Journal of Applied Research in Memory and Cognition 7 (2018) 540–551

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Testing Judgments of Learning in New Contexts to Reduce Confidence

Gabriel D. Saenz*, Steven M. Smith*

Texas A&M University, United States

Students often prepare for exams by restudying slides or notes, but the familiarity of these repeatedly reinstated contexts may influence students to be unduly confident in their knowledge, and cease study prematurely. Experiment 1 investigated the effects of reinstatement of encoding contexts during delayed judgments of learning (JOLs) in a laboratory paradigm. At encoding, word pairs were superimposed over unrelated 5-s video contexts, with one video context per target. Delayed judgments of learning for cue–target pairs were given with encoding contexts either changed or reinstated. When the encoding context was changed, judgments of learning decreased relative to the reinstatement condition, without a concomitant decrease in recall (tested later with no context cues), thus producing a metamemory illusion. Experiments 2 and 3 replicated these results with different designs. Although context manipulations did not improve metacognitive accuracy, *unfamiliar* contexts reduced JOLs, a desirable impact that is associated with increased study choice.

General Audience Summary

When students predict their performance on an exam, they are consistently overconfident. Further, lower-performing students are more overconfident in their predictions. Many therefore hypothesize that overconfidence may lead students to perform poorly, suggesting that overconfident students stop studying for their exams prematurely. We investigate one aspect of predictions which may be causing overconfidence, the reinstatement of study context during predictions. We refer to context as any setting or condition under which students learn that may be repeated during study time. If a student rereads their notes as a method of studying, they may have reinstated context because they worked with their notes both during class and during study time. We found that the reinstatement of context causes people to overestimate themselves as they assess how much they know, potentially because the repeated context setting makes the studied information seem more familiar than it actually is. Educators may try to avoid overconfidence, and thereby improve exam performance, by promoting the use of varied context during study.

Keywords: Judgments of learning, Context, Reinstatement, Memory illusion, Overconfidence, Students

Overconfidence can lead learners to undervalue additional study time, a metacognitive failure with many potential causes (see Metcalfe & Finn, 2008a; Saenz, Geraci, Miller, & Tirso, 2017). If students overestimate their learning, they might prematurely stop studying and underperform at test. How can we reduce overconfidence in students? We expect that students might be less likely to judge material as "already learned" if their judgments of learning (JOLs) are made in novel contexts. Context is here defined as sensory information associated with the physical environment under which a person interacted with learned material. Also known as *environmental context*, this information has been shown to affect recall (Smith, Glenberg,

Author Note

* Correspondence concerning this article should be addressed to Gabriel D. Saenz & Steven M. Smith, Department of Psychology, Texas A&M University, College Station 77843, TX, United States. Contact: gabriel.s@tamu.edu, stevesmith@tamu.edu

& Bjork, 1978; see also Smith, 1988, 1994, 2007). By reinstating contexts during judgments of learning (e.g., rereading one's own notes, reviewing material in one's study context), people may unwarily tap into context-enhanced memory retrieval, mistakenly perceive learning to be greater relative to a new-context judgment of learning, and subsequently reduce studying behavior. Three experiments tested the hypothesis that prompting judgments of learning in new contexts reduces judgments of learning, but does not affect subsequent memory of newly learned material. If novel contexts reduce judgments of learning relative to reinstated-context judgments of learning without affecting memory, students should be encouraged to use unfamiliar contexts during study to increase study choice behavior.

Knowledge Monitoring in Students

As students study, they must identify when they are "ready" for their test—a form of metamemory judgment where they judge their level of learning. Metamemory judgments, such as judgments of learning and feelings-of-knowing (FOKs), are influenced by several heuristics including cue familiarity (Metcalfe & Finn, 2008a; Metcalfe, Schwartz, & Joaquim, 1993; Reder & Ritter, 1992; Schwartz & Metcalfe, 1992), processing fluency (Alter & Oppenheimer, 2009; Susser & Mulligan, 2015), the accessibility of target-related material (e.g., Koriat, 1993), and the retrieval fluency of memory targets (e.g., Benjamin, Bjork, & Schwartz, 1998; Schwartz, Benjamin, & Bjork, 1997). A new theory has challenged this fluency explanation (e.g., Bjork, Dunlosky, & Kornell, 2013; Hu et al., 2015; Mueller, Dunlosky, Tauber, & Rhodes, 2014; Mueller, Tauber, & Dunlosky, 2013), suggesting that belief, rather than fluency, may better explain such metamemory illusions. However, further evidence for the influence of fluency on judgments of learning has emerged (Besken & Mulligan, 2013; Undorf & Erdfelder, 2015; Undorf, Zimdahl, & Bernstein, 2017). Despite this ongoing conversation, the theoretical motivation for these studies stems from the aforementioned findings that broadly indicate that more favorable or fluent memory conditions produce more optimistic self-assessments. In the present experiments we used judgments of learning to test this prediction; such a metamemory illusion could negatively impact students.

Judgments of learning can affect students' choices about whether to study more and what to study (Dunlosky & Hertzog, 1998; Metcalfe, 2002; Metcalfe & Finn, 2008b), and greater confidence in one's learning can lead to less studying behavior. Furthermore, studies have found metacognitive inaccuracy—particularly overconfidence—to be associated with low academic performance (e.g., Hacker, Bol, Horgan, & Rakow, 2000; Miller & Geraci, 2011). It is clear that accurate knowledge monitoring, or lower confidence at least, may improve test scores by increasing studying behavior.

There are several conditions that can affect metamemory judgments and, by proxy, studying behavior. Since increased familiarity, perceptual fluency, and target accessibility can increase judgments of learning, one may conclude that conditions that favor or at least appear to favor recall also increase one's metamemory confidence. For example, reinstatement of the encoding contexts of target events enhances episodic memory retrieval, or recollection (e.g., Smith, 1979; Smith et al., 1978; Smith & Manzano, 2010; Smith & Vela, 2001). Therefore, students may choose to study more if they vary their study contexts, because new contexts reduce confidence in one's learning.

Context and Metamemory

We hypothesized that testing judgments of learning (JOLs) for newly learned associations in new contexts would be lower, relative to JOLs given in reinstated contexts, without concomitant decreases in recall (given that there are no context cues provided at recall). In addition to existing pilot research (Nichols & Smith, 2011), there have been other findings of context effects on metamemory judgments. Hockley (2008) found that study–test changes in the background color, word location, or background image of studied words not only decreased hits and false alarms, but it also decreased confidence of recognition decisions. Context change has also been found to decrease feeling-of-knowing judgments (Hanczakowski, Zawadzka, Collie, & Macken, 2017).

