Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 2, 553–558

© 2014 American Psychological Association
0278-7393/15/$12.00 http://dx.doi.org/10.1037/a0038388

RESEARCH REPORT

Judgments of Learning as Memory Modifiers

Nicholas C. Soderstrom, Colin T. Clark,
Vered Halamish, and Elizabeth Ligon Bjork
University of California, Los Angeles

A frequent procedure used to study how individuals monitor their own learning is to collect judgments of learning (JOLs) during acquisition, considered to be important, in part, because such judgments are assumed to guide how individuals allocate their future learning resources. In such research, however, a tacit assumption is frequently made: Namely, that asking for such metacognitive judgments does not affect the learning process per se. In 3 experiments, the present research addressed the accuracy of this assumption and tested a possible account—based on aspects of Koriat's cue-utilization approach to JOLs (Koriat, 1997) and de Winstanley, Bjork, and Bjork's (1996) transfer-appropriate multifactor account of generation effects—for why the mere act of making JOLs might enhance later memory for the information so judged. Potential implications of the present findings for the future conduction of research using metacognitive measures as well as for students studying for exams is discussed.

Keywords: judgments of learning, memory, metacognition, metamemory, monitoring

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

For practical as well as theoretical reasons, the ability of individuals to judge their own learning during acquisition is of interest, in part, because such judgments can play a central role in determining how individuals allocate their future learning resources (for reviews, see Bjork, 1999; Bjork, Dunlosky, & Kornell, 2013; Soderstrom, Yue, & Bjork, in press). When, for example, individuals are asked to monitor their degree of learning during initial study and are then asked to make decisions about future study activities (e.g., which items they want to study again and for how long), they tend to choose items they previously judged to be more difficult or to have learned less well (e.g., Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; Soderstrom & Bjork, 2014; for a review, see Son & Metcalfe, 2000). Furthermore, as demonstrated by Kornell and Metcalfe (2006), learners tend to perform better when allowed to restudy those items they choose to restudy versus ones they do not.

To examine learners' assessments of their own learning, researchers often use a straightforward method: They ask participants to make judgments of their learning (JOLs) during acquisition. To illustrate, participants might be given a list of cue-target word pairs to learn (e.g., English–Spanish equivalents; pairs of synonyms) and, following presentation of each pair, be asked to judge the likelihood they will be able to recall the target-word given the cue-word on a later test. Experiments using such a procedure have shown that participants' JOLs can, under many circumstances, be fairly accurate in predicting future memory performance (e.g., Dunlosky & Nelson, 1994; Lovelace, 1984).

Importantly, and of most relevance to the present investigation, there is often a tacit assumption in research using such metacognitive measures that asking for such judgments does not affect the learning process per se. Consistent with this assumption, some studies in which no-judgment conditions have been included have not found a difference on later memory performance between such conditions versus those requiring JOLs (e.g., Benjamin, Bjork, & Schwartz, 1998; Kelemen & Weaver, 1997; Tauber & Rhodes, 2012). Other findings, however, have suggested that the very act of making JOLs may trigger processes in the learner that alter the memorial representation of the items being judged, often in ways that make them more recallable in the future—analogous to the way that retrieving an item not only reveals that it was accessible, but also makes that item more recallable in the future than it would have been otherwise (Bjork, 1975; Spellman & Bjork, 1992; see Roediger & Karpicke, 2006, for a review). Dougherty, Scheck, Nelson, and Narens (2005; Experiment 2) and Zechmeister and Shaughnessy (1980) reported better retention of items that participants were required to judge versus those they did not judge. These findings are difficult to interpret, however, because available study time was longer for the items participants judged versus those they simply studied. Consequently, the memory advantage for JOL items could have been driven by additional study time rather than any processes engendered by making the JOLs.

This article was published Online First December 22, 2014.

Nicholas C. Soderstrom, Colin T. Clark, Vered Halamish, and Elizabeth Ligon Bjork, Psychology Department, University of California, Los Angeles.

Colin T. Clark is now at the Design Media | Arts Department, University of California, Los Angeles and Santa Monica College; Vered Halamish is now at the Psychology Department, The Max Stern Yezreel Valley College and Tel-Aviv University.

Grant 29192G from the James S. McDonnell Foundation supported this research. We thank Robert A. Bjork and members of CogFog for insightful comments.

