Conditions of highly specific learning through cued recall

Steven C. Pan[1,2] · Jarrett Lovelett[2] · Derek Stoeckenius[2] · Timothy C. Rickard[2]

Abstract

After studying a stimulus (e.g., a word triplet such as *gift*, *rose*, *wine*), taking a cued recall test on that stimulus (e.g., *gift*, *rose*, *?*) improves later recall of the retrieved term (e.g., *wine*) relative to a restudy control. That testing effect, however, is specific to the tested term: later recall of a previously untested term from the same stimulus (e.g., *gift* or *rose*) is not enhanced. In the present research, two possibilities for that highly specific learning were investigated: (a) learning through cued recall is always highly specific to the tested term, or, alternatively, (b) learning specificity is unique to the case of retrieving a term from an episodic memory of a study event. We addressed those possibilities by using the pretesting paradigm, in which there is no study opportunity prior to cued recall testing, and hence retrieval occurs through semantic memory. The results of two experiments supported the latter hypothesis. Thus, it is not the recall attempt per se that produces highly specific learning, but rather the attempt to recall the response by accessing an episodic memory of a particular study event. Theoretical and practical implications for learning through cued recall are discussed.

Keywords Cued recall · Transfer · Testing effect · Pretesting effect · Retrieval practice · Feedback

The testing effect, or test-enhanced learning, is the finding that retrieval practice is typically a more potent learning method than is an equivalent amount of time spent restudying target materials or no reexposure to those materials (Roediger & Karpicke, 2006). For the case of cued recall compared with restudying, the testing-effect paradigm (henceforth, *testing paradigm*; see Fig. 1) commonly involves three phases: (a) a study phase, wherein all items (e.g., word triplets or biology facts) are studied; (b) a training phase, wherein half of the items are restudied and the other half are presented for a cued recall test (commonly followed by correct answer feedback); and (c) a final test, wherein all items are tested again through cued recall. This paradigm thus compares the effects of cued recall tests versus restudying on memory. The testing effect for cued recall is robust; in their theoretical review, Rickard and Pan (2017) found that 96% of 114 experiments exhibited test-enhanced learning on the final test.

However, for cases in which two or more cues are presented on the training test and a target word or term has to be recalled (as on a short answer or fill-in-the-blank test), the memory enhancement that occurs through cued recall does not exhibit transfer to a new cue–response arrangement of the same item. For example, consider an experiment wherein subjects first study a set of word triplets (e.g., *gift*, *rose*, *wine*). The training phase involves recall of a single word from each triplet when given the other two words as cues (e.g., recalling *wine* when given *gift*, *rose*, *?*, as cues). As shown in Fig. 2a, that training yields improved final test performance (relative to a restudy control condition) for questions involving identical cues and responses (e.g., when *gift*, *rose*, *?* is again presented on the final test, with the correct response again being *wine*; henceforth, the tested-same condition), but no transfer of that effect (i.e., no improved recall relative to the restudy control) for the case of rearranged cues and response (e.g., presentation of *gift*, *?*, *wine* on the final test, with the correct response being *rose*; henceforth, the tested-different condition). That specificity of learning has been demonstrated for word triplets (Pan, Wong, Potter, Mejia, & Rickard, 2016), multiterm history and biology facts (Pan, Gopal, & Rickard, 2015; as shown in Fig. 2b), term-definition facts (Pan & Rickard, 2017), and biology concepts (Hinze & Wiley, 2011; Pan, Hutter, D'Andrea, Unwalla, & Rickard, 2018).

