiment 5), nor do self-paced study times mediate the relationship between JOLs and font size (Mueller et al., 2014, Experiment 2), or between JOLs and word pair relatedness (Mueller et al., 2016, Experiment 1).

In light of these results, researchers have recently proposed the *analytic-processing theory* of metacognitive judgements (Mueller & Dunlosky, 2017; Mueller et al., 2014). This theory proposes that participants engage in an analytic process to make a JOL based on their subjective beliefs about how a given cue will affect their memory performance, which serves to reduce their uncertainty in predicting how they will perform on a future test. They may think, for example, that related (compared with unrelated) word pairs are more memorable because of a belief that memory is better for things that are more strongly connected, so they will give higher JOLs to related than to unrelated word pairs. This theory does not propose that processing fluency fails to influence metacognitive judgements, only that

analytic processing can operate independently and affect these judgements even when processing fluency does not. Given the results of Experiments 1 and 2 showing that processing fluency—reflected in self-paced study times—does not appear to underlie metacognitive judgements in foreign language vocabulary learning, analytic processing is a viable alternative explanation. Experiment 3 was designed to provide a more direct test of analytic processing theory.

Experiment 3

Experiment 3 set out to determine whether metacognitive judgements in foreign language vocabulary learning are driven by participants' subjective beliefs about learning. In a between-subjects design, participants learned either the full list of Swahili words with either picture cues or English translation cues, and then completed a free recall test of the Swahili words. Afterward, they were given a description of the full design—informing them that some participants learned the Swahili words from pictures, and some from English translations—and asked which method they thought would be more effective for learning the Swahili words.

Participants' perceived effectiveness of the two cue types provides information about their metacognitive judgements without having directly experienced both cues during learning. Because processing fluency depends upon the experienced ease with which items are processed during learning, it would be more likely to influence judgements for the list that participants experienced during learning—either the picture cues or the English translation cues, but not both. As such, any effects of cue type on participants' perceived effectiveness of cues that were not experienced may be more attributable to their beliefs about learning.

Indeed, the same logic has been applied in the use of prestudy JOLs, in which participants are asked to estimate their own learning from a description of a manipulation *before* they experience it (see Mueller et al., 2016; Mueller et al., 2014; Mueller et al., 2013; Witherby & Tauber, 2017), or by providing participants with a hypothetical description of an experiment and asking them to estimate how the described manipulation (e.g., font size) would affect memory (Kornell et al., 2011; Mueller et al., 2014). If processing fluency underlies metacognitive judgements for picture cues, then participants who directly experience the pictures, compared with those who do not, should be more likely to deem the picture cues more effective. Alternatively, if analytic processing underlies these judgements, then a comparable proportion of participants in both groups—irrespective of which cue type they experienced during learning—should deem the picture cues more effective.

Participants

Sixty-two new participants were recruited from the same participant pool as in Experiments 1 and 2. Participants were randomly assigned to learn the Swahili words with either picture cues (the Picture Group, $n=31$) or English translation cues (the Translation Group, $n=31$). The sample size was determined from previous laboratory-based studies using prestudy JOLs or hypothetical descriptions of experimental procedures (Mueller et al., 2016; Mueller et al., 2014; Mueller et al., 2013) where sample sizes ranged from 24 to 48 per group. Data were discarded from one participant in the Picture Group who left the experiment before starting the memory test. One additional participant was added to the Picture Group as a replacement.

Materials, design, and procedure

Both groups of participants were informed that they would be learning some words in Swahili, and would later be given a memory test over those words. The instructions indicated that participants would be learning the meanings of the Swahili words by being shown a picture of the meaning of the word (for participants in the Picture Group) or by being shown the English translation for the word (for participants in the Translation Group). An example was given to each group using an item that was not among the experimental stimuli (*Train: Reli*).

Participants learned the same 42 pairs as in the previous experiments, in addition to one extra pair (*Fence: Uwa*) that was part of the original stimuli set (see Carpenter & Olson, 2012, Experiment 1) but was not included in Experiments 1 and 2 to achieve an equal number of items per condition. Both groups in Experiment 3 were first given a presentation of all 43 stimuli, for 6 s each, in a unique random order. Following this presentation, an additional identical presentation was given, with the items presented in a new random order for each participant.[4] Following the second presentation, participants were given 5 min to type in all of the Swahili words that they could remember.