Correspondence concerning this article should be addressed to Nicholas C. Soderstrom, Department of Psychology, 1285 Franz Hall, University of California, Los Angeles, CA 90095-1563. E-mail: nsoderstrom@psych.ucla.edu

Our goals in the present research were twofold. First, we wanted to test in a straightforward manner whether simply requiring learners to make JOLs during initial acquisition or study can, indeed, enhance the later memorability of the judged information when time on task is controlled. Second, we wanted to test a conjecture as to why the requirement to make JOLs during study or acquisition might have such a consequence, a conjecture that is derived from aspects of Koriat's cue-utilization approach to JOLs (Koriat, 1997) and de Winstanley et al.'s (1996) transfer-appropriate multifactor account of generation effects.

According to Koriat's (1997) cue-utilization approach, the learner is assumed to draw upon cues in making JOLs, rather than directly accessing the strength of an item's representation in memory. Such cues include *intrinsic* cues (e.g., the perceived association between cues and targets when both are present during study) and *extrinsic* cues (e.g., the number of study repetitions and presentation durations as well as encoding operations), but with learners tending, at least during original study, to be more sensitive to intrinsic cues in making JOLs. The predictive accuracy of such cues will be high, Koriat argued, to the extent that performance on the criterion test is dependent on the same cues—say, for example, in the case where JOLs and later memory performance are affected in similar ways by cue-target associative strength.

Similarly, de Winstanley et al.'s (1996) account of generation effects assumes that the act of generation strengthens whatever type of information is used by the learner to complete the generation task, meaning that generation effects will be observed if the later criterion test requires access to that strengthened information. To illustrate, assume learners are given a list of cue-target pairs to learn with some pairs presented intact (e.g., true–false), whereas, for others, the target-word has to be generated (e.g., hot–???) with learners instructed to generate the opposite of the cue-word. As a consequence, cue-target relational information would be strengthened for generated pairs and, thus, a memory advantage for them versus the simply read pairs would be expected on a later cued-recall test, given its sensitivity to such information. Indeed, such generation effects have been shown to be robust and reliable (e.g., Slamecka & Graf, 1978; for a review, see Bertsch, Pesta, Wiscott, & McDaniel, 2007).

In the present research, we combined the ideas of these two approaches by further assuming that when a learner is required to make a JOL, the act of doing so can result in the strengthening of the cues or information used as the basis of arriving at such a judgment—similar to the strengthening that occurs for information used to complete a generation task. If so, later memory performance should be higher for items judged than for items not judged on tests sensitive to the strengthened information. Thus, even when some cue-target pairs are generated whereas others are only read, the generation effect itself could be eliminated, or at least substantially attenuated, if the nongenerated read pairs receive JOLs.

In summary, the present research was concerned with two primary questions. First, when controlling for the presence of confounding variables that in previous research may have favored the later retention of to-be-judged items, will making JOLs still enhance subsequent memory performance? Second, if such an enhancement is found, can it be attributed to the strengthening of the information used in making the JOLs (e.g., strengthening of the cue-target relational information existing between two pairs of words)? We report three experiments addressing these questions, with the first two (Experiments 1a and 1b) focusing on the issue raised in the first question, and the second focusing more on the issue of the second question.

Experiment 1a

In Experiment 1a, two groups of participants studied the same list of cue-target word pairs for the same duration, but with one group required to make a JOL to each pair while the other group made no JOLs. In addition, because previous research has indicated the degree of associative relatedness between members of a pair to be an intrinsic cue used by learners for making JOLs (e.g., Arbuckle & Cuddy, 1969; Soderstrom & McCabe, 2011), the list we presented contained two types of cue-target word pairs: ones that were strongly related and ones that were weakly related. Using such pairs (a) ensured that the difference in associative relationship between pairs would be salient to participants and, thus, the cue most likely to be used by participants in making their JOLs, and (b) allowed us to test whether the act of making a JOL based on the intrinsic cue of relatedness would result in differing amounts of strengthening of the memorial representation of that relationship. Specifically, we expected a meaningful relationship to be fairly easy to discern, even given a brief exposure, for the strongly related pairs and, thus, for this associative information to be strengthened in the process of making a JOL. In contrast, we expected that participants would be relatively less likely to be able to come up with a meaningful relationship during study of the weakly related pairs and, thus, there would be little, if any, consequent strengthening of the cue-target relational information for such pairs. If such dynamics are in play during study, then we would expect to observe an interaction between the requirement to judge and the degree of associative relationship of the cue-target pairs, with judged strongly related pairs being better remembered on a later cued-recall test than nonjudged strongly related pairs, whereas there would be little or no difference in the later memorability of the judged and nonjudged weakly related pairs.