Although previous studies have not examined context effects with judgments of learning, a number have investigated similar effects of other cues on feeling-of-knowing and judgments of learning. For example, Metcalfe et al. (1993) presented participants with multiple varieties of cue-target (A–B) pairs in which the original cues were presented multiple times with different targets; judgments of learning were influenced by the number of presentations of the cue despite no change in target recall (see also Besken & Mulligan, 2013; Eakin, 2005; Maki, 1999; Undorf & Erdfelder, 2015; Undorf et al., 2017). In these studies, manipulations of intentionally encoded cues affected metamemory judgments without concomitant effects on the memory targets associated with the cues. In the present investigation, *contexts* during judgments of learning were manipulated instead of the verbal cues that are essential for the memory test. Manipulations of contexts allow the same learned target items (such as cue–target pairs) to be presented in various ways. In the present study, we presented and repeated the same cue–target pairs in reinstated or new contexts. These conditions may mimic those under which students normally study and monitor their learning, because as they reread their texts or notes, they may unconsciously bias themselves to be overconfident about their level of learning and cease study prematurely.

The Present Experiments

We tested the hypothesis that context change, relative to context reinstatement, causes a decrease in delayed judgments of learning. Participants first studied word pairs, made judgments of learning with reinstated or changed contexts, and took a cued recall test without contexts (Figure 1). During study, each word pair was shown superimposed over a different video context. Delayed judgments of learning were given for each pair either in the same context seen at encoding, or with a different context. Before the judgment of learning task, participants were informed that no video contexts would be presented on the subsequent

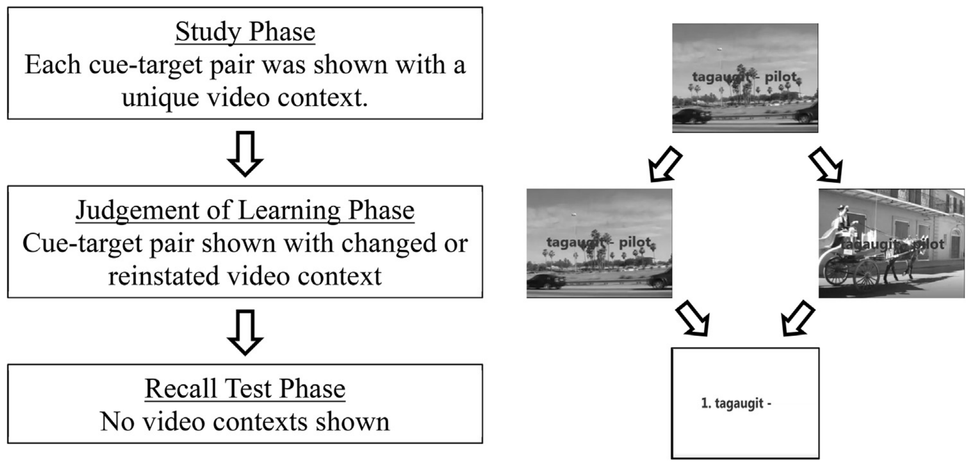

Figure 1. General procedure observed in all experiments.

recall test. Thus, we predicted that we would find weaker (or less frequent) judgments of learning when the judgments were made with changed contexts compared to reinstated contexts. We expected no difference in final recall scores because no contexts were presented at final recall.

In Experiment 1 we investigated context effects during judgments of learning using percentage-based judgments of learning (0–100), and manipulated reinstatement within-subjects. Experiment 2 replicated Experiment 1, except that judgments of learning were made with binary judgments (yes–no). Binary judgments of learning can produce different results (Hanczakowski, Zawadzka, Pasek, & Higham, 2013; Higham, Zawadzka, & Hanczakowski, 2016; Metcalfe & Finn, 2013; Zawadzka & Higham, 2015) and may provide converging evidence for the effects of context manipulations. Experiment 3 replicated Experiment 2, but manipulated context reinstatement between subjects, as the within-subjects manipulation may have called attention to context manipulations, thus producing changes in judgments of learning as a demand characteristic.

Experiment 1

The contexts used in the reported experiments were 5-s video recordings of various places. None of the scenes were known by participants, but the types of scenes were familiar to participants, such as a restaurant, a stairway, an athletic field, an elevator, windmills, or a busy street. Each video included motion and sound, but no audible dialog. Verbal stimuli appeared superimposed in large letters over the video contexts. The method is described in detail by Smith and Manzano (2010) and Smith, Handy, Angello, and Manzano (2014).

Whereas most cue–target experiments in judgment of learning literature have presented only cues during delayed judgments of learning (e.g., Metcalfe & Finn, 2008a; Nelson & Dunlosky, 1991, 1992), the present experiments showed both cues and targets for delayed judgments of learning, as was done by Dunlosky and Nelson (1992). Including memory targets at the delayed judgments of learning ensured that all memory targets were seen twice before the final recall test, regardless of treatment conditions. Alternatively, the inclusion of only cues during judgments of learning may affect recall inconsistently, as some targets may be recalled and others may not. Given that judgments of learning may affect recall (Mitchum, Kelley, & Fox, 2016; Naveh-Benjamin & Kilb, 2012; Nguyen & McDaniel, 2016; Soderstrom, Clark, Halamish, & Bjork, 2015; Spellman & Bjork, 1992), the inclusion of cue–target pairs ensured that each item was studied to a similar extent. Studies that present only cues (and not memory targets) at the time of a delayed judgment of learning are likely to find that retrieval successes during judgments of learning lead to higher judgments of learning and better subsequent recall, whereas retrieval failures lead to lower judgments of learning and do not enhance subsequent recall (see Metcalfe & Finn, 2008a). As in the Dunlosky and Nelson (1992) study, or the Benjamin et al. (1998) study in which participants judged their future recall of memory targets that they had just successfully generated, the methods we used allowed us to equalize exposure to critical memory targets during judgments of learning.