The use of a free recall test over the Swahili words provides additional insights into what participants may be doing while they study Swahili words in the context of pictures versus English translations. Carpenter and Olson (2012) reasoned that a picture cue (compared with an English translation cue) may create a false sense of confidence that undermines the processing of the Swahili word paired with it. Even with invariant study times between picture–Swahili pairs and English–Swahili pairs, a picture may "steal" attention away from the Swahili word paired with it, resulting in a relative memory deficit for Swahili words paired with pictures compared with English translations. Indeed, in Carpenter and Olson's Experiment 1, participants who studied these item pairs showed significantly better later free recall of the pictures compared with the English translations, suggesting that the picture cues might receive enhanced processing, possibly at the expense of the Swahili words paired with them. As the type of cued recall test used in Experiments 1 and 2 of the current study

contains some elements of memory for the cues themselves, as well as the Swahili words and the connections between them, a free recall test of just the Swahili words provides a more direct measure of the encoding quality of the Swahili words according to the type of cues with which they were presented.

To further understand participants' approaches to learning the materials, they were given a strategy questionnaire after the free recall test. At the end of the 5-min free recall period, participants were informed via instructions on the computer that time was up, and they were asked to complete a questionnaire. The first item on the questionnaire inquired about participants' strategies that they used to learn the Swahili words. Participants in both groups were asked, "What type of strategy did you use to try to learn these materials? (check all that apply)," with the options for the Picture Group "(a): I looked at the picture while mentally repeating the Swahili word (e.g., "Kelb," "Kelb"), (b): I tried to name the picture and then repeated that word with the Swahili word (e.g., "Dog—Kelb," "Dog—Kelb"), (c): I tried to form a mental image of the Swahili word with the picture (e.g., a dog running on the beach with "kelp" in its mouth"), and the options for the Translation Group "(a): I mentally repeated the two words together (e.g., "Dog—Kelb," "Dog—Kelb"), (b): I tried to form a mental image of the English word (e.g., a big yellow dog) and then thought about that image while repeating the Swahili word (e.g., "Kelb," "Kelb," "Kelb"), (c): I tried to form a mental image of the English word and Swahili word together (e.g., a dog running on the beach with "kelp" in its mouth)." Both groups were also given the option "(d): Describe any other strategies you used that are not listed here."

For the second item on the questionnaire, participants were given a description of the full experimental design, along with the example stimuli shown in Figure 1. They were informed that people in this experiment learned the Swahili words in one of two ways—by seeing the Swahili word with a picture or by seeing it with its English translation—and asked which method they thought would lead to better learning of the Swahili word. They were encouraged to choose only one option: (a) picture, (b) English translation, or (c) both the picture and English translation would be equally effective. Upon completing these questions, participants were thanked and debriefed.

Results and discussion

Free recall performance. Participants' scores on the free recall test did not differ significantly between participants in the Picture Group ($M=0.17$, $SD=0.10$) and the Translation Group ($M=0.20$, $SD=0.14$), $t(60)=0.891$, $p=.376$. This confirms the lack of an advantage for picture cues over English translation cues on an initial free recall test using a between-subjects design, which is consistent with the results of Experiments 1 and 2 based on a cued recall within-subjects design. The small and non-significant advantage of the Translation Group over the Picture Group suggests that if picture cues do "steal" attention away from Swahili words during learning, they do not appear to do so to a large degree.

Perceptions of cue effectiveness. Analyses of questionnaire responses exclude data from seven participants (three in the Picture Group and four in the Translation Group) who did not complete the questionnaire. We first calculated the proportion of participants from each group indicating which cue they thought would be most effective: the picture, the English translation, or both. The proportions for the Picture Group (respectively, 61%, 14%, and 25%) were highly comparable with those for the Translation Group (respectively, 74%, 15%, and 11%). The majority of participants in both groups indicated that the picture would be more effective. This proportion did not depend on group, $\chi^2(1)=1.110$, $p=.391$, indicating that pictures were believed to be more effective cues regardless of whether or not those cues had been directly experienced during a learning phase.