Method

Participants, materials, and procedure. Forty undergraduate students (27 women, 13 men) participated for partial credit in a course taught at the University of California, Los Angeles. During the study phase, 60 cue-target word pairs were studied for a later memory test. Of these pairs, half were strongly related (e.g., *blunt–sharp*; *pledge–promise*) and half were weakly related (e.g., *boxer–terrible*; *mercy–justice*), according to the MRC Psycholinguistic Database (Coltheart, 1981).[1] These pair types were randomly intermixed during study, with pairs exposed individually for 8 s on a computer screen. During exposure of each pair, the 20 participants assigned to the judgment condition were required to make a JOL to that pair by estimating the likelihood, on a 0–100% scale, of successfully recalling the pair on a later test, being prompted to do so half way through the exposure duration (i.e.,

[1] Specifically, according to the MRC database, the mean proportion of occurrence was .32 and .01 for strongly and weakly related cue-target pairs, respectively; meaning that, on average, 32% and 1% of respondents gave that response as a free association to the cue for the strongly related and weakly related pairs, respectively.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

after 4 s); whereas, the 20 participants assigned to the no-judgment condition made no JOLs during the 8-s exposure of each pair. Following the presentation of all pairs, participants played Tetris for 3 min and were then given a cued-recall test, in which they were presented with a sheet of paper containing all of the cue-words (in a randomized order) and asked to supply the appropriate target-word for each cue, being allowed 2.5 min for this test.

Results and Discussion

To test whether, as hoped, the construction of our materials would make the relatedness between cue-target words a salient intrinsic cue and, for participants required to make JOLs, the one on which they would largely base their JOLs, we compared the average magnitude of JOLs made to strongly related pairs (M = 61.88) versus weakly related pairs (M = 42.94) via a planned-comparison t test, which revealed this difference to be significant in support of our assumption, $t(19) = 9.92$, $p < .001$, $d = 1.11$.

Figure 1A presents cued-recall performance for Experiment 1a. We performed a 2 (cue-target association: high vs. weak) × 2 (judgment condition: judgment vs. no-judgment) mixed-design analysis of variance (ANOVA), which revealed a significant main effect of cue-target association with strongly related pairs being recalled at a higher rate than weakly related pairs (M = .61 vs. .34), $F(1, 38) = 161.27$, $p < .001$, $\eta_p^2 = .81$, but no significant main effect of judgment condition, $F(1, 38) = 2.42$, $p = .13$.

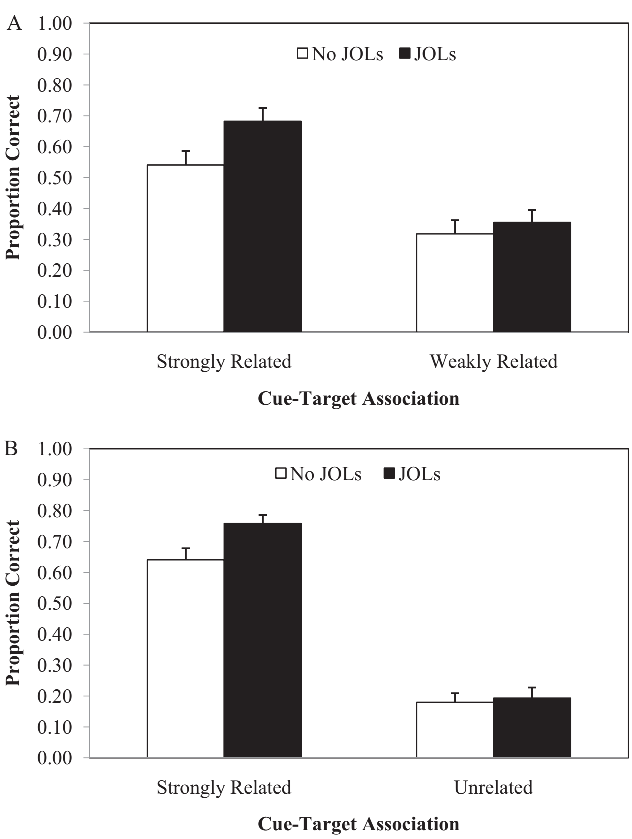

Figure 1. Mean cued-recall performance as a function of judgment condition and cue-target association in Experiments 1a (A) and 1b (B).