✉ Steven C. Pan
  stevencpan@psych.ucla.edu

1 Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA

2 Department of Psychology, University of California San Diego, San Diego, CA, USA
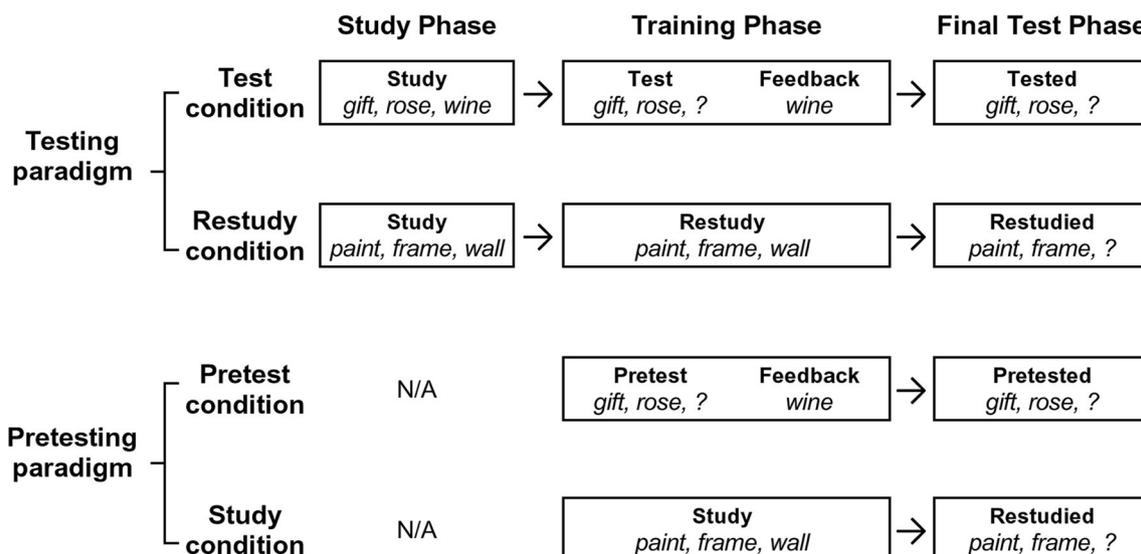
	Study Phase		Training Phase		Final Test Phase
Testing paradigm — **Test condition**	**Study** *gift, rose, wine*	→	**Test** *gift, rose, ?* / **Feedback** *wine*	→	**Tested** *gift, rose, ?*
Restudy condition	**Study** *paint, frame, wall*	→	**Restudy** *paint, frame, wall*	→	**Restudied** *paint, frame, ?*
Pretesting paradigm — **Pretest condition**	N/A		**Pretest** *gift, rose, ?* / **Feedback** *wine*	→	**Pretested** *gift, rose, ?*
Study condition	N/A		**Study** *paint, frame, wall*	→	**Restudied** *paint, frame, ?*

Fig. 1 Testing and pretesting paradigms. In the testing paradigm (top), all items are first studied (study phase); then, half of the items are studied again and the other half presented for a cued recall test followed by correct answer feedback (training phase); and finally, all items are tested (final test phase). The pretesting paradigm (bottom) is identical, except that there is no study phase. Consequently, accuracy on the pretest is approximately zero, and learning of the correct answer in the pretested condition occurs entirely through feedback

Based on that and related evidence, Rickard and Pan (2019) concluded that retrieval of a target through a set of two or more cues in the testing paradigm yields a new episodic memory of the testing event (which we refer to as test memory) that can subsequently be accessed only on highly similar test trials. Specifically, their data suggest that test memory in that case is functionally an inclusive-OR gate, such that the memory can be accessed on a later final test only if the originally presented cues, *or* a proper subset of them, are represented and the required response is the same. Introduction of the hypothesized inclusive-OR gate into Rickard and Pan's (2017) dual-memory model of the testing effect (wherein study and test events are assumed to create separate episodic memories) allowed that model to capture not only the general learning specificity effect but also the finding over multiple experiments of equivalent performance in the tested-different and restudy conditions (e.g., the results shown in Fig. 2).

Moreover, that account assumes that an inclusive-OR test memory forms on both correct and incorrect training test trials. Rickard and Pan (2019) found support for that assumption by showing that the predicted performance equivalence between the tested-different and restudy conditions holds across 17 experiments in the literature wherein proportion correct on the training phase test ranged from 0.27 to 0.93. Hence, the inclusive-OR account, and the corresponding performance equivalence in the tested-different and restudy conditions, appears to hold for both correct and incorrect training test trials in the testing paradigm.

Despite that empirical progress, neither the theoretical basis of that learning specificity effect nor its generality beyond the testing paradigm are fully understood. Toward that aim, we consider here two straightforward hypotheses. First, it may be that high specificity of learning occurs under all recall circumstances wherein retrieval is attempted from two or more cues, constituting a new and general principle of learning through cued recall. We are aware of no prior work in the literature that addresses that possibility. Alternatively, it may be that the learning specificity is unique to the study–test sequence in the testing paradigm (and analogous ecological contexts), wherein the training phase test prompts recall of an episodic memory of a prior study event. It may not be observed, for example, for the case of cued recall from semantic memory, as opposed to recall from episodic study memory. We defer further theoretical treatment of those possibilities to the General Discussion, and focus first on empirical tests of them.