Strategies. Table 2 displays the proportion of participants who reported using a particular strategy to learn the Swahili words. The majority of participants reported using a verbal rote rehearsal strategy, in which they repeated the two words together (74% in the Translation Group) or mentally named the picture and then repeated the two words together (68% in the Picture Group). Fewer participants reported using interactive mental imagery, and fewer still reported using a more elaborate strategy such as constructing phrases or stories, or relating the words to their own lives.

These data reveal important insights into students' strategies for learning foreign language vocabulary. In particular, the majority of participants appeared to favour an approach based on rote rehearsal, in which they mentally repeat the stimuli over and over. This was true whether participants learned the Swahili words with picture cues or with English translation cues, suggesting that rote rehearsal is a common strategy regardless of cue type. The tendency to favour rote rehearsal could partially explain the fairly low levels of recall of foreign language words on the initial tests in these and other experiments (Carpenter & Olson, 2012). It could also explain why, even if picture cues receive enhanced processing during learning (Carpenter & Olson, 2012, Experiment 1), this does not necessarily result in a significant deficit for the Swahili words paired with them, as there could be some adequate level of rote rehearsal applied to the Swahili words regardless of cue type.

Consistent with the analytic processing view, these data show that—regardless of which type of cue participants experienced during learning—afterward they were more likely to endorse picture cues as more effective than English translation cues. Directly manipulating which type of cue participants received during learning, and observing

Table 2. Self-reported strategies used to learn the Swahili words in Experiment 3.

Strategies	Picture group (%)	Translation group (%)
Look at picture while mentally repeating Swahili word	54	–
Name picture and mentally repeat with Swahili word	68	–
Form interactive image of Swahili word with picture	32	–
Mentally repeat the two words together	–	74
Form mental image of English word and repeat Swahili word	–	11
Form interactive image of English word and Swahili word	–	30
Other strategies		
Circular (e.g., repeat words out loud, relate words together)	0	11
Phonetic (e.g., transform Swahili word into similar-sounding word in English or another language)	18	7
Constructive (e.g., create phrases or stories)	11	0
Self-reference (e.g., link to own life/experiences)	4	4
Unclassified (relate things to picture, use people to remember, remember appealing words)	4	4

The write-in strategies labelled as "other" were classified as circular (reflecting one of the options already provided, such as simply repeating the two words), phonetic (a sound-based strategy), constructive (a strategy based on creating sentences or phrases), self-reference (a strategy involving connecting the stimuli to oneself), or unclassified (strategies that could not be classified given the participants' descriptions).

no effects of this experience on later judgements of cue effectiveness, indicates that such judgements do not appear to be based on direct experience with the cue during learning. Although participants received examples of both cue types during the pre-experiment instructions and thus could have based their judgements to some degree on these experiences, during the learning phase they received more extensive experience with one type of cue or the other. As participants' judgements about the effectiveness of each cue did not match the cue with which they had the most experience during learning, these data support the idea that the judgements of cue effectiveness are based more on participants' beliefs about which cue is better for learning than on their direct experiences with one type of cue or the other during learning.

It is possible, however, that the experience of learning the items in Experiment 3 could have affected participants' judgements in unknown ways. Thus, Experiment 4 was conducted to further examine participants' judgements about the effectiveness of picture versus English translation cues without first experiencing either type of cue during a learning phase.

Experiment 4

Participants

Fifty new participants were recruited from the same participant pool as in the previous experiments.

Design and procedure

Participants in Experiment 4 were given the same basic question about cue effectiveness from Experiment 3, including a description of the experimental design and the stimuli shown in Figure 1. Participants in Experiment 4, however, did not learn any of the words. Instead, they were informed that a separate group of people participated in a study in which they learned the Swahili words in one of two ways—by seeing the Swahili word with a picture, or by seeing it with its English translation. They were then asked to indicate which method they thought would lead to better learning of the Swahili word (choosing only one option): (a) picture, (b) English translation, or (c) both the picture and English translation would be equally effective.