Importantly, however, a significant interaction between cue-target association and judgment condition did obtain, $F(1, 38) = 5.61$, $p = .02$, $\eta_p^2 = .13$. Consistent with our expectations, planned-comparisons confirmed cued-recall to be greater for judged strongly related pairs than for nonjudged strongly related pairs, $t(38) = 2.26$, $p = .03$, $d = .74$, but not for judged versus nonjudged weakly related pairs ($p > .05$).

Thus, the pattern of results obtained in Experiment 1a seems clearly to indicate that the making of JOLs enhances subsequent memory performance even when controlling for the presence of the confounding variable (i.e., different duration times) that could be thought to favor such a result in previous research. Nonetheless, it seemed prudent to see if this result could be replicated in a second highly similar experiment using a different population of participants and slightly different materials.

Experiment 1b

Method

Participants, materials, and procedure. Much of the design and procedure of Experiment 1b was the same as that of Experiment 1a. Participants studied cue-target word pairs for a later memory test with a random half of the participants assigned to a judgment condition and the other half assigned to a no-judgment condition. As in Experiment 1a, each pair was exposed for 8 s on a computer screen, with those participants in the judgment condition being required to make a JOL to each pair (using a 100% scale) when prompted to do so half way through the exposure duration, whereas those participants in the no-judgment condition made no JOLs for any pairs. Also, as in Experiment 1a, participants played Tetris for 3 min following the presentation of all pairs before administration of the final cued-recall test.

The differences between Experiments 1a and 1b were the participant population, the exact cue-target pairs used, and the method of administering the final cued-recall test. Whereas in Experiment 1a, undergraduates participated in the laboratory for course credit, the 60 participants in Experiment 1b (29 women, 31 men; median age = 32 years, range = 18–67 years) were paid $1.00 for completing the experiment and were recruited online via Amazon's Mechanical Turk, a site that allows users to complete tasks for money. Studies using Mechanical Turk have replicated the results of many laboratory-based methods of data collection (see, e.g., Mason & Suri, 2012). Participants were fluent English speakers and lived in the United States. Also, rather than being strongly and weakly related cue-target pairs selected from Coltheart's (1981) MRC Psycholinguistic Database, the cue-target pairs in Experiment 1b were constructed to be strongly related (e.g., *loaf–bread*) and unrelated (e.g., *sack–flag*) pairs according to Nelson, McEvoy, and Schreiber (1998). Lastly, rather than using a paper-and-pencil final cued-recall test, participants in Experiment 1b were presented with each of the 60 cue-words individually for 5 s in a randomized order and asked to type in the target-word within that time.

Results and Discussion

As in Experiment 1a, we compared the average magnitude of JOLs made to strongly related pairs (M = 75.80) versus unrelated

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

pairs ($M = 23.00$) via a planned-comparison t test, which revealed this difference to be significant, $t(29) = 12.65$, $p < .001$, $d = 3.08$. Thus, as in Experiment 1b, the relatedness between cue-target words seems to be the cue on which participants largely based their JOLs.

Figure 1 (B) presents cued-recall performance for Experiment 1b. A 2 (cue-target association: high vs. unrelated) × 2 (judgment condition: judgment vs. no-judgment) mixed-design ANOVA conducted to assess cued-recall performance revealed a significant main effect of cue-target association such that strongly related pairs were recalled at a higher rate than unrelated pairs ($M = .70$ vs. .19), $F(1, 58) = 692.88$, $p < .001$, $\eta_p^2 = .92$, but no significant main effect of judgment condition, $F(1, 58) = 2.56$, $p = .12$. Importantly, however, the significant interaction seen in Experiment 1a between cue-target association and judgment condition occurred again in Experiment 1b, $F(1, 58) = 7.17$, $p = .01$, $\eta_p^2 = .11$, with planned-comparisons confirming cued-recall to be greater for judged strongly related pairs than for nonjudged strongly related pairs, $t(58) = 2.58$, $p = .01$, $d = .69$, but not for judged versus nonjudged unrelated pairs ($p > .05$).