We sought to approximate, in a simplified, controlled way, the conditions a student might face when preparing for an exam. The experiment consisted of three phases: an encoding phase, a delayed judgment of learning phase, and a cued-recall test for the studied words (see Figure 1). Importantly, participants were informed before encoding, and again before the judgment of learning phase of the experiment, that no contexts would be shown on the final cued recall test. By making judgments of learning before taking the exam, we imitate the conditions under which a student might evaluate themselves while preparing to take a test, and by testing recall without any contexts, we can obtain a clear picture of the process producing change in judgments of learning. If context change decreased judgments of learning without an associated change in recall relative to the reinstatement condition, we can conclude that context can directly affect metamemory judgments. Alternatively, if context change produces a decrease in both judgments of learning and recall, context change may be limited to affecting judgments of learning through changed recall, a result that would indicate that context reinstatement during study is beneficial because it increases recall at test.

Method

Participants. A total of 122 undergraduate students at a large state university participated in Experiment 1 in return for partial course credit. Participation was voluntary and alternative credit options were made available. Data from one participant who was fluent in Tagalog was excluded from the analyses. Participants were randomly assigned to counterbalancing conditions (2×2: encoding order and reinstated items). Participant quantity was chosen to ensure at least 30 participants in each of the four counterbalancing conditions. There were no significant effects or interactions of counterbalancing conditions ($p > .306$), which were thus excluded from analyses. Multiple participants were run per study session, where the number of participants in each session varied from 5 to 10 participants per session, depending on random online enrollment by participants.

Materials. Forty-eight Tagalog–English cue–target pairs (items) were used in conjunction with 72 context videos. The Tagalog language was chosen because of its use in previous studies (Smith & Handy, 2016; Schwartz & Frazier, 2005), and because all Tagalog words were pronounceable using English conventions and letters. Any obvious connections within the word pairs (such as cognates), as well as between the word pairs and their accompanying context videos, were avoided. English word cues were all concrete nouns, with an average word frequency of 1972.79 (Gardner & Davies, 2013) and an average length of 5.20 characters. Tagalog targets had an average character length of 5.80. For each word pair, the Tagalog word served as the cue and the English word was the to-be-recalled target. A total of 40 5-s context videos were used in this study.

Each of three study phases presented a video including context and word pairs for encoding, context and word pairs for judgments of learning, or cues without targets or context for a recall test (Figure 1). The context video scenes were of environments encountered on a daily basis (e.g., a carwash, a stairway, a ball-field, a museum), and included no discernible verbal audio or text. The word pairs were superimposed over the context videos using Microsoft Movie Maker's *Titles* function and appeared in a large red font in the center of each video scene (for examples of the videos, see http://people.tamu.edu/ stevesmith/BRM_videos/; Smith and Manzano, 2010). In all phases, items were presented in bold red text for 5 s with 3 s of a blank white image in between each item.

Design. Context reinstatement was manipulated between subjects, so that each subject responded to 24 items with a reinstated context during judgments of learning, and 24 items with a novel context during judgments of learning. The same 48 items were used in all conditions, but the order of items, item–context pairing, and item–context reinstatement or non-reinstatement, was counterbalanced between testing sessions. Items were also presented in a different, random order in each phase of the study.

Procedure. Participants were seated in a group in front of a large video screen. They were read instructions indicating they would perform a task that would test their ability to learn new material and to predict their future performance. They were then told that they would observe a series of word pairs that would be superimposed, one at a time, over background videos, and that they would later be tested on their memory of the words. They were also warned that no background videos would be shown at the final test.

Participants first completed encoding, in which each of 48 word pairs were shown in bold red text for 5 s, superimposed upon a context video. Participants then completed a judgment of learning phase, wherein half (24) of the items were shown with reinstated contexts, and the other half shown with novel contexts. During the judgment of learning phase, participants observed each item with its reinstated or novel context for 5 s, and were then presented with a blank white screen for 3 s during which they made judgments of learning on a paper and pencil response packet. When producing judgments of learning, participants were told to indicate "how confident" they were that they would be able to recall the target English word when given the cue Tagalog word. Participants were told to indicate their confidence on a 0–100 percent scale, with 100 indicating that they were "100% sure" that they would recall the target word. Before both the encoding and judgment of learning phase, participants were briefly informed of the overall procedure of the study and were warned that the final recall exam would not present contexts.

After completing the judgment of learning phase, participants worked through maze puzzles for 10 min before completing a final recall test. As in the previous phases, the recall test was presented via video, during which each cue was presented in bold red text, on a white background for 5 s without its associated target or any context video. Participants were encouraged to guess if they were unsure of the English answer, and they recorded their answers on a paper and pencil response packet.

Results

Judgments of learning. We investigated whether context change during judgments of learning decreased metamemory judgments. An ANOVA examining the effects of judgment of learning context (reinstated vs. changed, a within-subjects variable) using mean judgment of learning (between 1 and 100) as a dependent variable. A main effect of context change on judgments of learning was found, $F(1,121) = 4.175$, $p = .043$, $MSE = 22.156$, partial $\eta^2 = .033$. Context change ($M = 30.583$, $SD = 14.717$) resulted in lower judgments of learning than context reinstatement ($M = 31.814$, $SD = 15.402$; Figure 2).

Cued recall. To ensure that the changes in judgments of learning represented an illusory change in memory, we checked for corresponding differences in recall. If recall decreased with context change, then participants had a good reason to lower judgments of learning, but if recall did not decrease, participants had been fooled into thinking they had learned less depending on the context in which they judged themselves. A within-subjects ANOVA (judgment of learning context: reinstated vs. non-reinstated) analyzed final cued recall scores. There was no significant difference in recall proportion between the two context conditions (changed: $M = 35.314$, $SD = 20.934$; reinstated: $M = 33.846$, $SD = 21.973$), $F(1,121) = 1.686$, $p = .197$, $MSE = 131.560$, partial $\eta^2 = .014$. The mean proportion correct

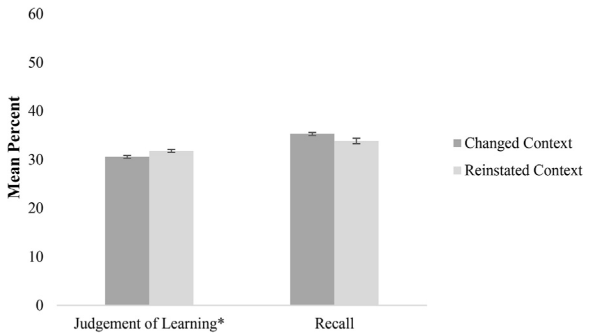

Figure 2. Predicted versus actual performance in Experiment 1. Mean judgments of learning and cued recall shown as a function of context condition during judgments of learning in Experiment 1. Error bars show within-subjects standard error (Loftus & Masson, 1994; Masson & Loftus, 2003). * $p < .05$.

recall for the context-change condition was very similar to the context-reinstated condition (Figure 2).