Results and discussion

Data were discarded from two participants who did not provide an answer to the question. From the remaining 48 participants, 38 (79%) indicated that the picture cue would be more effective, five (10%) indicated that the English translation cue would be more effective, and five (10%) indicated that both cues would be equally effective. The proportion of participants endorsing the picture cues was significant, $\chi^2(1) = 16.33$, $p < .001$, indicating that even in the absence of a learning phase providing experience with either type of cue, participants had a strong tendency to believe that the picture cues would lead to better learning of the Swahili words.

General discussion

The current experiments shed new light on the factors underlying metacognitive judgements in foreign language vocabulary learning. Specifically, the inflated confidence for picture cues, shown here and elsewhere (Carpenter & Olson, 2012, Experiment 2), does not appear to be driven by processing fluency as measured by self-paced study times. Self-paced study times were not faster in either Experiment 1 or Experiment 2—and in fact, were significantly slower in Experiment 1—for Swahili words encoded

with picture cues compared with English translation cues. Other studies have employed self-paced study procedures and found that study times do not differ as a function of manipulations traditionally attributed to fluency, such as concreteness (Witherby & Tauber, 2017), and self-paced study times do not mediate the relationship between font size and JOLs (Mueller et al., 2014), or between word relatedness and JOLs (Mueller et al., 2016).

It is worth noting, however, that self-paced study times might reflect things other than, or in addition to, processing fluency. Pictures, for example, might be more complex or contain more information compared with English translations, and the time that participants spend studying Swahili words accompanied by pictures (compared with English translations) could partly reflect the processing of this extra information. Here the data from Experiment 3 can be useful by revealing information about participants' self-reported strategies. In both cases (learning from picture cues or English translation cues), the majority of participants reported using a rote rehearsal strategy in which they named the picture and repeated it with the Swahili word (in the Picture Group) or repeated the English word with the Swahili word (in the Translation Group). Given that verbal rote rehearsal is the most common strategy for both types of cues, exploring the time needed to execute this rote rehearsal strategy can help reveal what might be contributing to the self-paced study times in both conditions.

In Carpenter and Olson's (2012) original study, the stimuli were pilot tested by giving separate groups of participants just the pictures or just the English words, and asking participants to name them as quickly as possible. For the pictures, this meant producing the name of the object that participants believed was in the picture, and for the English words, this meant simply reading the word. Participants produced the expected names for the pictures more than 95% of the time, verifying that the objects in the pictures could be successfully identified at a high rate. Of interest for present purposes, these participants' average median correct naming latencies for pictures ($M = 1{,}874.86$, $SD = 369.13$) were slower than for words ($M = 1{,}442.04$, $SD = 349.55$), $t(40) = 3.64$, $p = .001$. On average, it took about half a second longer to name the pictures than to read the words (a difference of 432.82 ms). This is consistent with previous research showing that it takes longer to name a picture than to read its associated word (Kroll & Stewart, 1994). When using a verbal rote rehearsal strategy, self-paced study times might therefore be slowed down for picture cues relative to English translation cues because of the extra step (i.e., naming the picture) that is involved in studying picture cues. Thus, even if picture cues were processed more fluently (i.e., more quickly) than English translation cues, the extra time needed to name the picture could have obscured what would have otherwise been faster and more fluent processing of the picture cues.

Based on the self-paced study times in Experiments 1 and 2, however, even after adjusting for the 433-ms "head start" for English translation cues (i.e., by subtracting this time from the average self-paced study time for picture cues for each participant), the difference in self-paced study times still did not differ significantly between picture cues and English translation cues in Experiment 1, $t(22) = 0.47$, $p = .65$, or in Experiment 2, $t(45) = 0.18$, $p = .86$. Thus, even after controlling for potential time-sensitive differences in the use of a rote rehearsal strategy between the two cue types, we observed no significant differences in self-paced study times—the measure of processing fluency adopted in the current experiments—between picture cues versus English translation cues.