In the work reported here, we tested those two competing hypotheses using the pretesting paradigm (Hays, Kornell, & Bjork, 2013; Huelser & Metcalfe, 2012; Kornell, Hays, & Bjork, 2009; Potts & Shanks, 2014; Richland, Kornell, & Kao, 2009). In that paradigm there is no initial study phase, but there are training and final test phases, just as in the testing paradigm (see Fig. 1 for a comparison of the two paradigms). On the training phase test (the pretest), subjects have to guess the correct response (e.g., supply a missing term), followed by correct answer feedback. Prior work using latent semantic analysis indicates that retrieval on the pretest occurs through semantic memory (Huelser & Metcalfe, 2012; see also Grimaldi & Karpicke, 2012). By design in that paradigm, proportion correct on the pretest is virtually zero (i.e., retrieved responses, though often bearing some semantic relationship
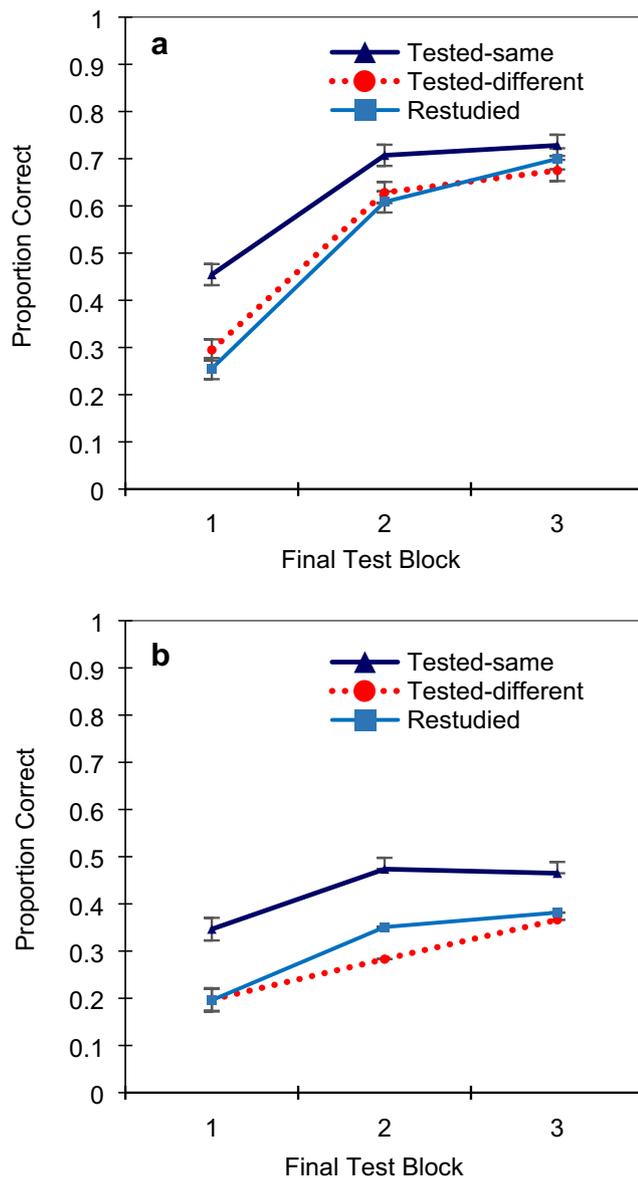
Fig. 2 Specificity of learning following cued recall testing as evident on a delayed final test. Evidence from word triplets (**a**) and multiterm facts (**b**), reproduced from Pan, Wong, Potter, Mejia, and Rickard (2016, Experiment 1) and Pan, Gopal, and Rickard (2015, Experiment 1) data. *Tested-same* refers to the case of identical correct answer responses on initial and final tests, *tested-different* refers to the case of different responses on initial and final tests, and *restudied* refers to items that were restudied during training

In both experiments in this study, two of the three words of a triplet were presented as cues on pretest trials. On the final test, two words were again presented on each trial, with a third word to be retrieved. Hence the cue configurations of the training and final test events were identical to those of the aforementioned testing paradigm for triplets (see Fig. 1). There were three conditions on the final test, all matching those of the testing paradigm: pretested-same, pretested-different, and studied (see Fig. 3). That design is thus well suited for testing the two hypotheses outlined above. If, as a general principle, a cued recall attempt with feedback always yields high learning specificity when two or more cues are presented, then the same learning specificity that has been observed in the testing paradigm should be observed in the current pretesting experiments (i.e., that pattern might hold regardless whether retrieval occurs through episodic or semantic memory). Alternatively, if specificity of learning through cued recall is unique to the study–test sequence of the testing paradigm, then that specificity should not be observed in the current pretesting experiments, in which there is no study phase.