We investigated two measures of metacognitive accuracy (resolution and calibration) because changes in judgments of learning may lead to changes in metacognitive accuracy, or self-evaluative efficacy. Metacognitive resolution measures the association between a person's metamemory judgments and their performance, and is generally analyzed using correlational methods such as the gamma coefficient. Metacognitive calibration measures a person's overall level of self-evaluative efficacy, and is generally measured by calculating bias calibration (average JOL – average Recall) and/or magnitude calibration (i.e., the magnitude of the bias calibration). Although bias calibration can show the direction of metacognitive error, it is susceptible to analytical errors, as overconfident and underconfident errors may cancel out. Alternatively, magnitude calibration clearly shows the magnitude of error in metamemory judgments, but it cannot describe the direction of the error (Dunlosky & Bjork, 2008).

Resolution. Gamma correlations were calculated for each participant. Across context conditions mean the mean gamma was strong ($G = .293$). There was a significant difference in prediction accuracy between context conditions, $t(121) = 6.670$, $p < .001$. Items in the context-changed condition were less accurately judged (*Mean* $G = .270$, $SD = .057$) than the context-reinstated condition (*Mean* $G = .317$, $SD = .053$).

Calibration. A paired samples t test comparing bias calibration found that items in the context-changed condition were significantly less accurately, and more underconfidently evaluated ($M = -4.731$, $SD = 20.191$) than items in the context reinstated condition ($M = -2.031$, $SD = 20.138$), $t(121) = 2.250$, $p = .026$. However, an analysis of magnitude calibration found no such significant difference (changed: $M = 16.417$, $SD = 12.591$; reinstated: $M = 16.481$, $SD = 11.656$), $t(121) = .061$, $p = .951$.

Discussion

Our results indicate that reinstating encoding context during judgments of learning influenced participants into erroneously inflating their judgments of learning relative to the reinstatement condition, as there was no associated effect of reinstatement on recall. We obtained these results under conditions designed to resemble real educational situations. As a student prepares for an exam, they may reread their notes or their textbook, sparking a reinstated context condition in which they may erroneously judge themselves to be prepared for an exam, when in reality, without the context of their notes or their textbook, their performance on the test will be lower than they expected.

We also investigated how metacognitive accuracy was affected by context condition. Both resolution and bias calibration analyses suggested that context change reduced metacognitive efficacy relative to the context-reinstated condition. Intuitively, one might expect better metacognitive accuracy to be more beneficial, but that might not always be the case. First, our changed-context manipulation was intended to lower metacognitive confidence, an outcome that has been associated with increased study choice (Metcalfe, 2002; Metcalfe & Finn, 2008b). Because metacognitive judgments were generally underconfident in this experiment, such judgments are less accurate, given that confidence was reduced without an associated change in recall. Second, magnitude calibration results were not significantly affected by our context manipulations, indicating that the overall amount of metacognitive error did not change. This outcome would be expected with roughly equal amounts of under- and overconfidence, consistent with the large standard deviations and close-to-zero bias calibration results of the current data.

Experiment 2

Having demonstrated an erroneous rise in confidence resulting from manipulating context during judgments of learning, we sought to replicate our results under more conservative methods. During Experiment 1, participants were asked for a percentage (i.e., 0–100) response to indicate their judgment of learning, which may have magnified our results. Zawadzka and Higham (2015) propose that percentage based judgments of learning may offer a clear picture of how a person views their own memory. With its wider range of response, percentage based judgments of learning may be more acutely sensitive to confidence manipulations than people actually are.

Instead, binary judgments of learning have been found to produce different results from percentage-based judgments of learning (Hanczakowski et al., 2013; Metcalfe & Finn, 2013). Zawadzka and Higham (2015) investigated binary versus percentage-based judgments of learning and found that binary judgments may be better for tasks related to calibration. Further, when binary judgments of learning are used, they are recorded as ones and zeroes, indicating yes and no responses, but these judgments are averaged within participants' context condition, resulting in mean judgments of learning that are still on a percentage scale (0–100). Binary judgments of learning produce data and figures that are visually similar to percentage-based judgments of learning, but they may reflect different mental processes than percentage-based judgments of learning. Particularly, there is evidence that while percentile JOLs may reflect differences in confidence rankings for items, binary JOLs more closely reflect subjective probability of recall (Hanczakowski et al., 2013). Testing the contextual reinstatement illusion using binary judgments of learning has two possible outcomes. Binary

JOLs may still produce a metamemory illusion when given under context reinstatement, replicating our results while speaking to the robustness of this potentially detrimental illusion. Alternatively, if testing our metamemory illusion using binary JOLs eliminates its effect, then we will have identified a simple intervention that students may use to prevent themselves from falling prey to this illusory effect.

One of the most common findings in the judgment-of-learning literature is that judgments of learning are generally overconfident in predicting recall performance (Metcalfe & Finn, 2008a). In the present experiment, we found the opposite effect: judgments of learning ($M = 31.199$, $SD = 14.691$) were numerically *lower* than recall scores ($M = 34.580$, $SD = 20.531$), $t(121) = -1.96$, $p = .052$. Although recall was not at floor level, the large number of cue–target pairs in Experiment 1 may have been overwhelming to our participants, promoting lower JOLs than would normally be expected. We therefore chose to decrease the number of study items in Experiment 2. Given the potential differences between binary and percentage JOLs, as well as the contemporary nature of percentage judgment-of-learning procedures, Experiment 2 repeated the design and procedure of Experiment 1, except using binary JOLs and a new, smaller set of cue–target pairs.

Method

Participants. A total of 44 undergraduate students from introductory psychology classes at a large state university participated in this experiment in return for partial course credit. This participant number was chosen because it had more than sufficient power ($1\text{-}\beta > .80$) to obtain our previous effect size, given the correlation between our context conditions on judgments of learning (partial $\eta^2 = .033$; $r = .903$; Faul, Erdfelder, Lang, & Buchner, 2007). Participation was voluntary, and other options were available to earn equal credit. Participants were randomly assigned to treatment conditions. The number of participants in each experimental session depended on the random online enrollment of the participants and varied from 4–10 participants per session.