Thus, the pattern of results obtained in both Experiments 1a and 1b support our conjecture that by making the intrinsic cue of relatedness salient, participants would use such information as the basis for making their JOLs, resulting in the strengthening of that association when such information was easily discernable (i.e., for the strongly related pairs in both Experiments 1a and 1b), but to a lesser extent, if at all, when such information was not easy to discern (i.e., for the weakly related and unrelated pairs in Experiments 1a and 1b, respectively), leading, in turn, to enhanced later recall for judged versus nonjudged strongly related pairs, but not for judged versus nonjudged weakly related or unrelated pairs.

As previously discussed, the strengthening of cue-target relational information has also been implicated in producing the generation effect when such information must be used to complete the generation task. For example, in a study involving the learning of cue-target pairs for which some targets were presented intact (i.e., read items) and some had to be generated, de Winstanley et al. (1996) observed a generation advantage in participants' performance on a later cued-recall test when conditions of learning were manipulated during study such that participants were forced to rely on cue-target-relational processing in order to generate the targets. Thus, under conditions structured so that both the act of making JOLs and the act of generation would make use of cue-target relational information, we would expect the generation effect to be eliminated, or at least significantly reduced, if participants were required to make JOLs to read items. Experiment 2 tested this prediction.

Experiment 2

Method

Participants, materials, design, and procedure. Forty participants (17 women, 23 men; median age = 30.5 years, range = 19–62 years) were paid $1.00 for completing the experiment and were recruited online via Amazon's Mechanical Turk. Participants were fluent English speakers and lived in the United States.

Participants studied 50 strongly related cue-target word pairs (according to Nelson et al., 1998) for a later memory test. In the no-judgment condition, participants ($n = 20$) completed a standard generation experiment in which half of the cue-target pairs were simply read and studied in their entirety for 8 s each; whereas, for the other half of the pairs, the targets were generated. For each generated pair, participants were shown the cue-word and the consonants of the target-word (e.g., *ORCHID—FL_W_R*) and were instructed to type in the entire target-word (*FLOWER*) within 8 s. Participants were informed that the target-word was strongly related to the cue-word. Read and generated items were intermixed during study and were counterbalanced across participants. The procedure for the judgment condition was identical to the no-judgment condition except that participants ($n = 20$) were required to make a JOL for each to-be-read item by estimating the likelihood, on a 0–100% scale, of successfully recalling the pair on a later test, being prompted to do so half way through the exposure duration (i.e., after 4 s). No JOLs were made for the to-be-generated pairs. Following presentation of all pairs, participants in both conditions played Tetris for 3 min before taking a cued-recall test. During the cued-recall test, all 50 cue-words were presented individually for 5 s each in a randomized order, and participants were instructed to type in the target-word within that time.

Results and Discussion

Figure 2 presents cued-recall performance for Experiment 2. To assess whether making JOLs eliminates or reduces the generation effect, we conducted a 2 (item type: read vs. generate) × 2 (judgment condition: judgment vs. no-judgment) mixed-design ANOVA, which revealed a significant main effect of item type such that, overall, generated pairs were recalled at a higher rate than read pairs ($M = .78$ vs. .63), $F(1, 38) = 47.89$, $p < .001$, $\eta_p^2 = .56$, but no significant main effect of judgment condition was revealed, $F(1, 38) = 1.89$, $p = .18$. Critically, however, a significant interaction between item type and judgment condition was shown to be reliable, $F(1, 38) = 11.80$, $p = .001$, $\eta_p^2 = .24$. Consistent with our expectations, planned-comparisons confirmed cued-recall to be greater for judged read pairs than for nonjudged read pairs, $t(38) = 2.42$, $p = .02$, $d = .79$, but no recall difference across conditions for the generated pairs ($p > .05$). We note, however, that while making JOLs did significantly attenuate the generation effect, making such judgments did not completely eliminate it, as generated items in the judgment condition were recalled more often than read items that received JOLs, $t(19) = 2.55$, $p = .02$, $d = .46$.