Experiment 1

The first experiment investigated pretesting and learning specificity using a 48-hr delayed final test, a retention interval that is consistent with the aforementioned testing paradigm studies.

Method

Subjects Fifty-two undergraduate students participated for course credit. Four subjects were excluded because they did not return for the second session, and one was excluded due to experimenter error. The target sample size in all experiments was 43, which yields statistical power of .85 to detect a ≥.05 proportion correct difference between tested-different and nontested conditions (based on a one-tailed *t* test of triplet data from Pan et al., 2016, Experiment 1).

Materials The stimuli were 36 three-word triplets (Pan et al., 2016). Word frequency and forward and backward associative strength, where data were available, averaged 82 and 0.08, respectively (Nelson, McEvoy, & Schreiber, 2004). A master list was used to create six training lists wherein assignment of triplets to pretesting or study, and the choice of to-be-retrieved word during pretesting (i.e., one of three words per triplet), was counterbalanced over subjects.

For each training list, a corresponding final test list was created in which each word per triplet was assessed once in each of three 36-trial blocks. Each block had six pretested-same, 12 pretested-different, and 18 studied trials. Pretested-same trials assessed a pretested triplet for the same word as
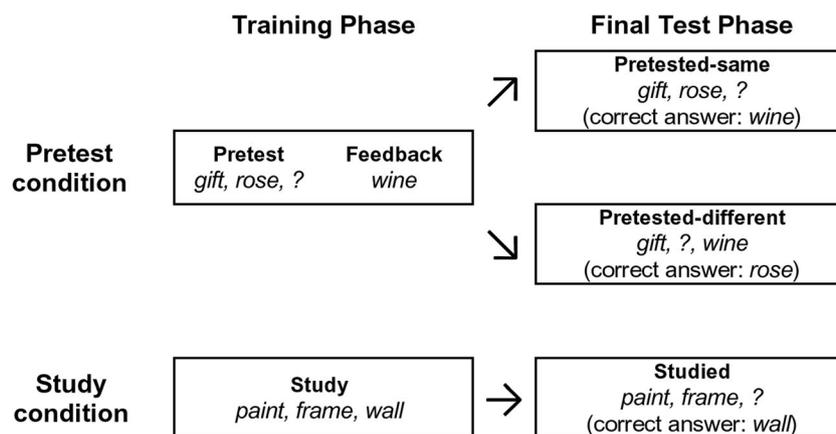
with the cue, are not the correct answers). Hence, the structure of the pretesting paradigm is most analogous to incorrect test trials in the testing paradigm. Final test performance (i.e., accurate recall of the correct responses as indicated during feedback on pretest trials) has been shown to be better in the pretested condition than in a study control condition (the pretesting effect). However, the question of whether pretesting yields learning that is transferrable (i.e., not highly specific), such as to the case of cue–response rearranged items, has yet to be addressed.

Fig. 3 Example training and final test trials used in Experiments 1 and 2. During training, subjects are asked to type a response on each pretest trial, followed by correct answer feedback, and to study intact triplets on each study trial. On the final test, pretested-same trials involve recalling previously pretested responses, pretested-different trials involve recalling cues from pretested triplets, and studied trials involved previously studied triplets

was missing during training, pretested-different trials assessed a pretested triplet for a different word as was pretested during training, and studied trials assessed previously studied triplets. Each subject trained using one training list and was assessed using a corresponding final test list.