Materials. A new set of 24 Tagalog–English word pairs were chosen as memory targets. Each of the 24 English target words were concrete nouns, with an average frequency of 5237.46 (www.wordandphrase.info), and an average character length of 5.9 (Gardner & Davies, 2013). All Tagalog targets were pronounceable using English conventions, with an average character length of 5.80. Thirty-six context videos were randomly selected from the previous study to be paired with the new set of items. Twelve contexts were presented at encoding and during judgments of learning as reinstated contexts, 12 were presented only at encoding as initial contexts, and 12 were presented only during judgments of learning as novel contexts. Again, all obvious relationships between critical word pairs and background context videos were avoided, and the word pairs were superimposed over the videos using the same technique as described for Experiment 1.

In addition to the new items, the window between each item was also widened to 5 s during the encoding, judgment of learning, and test phases. Participants were presented with 5 s of a blank white screen between to avoid any rollover context effects, and to allow participants to make thoughtful decisions during the judgment of learning phase.

When producing judgments of learning, participants were given a set of instructions for producing binary judgments of learning: "You will now provide a yes/no response after each word pair indicating your ability to recall the critical word in 10 minutes." Participants were told that a "yes" response indicated that they would be able to recall the response word in 10 min, whereas a "no" response indicated they would not be able to recall the response word in 10 min. Participants were told that they would have 5-s to make each judgment and that the instructions for what merited a "yes" or "no" response would appear after each word was presented. They were again told that no context video scenes would be shown at final test.

Judgments of learning were recorded as ones and zeroes, representing yes and no responses. These judgment of learning scores were averaged for each participant and context condition, and multiplied by 100 to produce mean judgments of learning that were analyzed in the same manner as percentage-based judgments of learning.

Procedure. The procedure described for Experiment 1 was again used in Experiment 2. See Procedure section in Experiment 1 above for details.

Results

Judgments of learning. We investigated whether our results replicated those of Experiment 1, that context change during judgments of learning produced lowered metamemory judgments relative to context reinstatement, despite using a binary judgment of learning paradigm. A within-subjects ANOVA examining the effects of context condition (change vs. reinstated) on judgments of learning was performed. A main effect of context condition on judgments of learning was found, $F(1,43) = 11.134$, $p = .002$, $MSE = 116.256$, partial $\eta^2 = .206$. Judgments of learning were significantly lower for the context-change condition ($M = 35.322$, $SD = 16.606$) than for the context-reinstated condition ($M = 42.992$, $SD = 19.163$; Figure 3).

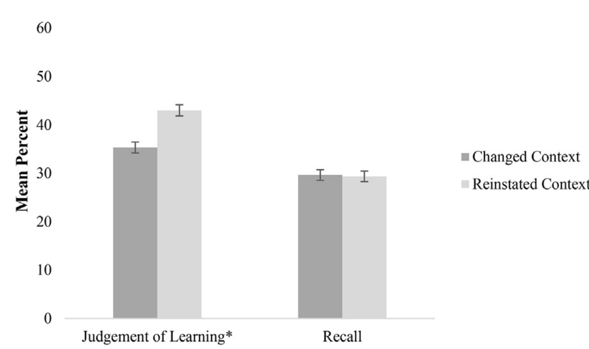

Figure 3. Predicted versus actual performance in Experiment 2. Mean judgments of learning and cued recall shown as a function of context condition during judgments of learning in Experiment 2. Error bars show within-subjects standard error (Loftus & Masson, 1994; Masson & Loftus, 2003). * $p < .05$.

Cued recall. The main effect of context condition on cued recall was not significant, $F(1,43) = .017$, $p = .898$, *MSE* 1.776, partial $\eta^2 < .001$. The mean proportion recalled was nearly the same for the context-changed condition ($M = 29.640$, $SD = 20.672$) as for the context-reinstated condition ($M = 29.356$, $SD = 19.625$; Figure 3).

Resolution. Gamma correlations were calculated for each participant. Across context conditions the mean gamma was strong ($G = .319$). There was a significant difference in prediction accuracy between context conditions, $t(43) = 2.497$, $p = .016$. Items in the context-changed condition were more accurately judged (*Mean* $G = .366$, $SD = .240$) than the context-reinstated condition (*Mean* $G = .273$, $SD = .287$).

Calibration. A paired samples *t* test comparing bias calibration (i.e., mean JOL – mean Recall) found that items in the context-changed condition were significantly more accurately and less overconfidently evaluated ($M = 5.682$, $SD = 23.017$) than items in the context-reinstated condition, ($M = 13.636$, $SD = 27.217$), $t(43) = 3.070$, $p = .004$. An analysis of magnitude calibration also found that items in the context-changed condition had less metacognitive error ($M = 18.940$, $SD = 13.992$) than items in the context-reinstated condition ($M = 24.242$, $SD = 18.157$), $t(43) = 2.128$, $p = .039$.

Discussion

As in Experiment 1, a contextually influenced judgment-of-learning effect was found in Experiment 2. The effect of context during judgments of learning was significant; paired associates shown in changed contexts during judgments of learning were judged to be less memorable than pairs tested in reinstated contexts. Final recall scores were not affected by context during judgments of learning. Thus, the effect of context change on judgments of learning again reflected a metamemory illusion, affecting metamemory judgments but not memory performance.

Our analyses also found that changing context during judgments of learning improved metacognitive accuracy, as measured by resolution, bias calibration, and magnitude calibration. Participants in Experiment 2 were generally overconfident; therefore, changing contexts, which lowered confidence, improved metacognitive accuracy. Because fewer study items were used in Experiment 2 relative to Experiment 1, the apparent difficulty of the task was lower, resulting in more overconfident judgments in Experiment 2. Because Experiment 2 involved mostly overconfident metacognitive error, our confidence-reducing paradigm resulted in improvements in both resolution and calibration.

Experiment 3

Some studies have found that factors manipulated within-subjects affect metamemory judgments, but when those factors are manipulated between subjects, they do not have the same effect. For example, Koriat, Bjork, Sheffer, and Bar (2004) asked participants to predict their performance after 1 min, 10 min, 30 min, 1 h, 1 day, or 1 week. In the between-subjects conditions, participants were asked about only one retention interval, whereas in the within-subjects condition participants were asked about all of the retention intervals. Koriat et al. (2004) concluded that because they were asked for judgments about all of the retention intervals, participants in the within-subjects condition likely took the retention intervals into account when making their judgments of learning, whereas those in the between-subjects condition, who were asked about only a single retention interval, were less likely to take the retention interval into account in their judgments. Thus, the within-subjects design called participants' attention to differences among the different treatment conditions.