Experiment 2 showed that the generation effect can be substantially reduced by having participants make JOLs for read items, suggesting that both the generation effect and the act of making JOLs in the current context enhanced memory, at least in large part, by strengthening cue-target relational information. That the read items in both the judgment and no-judgment conditions were related and that the attenuation of the generation effect resulted from judged read items being better remembered than nonjudged read items is consistent with the results of Experiment 1. Thus, the pattern of results in Experiment 2 accords with our current conjecture regarding the potential consequences of making JOLs.

General Discussion

The present research was undertaken to ascertain, first, whether simply the requirement to make JOLs to studied items would result in their enhanced recall on a later memory test compared to when

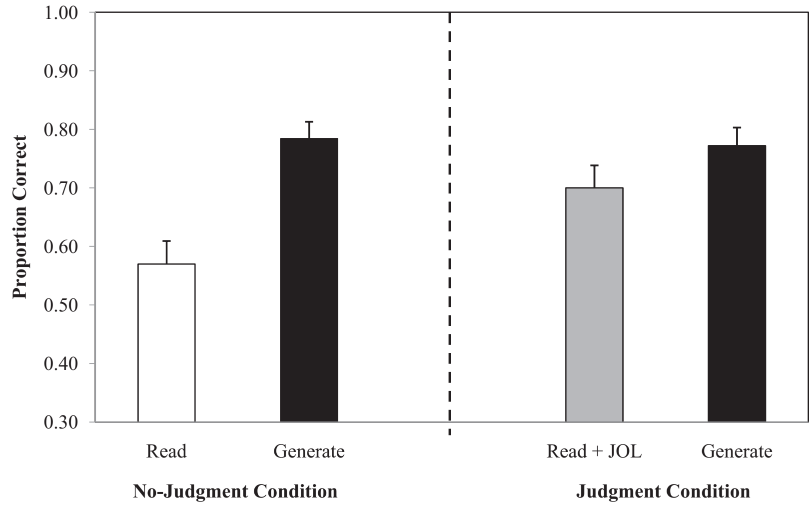

Figure 2. Mean cued-recall performance as a function of judgment condition and item type in Experiment 2.

the same items were studied for an equal duration but without any requirement to make JOLs. Second, if such a result emerged, could the observed enhancement be attributed to a strengthening of the cues or information used as the basis for making the JOLs? Based on the pattern of results obtained across all of the present experiments, the answer to both of these questions appears to be "yes."

That the strongly related pairs in Experiments 1a and 1b and the to-be-read items in Experiment 2 were better remembered when judged than when not judged implies that merely the act of making a JOL can modify the memorial representation of the judged item, which, as shown in these studies, can have performance-enhancing consequences. However, this performance-enhancing effect from making JOLs cannot be thought of as an inevitable outcome, given the interaction between degree of cue-target association and the requirement to judge observed in Experiments 1a and 1b. Thus, the pattern of results obtained across all of the present experiments is, we believe, consistent with our initial conjecture—which, as previously discussed, incorporates ideas from both Koriat's cue-utilization approach to JOLs (Koriat, 1997) and de Winstanley et al.'s (1996) transfer-appropriate multifactor account of generation effects—that the requirement to make a JOL strengthens the cues or information used as the basis for doing so, leading to better memory performance on the items so judged when a later criterion test is sensitive to such information.

In the present research, by making the intrinsic cue of relatedness highly salient, participants did indeed appear to use such information as a basis for making their JOLs, resulting in the strengthening of such information when it was easily discernable (i.e., for strongly related pairs), but to a lesser extent, if at all, when such information was not easily discernable (i.e., for weakly related and unrelated pairs), leading to the observed enhanced performance for judged strongly related pairs, but not judged weakly related pairs (Experiment 1a) or judged unrelated pairs (Experiment 1b), compared to their nonjudged counterparts. Replicating and extending the results of Experiments 1a and 1b, Experiment 2 demonstrated that making JOLs for related to-be-read items boosted their recall relative to nonjudged related to-be-read items, which resulted in a significant attenuation of the generation effect, suggesting that both the act of generation and the act of making JOLs strengthened cue-target relational information. That the generation effect was not completely eliminated in the current context suggests that generation may have strengthened such relational information even more so than making JOLs or, alternatively, that generation also strengthened item-specific information—in this case, for the target-word that was generated.