Experiment 3 showed that participants judge picture cues as more effective than English translation cues whether they had experienced these picture cues during a learning phase or not, and Experiment 4 showed the same preference for picture cues under conditions where participants experienced neither picture cues nor English translation cues during a learning phase. This is consistent with a number of recent studies showing that in the absence of a direct experience processing the stimuli during a learning phase—that is, by asking participants to make prestudy JOLs, or JOLs in response to a description of stimuli that have not been encoded—JOLs are higher for related than for unrelated cue–target pairs (Mueller et al., 2013), for identical than for related cue–target pairs (Mueller et al., 2016), for larger-font words than for smaller-font words (Kornell et al., 2011; Mueller et al., 2014), and for concrete words than for abstract words (Witherby & Tauber, 2017). In the absence of an opportunity to study the target stimuli during a learning phase, participants have no experience of the ease with which those stimuli are processed and are therefore more likely to base their metacognitive judgements on their beliefs about learning. Some of these metacognitive judgements may reflect pre-existing beliefs about learning, but beliefs may also be influenced by experience, in that experience with the list of items (or simply hearing about the characteristics of the items) may alter pre-existing beliefs or create new beliefs about learning.

Like Carpenter and Olson (2012), we found in Experiments 1 and 2 that picture cues were not significantly more effective than English translation cues at Test 1, but were significantly more effective at Test 2. It is likely that participants' performance at Test 1 provided critical diagnostic information about their own performance—that is, the "reality check"—that informed their subsequent judgements, resulting in improved performance at Test 2. This test-induced recalibration has been demonstrated in the literature, usually in the form of a shift from initial overconfidence to underconfidence following test practice (Koriat et al., 2002). The improved calibration is likely driven by performance on the initial test, with higher subsequent JOLs given to items that were recalled

successfully and lower JOLs given to items that could not be recalled. Previous research provides evidence for the use of this memory-for-past-test (MPT) heuristic and the role that it plays in the underconfidence-with-practice effect (Finn & Metcalfe, 2007).

Indeed, in the current study, participants appeared to make use of the MPT heuristic, as gamma correlations were stronger between JOLs made at Test 1 and performance on Test 1, compared with JOLs made at encoding and performance on Test 1. This was true for Experiment 1, respectively, $M = 0.90$, $SD = 0.12$; $M = 0.48$, $SD = 0.28$; $t(22) = 7.838$, $p < .001$, $d = 1.634$, and for Experiment 2, respectively, $M = 0.82$, $SD = 0.44$; $M = 0.42$, $SD = 0.39$; $t(43) = 4.585$, $p < .001$, $d = 0.691$. Evidence for MPT further supports the idea that metacognitive judgements in the current paradigm may have been driven by beliefs, as previous studies have shown that MPT tends to reflect participants' beliefs about how their past test performance is related to their future test performance (Serra & Ariel, 2014).

The reality check provided by Test 1 resulted in a significant performance advantage for picture cues over English translation cues at Test 2, possibly reflecting a shift in how participants approached the material after trying to recall it on Test 1, and in particular a shift to a more effective means of learning the Swahili words from picture cues (compared with English translation cues) during the feedback provided at Test 1. Although the current design does not permit any direct measures of how the processing may have differed for the two types of cues (i.e., the feedback presentation time was always fixed), exploring self-paced feedback is an interesting idea for future research that may reveal insights into processing- or strategy-related shifts in learning from the two types of cues following an initial test. Given that JOLs were generally higher for picture cues compared with English translation cues, it is possible that the superior downstream memory performance for picture cues—driven by a greater proportion of initial errors being corrected—is due to mechanisms similar to those underlying the hypercorrection effect (Butterfield & Metcalfe, 2001; Carpenter, Haynes, Corral, & Yeung, 2018).

Given the support for the idea that metacognitive judgements in foreign language vocabulary learning reflect beliefs about learning, it is worth considering where these beliefs might originate. One possibility is that participants have learned that pictures are generally more effective for memory than verbal materials. Indeed, decades of research on the *picture superiority effect* have revealed that pictures of familiar things are more easily recalled and recognised than the words corresponding to those things (Durso & Johnson, 1980; Hockley, 2008; D. L. Nelson, 1979; D. L. Nelson, Reed, & McEvoy, 1977; Paivio & Csapo, 1973; Paivio, Rogers, & Smythe, 1968; Smith & Magee, 1980). If participants have experienced this phenomenon in their own lives, and this experience informs the belief that pictures themselves are readily accessible in memory, that could lead to the intuitive (but not always accurate) assumption that pictures are also highly effective cues for remembering other things, such as new words in a foreign language.