The results of Experiments 1 and 2 may have been due to calling participants' attention to differences between changed versus reinstated context conditions during judgments of learning, because context had been manipulated within-subjects so far. Would participants still make lowered judgments of learning when observing exclusively changed versus reinstated contexts?

Because we found a contextually driven metamemory illusion in Experiments 1 and 2 using a within-subjects manipulation, we wanted to examine the effect of participants' experiencing exclusively changed or reinstated contexts during encoding and judgment of learning phases on metamemory judgments and accuracy. Such conditions may also better represent an applied setting, as students would most probably not naturally vary the reinstatement of context under which they prepare for an exam.

Method

Participants. A total of 84 undergraduate students at a large state university participated in Experiment 3 in return for partial course credit. Participation was voluntary, and other options were available to earn equal credit. This number of participants should provide sufficient power ($1\text{-}\beta > .80$) to detect a conservative estimate of the true effect size (partial $\eta^2 = .100$) in a between-subjects paradigm (Faul et al., 2007). Participants self-enrolled in the experiment, and group assignment was based on time of testing. The number of participants in each experimental session depended upon the random number of participants who opted to enroll for a session, with a minimum of two and a maximum of 15 participants enrolling per session. There were 44 participants randomly assigned to the context-change context condition, and 40 participants to the context-reinstated condition. Results for two participants in the context-change condition were excluded for failure to follow directions, leaving 38 participants in the context-change condition.

Materials and procedure. Experiment 3 closely followed the procedure used in Experiment 2, with the main difference being that Experiment 3 used a between-subjects design. Although the same 20 items from Experiment 2 were used, 10 new context videos were added to the pool, allowing for some participants to view exclusively reinstated contexts during judgments of learning in the reinstatement condition, and for other participants to view exclusively novel contexts videos during judgments of learning in the non-reinstatement condition.

Design. Context change during judgments of learning (changed during judgments of learning vs. reinstated during judgments of learning) was manipulated as a between-subjects variable; each word pair was shown with the same video context

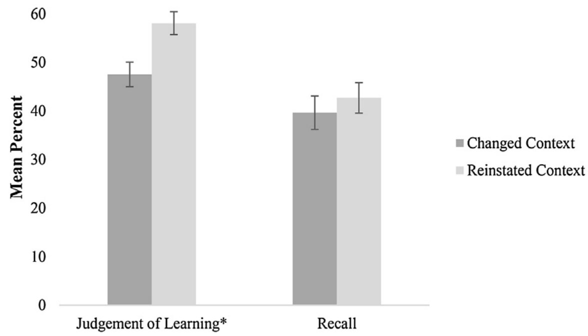

Figure 4. Predicted versus actual performance in Experiment 3. Mean judgments of learning and cued recall shown as a function of context condition during judgments of learning in Experiment 3. Error bars show standard error. * $p < .05$.

that been associated with that pair at encoding (in the reinstatement condition), or with a new context video, not previously seen in the experiment (in the non-reinstatement condition).

Results

Judgments of learning. We investigated whether context change lowered judgments of learning relative to context reinstatement, despite using a binary response scale and manipulating reinstatement between subjects. An ANOVA examined the effects of context condition on judgments of learning. A significant main effect of context condition on judgment of learning frequency was found, $F(1,80) = 8.523$, $p = .005$, $MSE = 266.962$, partial $\eta^2 = .096$. Judgments of learning were lower for the context-change condition ($M = 47.533$, $SD = 15.771$) than for the context-reinstated condition ($M = 58.097$, $SD = 16.813$; see Figure 4).

Cued recall. An ANOVA examining the effect of context condition on percent of final recall was peromed; the effect was not significant, $F(1,80) = .426$, $p = .516$, $MSE = .019$, partial $\eta^2 = .005$. Recall for the context-change condition ($M = 39.660$, $SD = 21.225$) did not differ significantly from that of the context-reinstated condition ($M = 42.700$, $SD = 20.829$; Figure 4). Since reinstatement condition did not affect recall, we can conclude that context reinstatement can produce a metamemory illusion directly causing people to erroneously believe they know more than they actually do, relative to a changed context condition.

Resolution. Gamma correlations were calculated for each participant. The mean gamma correlation across conditions was weak ($G = .074$). There was no significant difference in prediction accuracy between participants in the context changed (*Mean* $G = .030$, $SD = .206$) and context reinstated condition (*Mean* $G = .112$, $SD = .226$), $t(80) = 1.717$, $p = .09$.

Calibration. An independent samples t test found that there was no significant difference between participants in the context-changed condition and the context-reinstated condition on either bias calibration (changed: $M = 7.875$, $SD = 21.465$; reinstated: $M = 15.401$, $SD = 16.934$), $t(80) = 1.773$, $p = .080$, or magnitude calibration (changed: $M = 18.415$, $SD = 13.274$; reinstated: $M = 19.099$, $SD = 12.498$), $t(80) = .240$, $p = .811$.

General Discussion

Students, known to be inaccurate monitors of their own learning (Bol, Hacker, O'Shea, Allen, 2005; Everson & Tobias, 1998; Saenz et al., 2017), may overestimate their learning because the familiarity of their notes, textbooks, and study locations leads them unwittingly into feeling so confident that they end their study efforts prematurely. Three experiments found evidence of a context reinstatement illusion, that is, the illusion that one's learning is judged to be better with encoding contexts reinstated rather than changed. Thus, overconfidence in learning can be reduced if judgments of learning are made in unfamiliar contexts. When subsequent recall was tested without context cues, recall performance was no different whether judgments of learning had been done with reinstated or changed contexts. Delayed judgments of learning made with changed contexts were lower in all three experiments than judgments of learning made for items in reinstated contexts. As with other metamemory illusions, such as the illusion of knowing (e.g., Glenberg, Wilkinson, & Epstein, 1982), the immediate/delayed judgment of learning effect (e.g., Dunlosky & Nelson, 1992, 1997), the cue familiarity effect (e.g., Metcalfe et al., 1993; Miner & Reder, 1994), and the font size effect (e.g., Besken & Mulligan, 2013; Undorf & Erdfelder, 2015; Undorf et al., 2017), the context reinstatement illusion shows how a factor that influences metamemory judgments may not affect memory performance.