In addition to supporting our conjecture as to how making JOLs can improve later memory performance, the present results also clearly demonstrate that the making of JOLs cannot be viewed as an inconsequential act—that is, one that simply assesses an item's current state of learning without the potential of changing its state of learning. This potential effect of JOLs thus needs to be kept in mind both when designing and interpreting the results of studies employing the requirement to make JOLs and, perhaps, other metacognitive judgments as well (but see Dougherty et al., 2005, Experiment 2, for evidence that such memory-enhancing effects might be unique to JOLs). Memory researchers wishing to use such measures as one of their investigative tools may, for example, need to conduct their studies both with and without the collection of such judgments in order to obtain an undistorted assessment of how memory processes are affected by the other factors under investigation. Furthermore, given the present evidence that the effect of making JOLs can depend upon the nature of the items being judged and the cues or type of information they present for making such judgments, the inclusion of a judgment-free condition would seem necessary not only for accurate interpretation of memory results but also for the accurate interpretation of the observed correlation between JOLs and future memory performance.

Concluding Comments

In summary, the present results have a somewhat disturbing implication of which investigators need to be mindful when using and interpreting JOLs: Namely, truly accurate JOLs—in the sense of judgments that reflect one's current state of knowledge or degree of learning some information—may not be obtainable because the mere act of making the judgment is likely to affect that

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

which it is intended to assess and with consequences for later performance.

On a somewhat brighter side, however, the present findings would seem to have a potentially positive implication for education. When preparing for class examinations, students want to discriminate between what they already know well versus what they know less well in order to focus their study activities efficiently. Although only a small percentage of students report using self-administered tests to make such assessments (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007), most students probably make something akin to JOLs for the various topics or facts to be covered on the exams. The present results suggest that just the simple act of making such assessments may have some power to make the assessed material more memorable and thus, under the right conditions, increase students' later exam performance.

References

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81,* 126–131. http://dx.doi.org/10.1037/h0027455

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. http://dx.doi.org/10.1037/0096-3445.127.1.55

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35,* 201–210. http://dx.doi.org/10.3758/BF03193441

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII—cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64,* 417–444. http://dx.doi.org/10.1146/annurev-psych-113011-143823

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33,* 497–505. http://dx.doi.org/10.1080/14640748108400805

de Winstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory, 4,* 31–48. http://dx.doi.org/10.1080/741940667

Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory & Cognition, 33,* 1096–1115. http://dx.doi.org/10.3758/BF03193216

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565. http://dx.doi.org/10.1006/jmla.1994.1026

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19,* 126–134. http://dx.doi.org/10.3758/s13423-011-0181-y

Kelemen, W. L., & Weaver, C. A., III. (1997). Enhanced metamemory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1394–1409. http://dx.doi.org/10.1037/0278-7393.23.6.1394

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

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14,* 219–224. http://dx.doi.org/10.3758/BF03194055

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 609–622. http://dx.doi.org/10.1037/0278-7393.32.3.609

Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766. http://dx.doi.org/10.1037/0278-7393.10.4.756

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44,* 1–23. http://dx.doi.org/10.3758/s13428-011-0124-6

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.* Retrieved from http://w3.usf.edu/FreeAssociation/

Nelson, T. O., Dunlosky, J., Graf, E. A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207–213. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00502.x

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686. http://dx.doi.org/10.1037/0278-7393.14.4.676

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. http://dx.doi.org/10.1037/0278-7393.4.6.592

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73,* 99–115. http://dx.doi.org/10.1016/j.jml.2014.03.003

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1236–1242. http://dx.doi.org/10.1037/a0023548

Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (in press). Metamemory and education. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory*. New York: Oxford University Press.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 204–221. http://dx.doi.org/10.1037/0278-7393.26.1.204

Spellman, B. A., & Bjork, R. A. (1992). Technical commentary: When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3,* 315–316. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00680.x

Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *The Quarterly Journal of Experimental Psychology, 65,* 1376–1396. http://dx.doi.org/10.1080/17470218.2012.656665

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15,* 41–44. http://dx.doi.org/10.3758/BF03329756

Received March 4, 2012
Revision received September 15, 2014
Accepted October 2, 2014 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.