Although the current study shows that processing fluency measured via self-paced study times does not appear to underlie metacognitive judgements in foreign language vocabulary learning, it is worth noting that processing fluency has been shown to underlie metacognitive judgements in other paradigms. For example, Undorf and Erdfelder (2015) found that related word pairs (compared with unrelated word pairs) received higher JOLs and were processed faster during self-paced study, and the relationship between word pair relatedness and JOLs was partially mediated by self-paced study times. Along similar lines but with text materials, Ball et al. (2014) found that bolded (compared with unbolded) terms and definitions received higher JOLs and were processed faster, and the relationship between bolding and JOLs was also partially mediated by self-paced study times.

Although self-paced study provides one way of measuring processing fluency, other measures of processing fluency have been considered as well and have sometimes yielded different experimental outcomes. Some studies have found that metacognitive judgements are driven by other measures that reflect processing fluency, such as the time it takes to name a perceptually degraded stimulus (Undorf, Zimdahl, & Bernstein, 2017), the time it takes to identify a stimulus presented after a mask in a continuous identification (CID) paradigm (Yang, Huang, & Shanks, 2018), and the number of practice trials required to reach acquisition (Undorf & Erdfelder, 2015; but see Witherby & Tauber, 2017). Additional measures of processing fluency might be available via eye-tracking methodology, as there is evidence that participants fixate longer to text stimuli that are rendered objectively disfluent (e.g., presented in hard-to-read vs. easy-to-read font; Sanchez & Jaeger, 2015; Slattery & Raynor, 2010), and changes in pupil dilation can provide a reliable measure of subjective processing associated with cognitively demanding tasks (Goldinger & Papesh, 2012). Such methodology has yet to be integrated into systematic investigations of the factors underlying metacognitive judgements, but could prove useful for measuring real-time processing fluency. Thus, although self-paced study times are one way to measure processing fluency, they provide only part of the story about the contribution of processing fluency to metacognitive judgements, and should be considered alongside other measures of processing fluency as well.

In conclusion, using pictures as cues to learn foreign language vocabulary can inflate confidence, but not necessarily accuracy. The current experiments show that this tendency does not appear to be driven by processing fluency as measured by self-paced study times, but rather by people's beliefs that pictures serve as more effective cues

for memory than English translations. Given the heavy reliance on pictures in real-world learning of linguistic and scientific information, optimising the effectiveness of pictures in instruction depends on a thorough understanding of how students approach learning of these materials, and the important role of students' beliefs during all stages of the learning process.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This material is based upon work supported by the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition, Collaborative Grant No. 220020483.

ORCID iD

Shana K Carpenter https://orcid.org/0000-0003-0784-9026

Notes

1. We analysed these data using both mean and median judgments of learning (JOLs), and the same pattern of statistically significant results emerged in both cases.
2. We analysed these data using both mean and median self-paced study times, and using both parametric and non-parametric tests. In all cases, self-paced study times were never faster for picture cues compared with English translation cues (showing, if anything, longer study times in some cases for picture cues). Additional analyses of participants' self-paced JOLs revealed no significant differences in the time participants took to make their JOLs for picture cues compared with English translation cues.
3. See Note 1.
4. Initial pilot testing revealed that, after only one presentation, performance on the memory test was quite low. Two presentations were given to avoid potential floor effects that might obscure performance differences between the two groups.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting tribes of fluency to form a metacognitive nation. *Personality & Social Psychology Review*, *13*, 219–235.

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*, 336–348.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 1491–1494.

Carpenter, S. K., Haynes, C. L., Corral, D., & Yeung, K. L. (2018). Hypercorrection of high confidence errors in the classroom. *Memory*, *26*, 1379–1384.

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*, 161–172.

Carpenter, S. K., Northern, P. E., Tauber, S. K., & Toftness, A. R. (in press). Effects of lecture fluency and instructor experience on students' judgments of learning, test scores, and evaluations of instructors. *Journal of Experimental Psychology: Applied*.[AQ: 1]

Carpenter, S. K., & Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*, 92–101.