Explaining Context-Based Confidence

What caused the decrease in judgments of learning as a function of context change? Here, we consider four possible explanations: a contextual crutch hypothesis, a memory strength hypothesis, a belief hypothesis, and a fluency hypothesis. The contextual crutch hypothesis (Smith & Handy, 2014, 2015) states that context cues may assist, or even replace designated memory cues (in this case, Tagalog cue words). For example, in learning to associate names with faces, people might rely on context cues when face–name associations are too weak (Smith & Handy, 2014). If a memory test occurs without context cues in such cases, recall may fail. In the present study, the contextual crutch hypothesis suggests that participants may use the video contexts as "crutches" to cue memory of target words. To avoid this pitfall, we *repeatedly* warned our participants that there would be no video contexts on the final memory test. These warnings preceded both the encoding and judgment-of-learning phases of all three experiments; despite these warnings, however, participants may have relied on the context videos (as contextual crutches) during encoding. The context crutch presented during judgments of learning would inflate memory judgments, while its absence at test would result in lower recall. A contextual crutch explanation must therefore assume that the explicit warnings, stating that there would be no videos at test, were ignored.

The memory strength hypothesis posits that reinstated contexts increase judgments of learning because judgments of learning reflect memory strength, and repetitions in the same context may produce stronger memory traces than repetitions in different contexts. This hypothesis assumes that the judgment of learning, made with a reinstated or changed context, acts as a memory strengthening event. Consistent with this explanation are findings that repetitions in a constant, unchanging context produced better recall than repetitions in varied contexts (e.g., Bellezza & Young, 1989; Isarida, 2005; Isarida & Isarida, 2010; Smith & Handy, 2014, Experiment 4). Although *retrieval practice* in varied contexts leads to better recall (if recall is unsupported by context cues; Smith & Handy, 2014, 2015), *repetitions* may yield better recall if they are done in constant contexts, relative to varied contexts, (although see Smith, 1982, 1984; Smith et al., 1978; Smith & Rothkopf, 1984 for the opposite pattern). This memory strength hypothesis is, however, inconsistent with the recall results of the present experiments, all three of which showed that context change during judgments of learning had no effect on final recall performance, even though it did affect judgments of learning.

The belief hypothesis (e.g., Bjork et al., 2013; Hu et al., 2015; Mueller et al., 2013, 2014) has challenged the fluency hypothesis (Alter & Oppenheimer, 2009; Susser & Mulligan, 2015), suggesting that metamemory judgments are influenced by one's beliefs instead of memory fluency. In this case, participants believe that repetitions in reinstated contexts benefit memory more than repetitions in non-reinstated contexts. At the time of judgments of learning participants may realize they were observing some targets repeated in reinstated contexts and others in new contexts, and alter judgments of learning appropriately. The belief hypothesis is consistent with the results of Experiments 1 and 2, in which participants made judgments of learning for both changed and reinstated context word pairs; the contrast of the two treatments during judgments of learning could have tapped into participants' beliefs and personal metacognitive theories (e.g., Koriat et al., 2004) about how contexts might affect memory. The results of Experiment 3, however, weaken this belief explanation; in that experiment, context reinstatement during judgments of learning was manipulated between-subjects, prohibiting participants from comparing the two treatment conditions; the context reinstatement illusion was evidenced in that experiment, nonetheless.

The retrieval fluency hypothesis states that judgments of learning reflect the fluency with which correct responses come to mind. Reinstatement of video contexts can have a powerful effect on recall (Smith & Handy, 2015; Smith & Manzano, 2010; Smith et al., 2014; Staudigl & Hanslmayr, 2013), so reinstating context for some items during judgments of learning may have led participants to more fluently retrieve those targets, causing increased judgments of learning on reinstated items. Some may argue that, because participants were presented with both cues and targets during judgments of learning, participants were not required to perform retrieval. Although it may still be that participants perform a type of retrieval when presented with both target and cue (e.g. "checking" for a mental connection between the cue–target pair), it may be premature to suggest that retrieval fluency specifically may account for our results. Indeed, because the entire cue–target pair was presented at JOL, there is little evidence to suggest that recall may have occurred during JOL. Instead, context reinstatement at JOL may have led to a more fluent perception of the cue–target pair. Although we lack a critical test of this hypothesis, none of our results are inconsistent with the fluency hypothesis.

Several questions arise in regard to the present results. First, it is important to better explicate the mechanism(s) that produced this metamemory illusion, be it belief, fluency, or an alternative mechanism. Second, can we infer that lowered judgments of learning will translate into increased study choices (Dunlosky & Hertzog, 1998; Metcalfe, 2002; Metcalfe & Finn, 2008b)? A direct assessment of study choice behavior as a function of context is needed to establish an applied effect of our manipulation. Further investigations should directly observe study choice or study choice reports to confirm the proposed effects of context change. Third, why was final recall unaffected by the previous judgment of learning context? A judgment of learning trial may act as a learning trial (e.g., Mitchum et al., 2016; Naveh-Benjamin & Kilb, 2012; Nguyen & McDaniel, 2016; Soderstrom et al., 2015; Spellman & Bjork, 1992), and contextual variation at encoding, corresponding to our new-context judgment of learning condition, has been found to improve memory in some cases (e.g., Smith & Handy, 2014, 2016). It may be that, as found by Smith and Handy (2016), the varied context benefit is seen only after multiple practice trials. Finally, it is not clear whether context reinstatement increases judgments of learning, or if context change reduces judgments of learning. Further research is needed to properly address these questions.

Metacognitive Accuracy

Measurements of metacognitive accuracy such as such as resolution, bias calibration, and magnitude calibration may help us understand what effects context manipulation had on judgments of learning. Results from Experiments 1 and 2 indicated that context change can either improve or worsen metacognitive accuracy, depending on the perceived difficulty of the target task. Because context change improved metacognitive accuracy only in generally overconfidence-producing conditions (as in Experiment 2), our findings indicate that context manipulations consistently affect judgments of learning, but not calibration, per se. Whereas the value of lowered JOL confidence is to increase study choices, in general, the benefit of improved metacognitive accuracy is clearer when there is a limited amount of study time, and students must allocate time to the material that is most in need of improvement.

Could the differences in metacognitive accuracy change in Experiments 1 and 2 have been caused by the different dependent variables used therein (percentile vs. binary judgments of learning)? It is not apparent that judgment type can affect metacognitive accuracy (either calibration or resolution) in a single trial paradigm. Hanczakowski et al. (2013) found that binary judgments of learning can produce different results than percentile judgments of learning after at least one round of practice trials (e.g., the underconfidence with practice effect; Koriat, 1997). Because learning and studying often involve multiple practice trials, additional research is necessary to more thoroughly explicate these issues.

Conclusions

Three experiments found that reinstating encoding contexts at the time that delayed judgments of learning were made increased judgments of learning, but had no effect on subsequent memory performance, causing people to erroneously believe they know more than they do. We investigated this effect with both binary and percentage-based judgments of learning, as well as in between- and within-subjects designs, and we found that context change consistently reduced metamemory judgments relative to a reinstated context condition. Such a manipulation may be critical in learners, as they often re-read their notes or participate in review sessions in class, situations wherein circumstances are reinstated that may unreasonably inflate students' judgments of learning. Our data reinforce a number of existing studying recommendations, such as varying study location and time, as our findings suggest that not only may these be more productive study strategies, they may help students make better study choices. Alternatively, our data suggest that students should be particularly wary of studying by reviewing notes or rereading text material, because such strategies may lead to lowered study choice, and underpreparation. By studying and evaluating themselves in novel contexts, thereby reducing their confidence, learners may be driven to study more than they might otherwise. Educators wishing to reduce overconfidence and increase study choice may encourage students to prepare for exams under varied context conditions as well.

Conflict of Interest Statement

The authors declare no conflict of interest.

Author Contributions

Both authors jointly conceived and designed the studies, analyzed and interpreted the data, and wrote the manuscript.

References

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

Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 990–997.

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*, 55–68.

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*, 897–903.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.

Bol, L., Hacker, D. J., O'Shea, P., & Allen, D. (2005). The influence of overt practice, achievement level, and explanatory style on calibration accuracy and performance. *The Journal of Experimental Education*, *73*, 269–290.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah, NJ: Erlbaum.

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

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language*, *36*, 34–49.

Dunlosky, J., & Bjork, R. A. (2008). *Handbook of metamemory and memory*. Psychology Press.

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52*, 526–534.

Everson, H. T., & Tobias, S. (1998). The ability to estimate knowledge and performance in college: A metacognitive analysis. *Instructional Science*, *26*, 65–79.

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

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, *35*, 305–327.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, *10*, 597–602.

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.

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language*, *69*, 429–444.

Hanczakowski, M., Zawadzka, K., Collie, H., & Macken, B. (2017). Metamemory in a familiar place: The effects of environmental context on feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 59–71.

Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 39–62). New York, NY: Oxford University Press.

Hockley, W. E. (2008). The effects of environmental context on recognition memory and claims of remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1412–1429.

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*, 1–11.

Isarida, T. (2005). Study-time effect on free recall within and out of context. *Memory*, *13*, 785–795.

Isarida, T., & Isarida, T. K. (2010). Effects of simple-and complex-place contexts in the multiple-context paradigm. *The Quarterly Journal of Experimental Psychology*, *63*, 2399–2412.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349.

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

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

Maki, R. H. (1999). The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1011–1023.

Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*, 203.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131*, 349–363.

Metcalfe, J., & Finn, B. (2008a). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1084–1097.

Metcalfe, J., & Finn, B. (2008b). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*, 174–179.

Metcalfe, J., & Finn, B. (2013). Metacognition and control of study choice in children. *Metacognition and Learning*, *8*, 19–46.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 851–861.

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

Miner, A. C., & Reder, L. M. (1994). A new look at feeling of knowing: Its metacognitive role in regulating question answering. In J. Metcalfe, & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47–70). Cambridge: MIT Press.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*, 200.

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.

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.

Naveh-Benjamin, M., & Kilb, A. (2012). How the measurement of memory processes can affect memory performance: The case of remember/know judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 194–203.

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

Nelson, T. O., & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science*, *3*, 317–318.

Nguyen, K., & McDaniel, M. A. (2016). The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied*, *22*, 59.

Nichols, J. H., & Smith, S. M. (2011). Inflating judgments of learning with video context reinstatement. In *Presented at the meeting of the Psychonomic Society*

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–451.

Saenz, G. D., Geraci, L., Miller, T. M., & Tirso, R. (2017). Metacognition in the classroom: The association between students' exam predictions and their desired grades. *Consciousness and Cognition*, *51*, 125–139.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1074–1083.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, *6*, 132–137.

Schwartz, B. L., & Frazier, L. D. (2005). Tip-of-the-tongue states and aging: Contrasting psycholinguistic and metacognitive perspectives. *The Journal of General Psychology*, *132*, 377–391.

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 460–471.

Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition*, *10*, 405–412.

Smith, S. M. (1984). A comparison of two techniques for reducing context-dependent forgetting. *Memory & Cognition*, *12*, 477–482.

Smith, S. M. (1988). Environmental context-dependent memory. In G. Davies, & D. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13–33). New York, NY: Wiley.

Smith, S. M. (1994). Theoretical principles of context-dependent memory. In P. Morris, & M. Gruneberg (Eds.), *Aspects of memory: Theoretical aspects* (2nd ed., pp. 168–195). New York, NY: Routledge Press.

Smith, S. M. (2007). Context and human memory. In H. L. Roediger III, Y. Dudai, & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts* (pp. 111–114). Gosport, Hampshire, Great Britain: Oxford University Press.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353.

Smith, S. M., & Handy, J. D. (2014). Effects of varied and constant environmental contexts on acquisition and retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1582–1593.

Smith, S. M., & Handy, J. D. (2015). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory*, *24*, 1134–1141.

Smith, S. M., & Handy, J. D. (2016). The crutch of context-dependency: Effects of contextual support and constancy on acquisition and retention. *Memory*, *24*, 1134–1141.

Smith, S. M., Handy, J. D., Angello, G., & Manzano, I. (2014). Effects of similarity on environmental context cueing. *Memory*, *22*, 493–508.

Smith, S. M., & Manzano, I. (2010). Video context-dependent recall. *Behavior Research Methods*, *42*, 292–301.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction*, *1*, 341–358.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 553–558.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*, 315–317.

Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*, *23*, 1101–1106.

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, *22*, 1014–1019.

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., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, *92*, 293–304.

Zawadzka, K., & Higham, P. A. (2015). Judgments of learning index relative confidence, not subjective probability. *Memory & Cognition*, *43*, 1168–1179.

Received 25 August 2017;
received in revised form 14 May 2018;
accepted 13 July 2018
Available online 1 September 2018