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*, 1350–1356.

Chen, H.-C. (1990). Lexical processing in a non-native language: Effects of language proficiency and learning strategy. *Memory & Cognition*, *18*, 279–288.

Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.

Dunlosky, J., & Tauber, S. K. (Eds.) (2016). *The Oxford handbook of metamemory*. New York, NY: Oxford University Press.

Durso, F. T., & Johnson, M. K. (1980). The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning & Verbal Behavior*, *19*, 416–429.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *33*, 238–244.

Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students' experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review*, *27*, 567–586.

Foster, N. L., Was, C. A., Dunlosky, J., & Isaacson, R. M. (2017). Even after thirteen class exams, students are still overconfident: The role of memory for past exam performance in student predictions. *Metacognition & Learning*, *12*, 1–19.

Geller, J., Still, M. L., Dark, V. J., & Carpenter, S. K. (2018). Would disfluency by any other name still be disfluent? Examining the disfluency effect with cursive handwriting. *Memory & Cognition*, *46*, 1109–1126.

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, *21*, 90–95.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*, 160–170.

Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics? *Contemporary Educational Psychology*, *49*, 80–90.

Hockley, W. E. (2008). The picture superiority effect in associative recognition. *Memory & Cognition*, *24*, 202–216.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory & Language*, *52*, 478–492.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*, 147–162.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *2*, 787–794.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory & Language*, *33*, 149–174.

Lenzner, A., Schnotz, W., & Mueller, A. (2013). The role of decorative pictures in learning. *Instructional Science*, *41*, 811–831.

Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69.

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *37*, 502–506.

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about memory. *Journal of Memory & Language*, *93*, 245–258.

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

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 & Language*, *70*, 1–12.

Mueller, M. L., Tauber, S. K., & Dunlosky, J. (2013). Contributions of beliefs and processing fluency to the effect of relatedness on judgments of learning. *Psychonomic Bulletin & Review*, *20*, 378–384.

Nelson, D. L. (1979). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 45–76). Hillsdale, NJ: Lawrence Erlbaum.

Nelson, D. L., Reed, V. S., & McEvoy, C. L. (1977). Learning to order pictures and words: A model of sensory and semantic encoding. *Journal of Experimental Psychology: Human Learning & Memory*, *3*, 485–497.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOL) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*, 267–270.

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*, 176–206.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, *11*, 137–138.

Reber, R., & Greifeneder, R. (2017). Processing fluency in education: How metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist*, *52*, 84–103.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.

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

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, *137*, 131–148.

Sanchez, C. A., & Jaeger, A. J. (2015). If it's hard to read, it changes how long you do it: Reading time as an explanation for perceptual fluency effects on judgment. *Psychonomic Bulletin & Review*, *22*, 206–211.

Serra, M. J., & Ariel, R. (2014). People use the memory for past test heuristic as an explicit cue for judgments of learning. *Memory & Cognition*, *42*, 1260–1272.

Serra, M. J., & Dunlosky, J. (2010). Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory*, *18*, 698–711.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

Slattery, T. J., & Raynor, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology*, *24*, 1129–1148.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, *109*, 373–392.

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 & Learning*, *13*, 1–14.

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, *43*, 647–658.

Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *45*, 97–109.

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on

judgments of learning. *Journal of Memory & Language*, *92*, 293–304.

Unkelbach, C., & Greifeneder, R. (2013). *The experience of thinking: How the fluency of mental processes influences cognition and behavior*. New York, NY: Psychology Press.

Weaver, C. A., & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science*, *8*, 318–321.

Wiley, J., Sarmento, D., Griffin, T. D., & Hinze, S. R. (2017). Biology textbook graphics and their impact on expectations of understanding. *Discourse Processes*, *54*, 463–478.

Witherby, A. E., & Tauber, S. K. (2017). The concreteness effect on judgments of learning: Evaluating contributions of fluency and beliefs. *Memory & Cognition*, *45*, 639–650.

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

Yue, C. L., Castel, A. D., & Bjork, R. A. (2013). When disfluency is—and is not—a desirable difficulty: The influence of typeface clarity on metacognitive judgments and memory. *Memory & Cognition*, *41*, 229–241.