Design and procedure There were two experimental phases, training and final test. Training method (pretesting vs. study) was manipulated within subjects. During the training phase, half of the triplets were presented for pretesting and half were presented for study, in random order (for a total of 36 training trials, one per triplet). On pretest trials, two words per triplet were presented, and the third was replaced by "*?*" (e.g., *gift, rose, ?*). Subjects were instructed to guess the missing word, type it into a textbox, and to not reuse guesses across triplets. After 5 s, the "*?*" was replaced by the correct answer (e.g., *gift, rose, wine*), and the complete triplet was presented for another 5 s as feedback. On study trials, a complete triplet (e.g., *paint, frame, wall*) was presented for 5 s (hence, exposure time per complete triplet was equated, e.g., Huelser & Metcalfe, 2012). The spatial order of each word (and any "*?*") was randomized to top, middle, or bottom positions (columnar format) on each trial. Final test trials were identical, except for being self-paced with no feedback. The two phases were separated by a 48-hr delay.

Results and discussion

Training As expected, given that there was no prior study, and in line with prior results for this paradigm, pretest accuracy was low (3.1%). Following the pretesting literature, items corresponding to those correct trials were excluded from the final test analysis.

Final test The expected pretesting effect was observed (see Fig. 4a). Of most interest, there was substantial positive transfer to the pretested-different condition. As detailed in Table 1, a factorial analysis of variance (ANOVA) with factors of block (1–3) and condition (pretested-different vs. studied; the tested-same condition was excluded to yield the critical single degree of freedom test on the condition factor) confirmed that transfer. There was also a main effect of block (reflecting the fact that answers on Blocks 2 and 3 were presented as cues during preceding blocks; see Pan et al., 2016) and no apparent Block × Condition interaction. Thus, in diametric contrast to results from the testing paradigm, pretesting yields highly transferrable learning for triplets. This constitutes clear evidence that cued recall does not always yield specific learning.

Experiment 2

Experiment 2 exactly replicated Experiment 1, but with a 5-min. delay, which is consistent with the delay used in most pretesting studies.

Method

Subjects Forty-seven undergraduate students participated for course credit. Data from one subject was excluded due to a computer error.

Materials, design, and procedure All aspects of this experiment were identical to its predecessor, except for the 5-min. delay, during which subjects completed an unrelated visual distractor task (adapted from Kornell & Son, 2009).
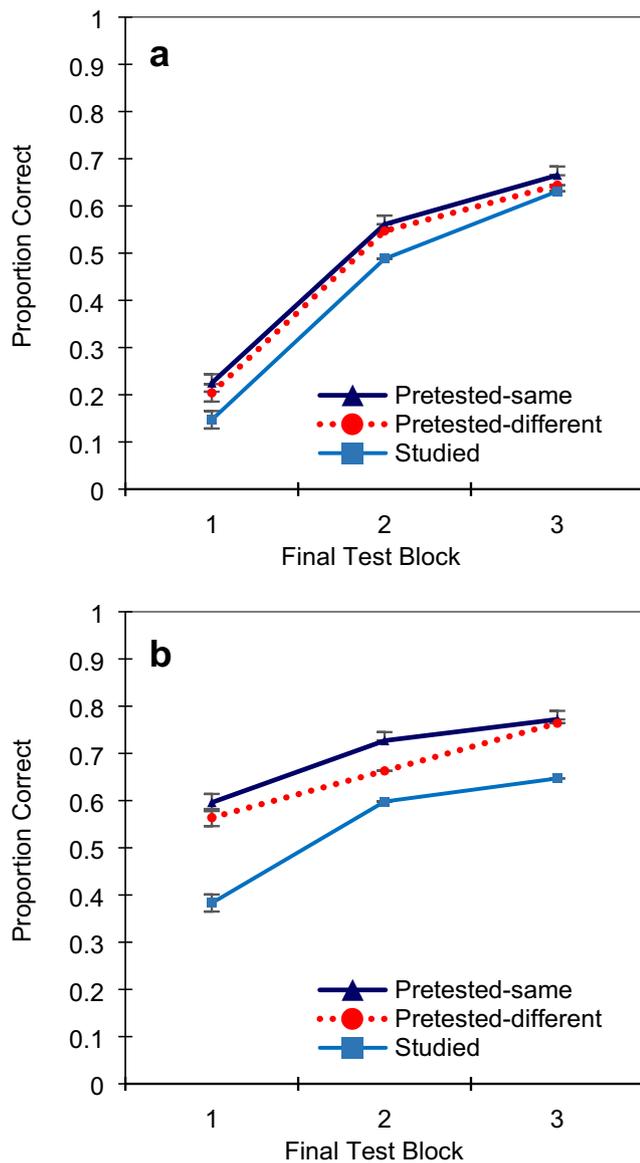
Fig. 4 Final test results of Experiments 1 (**a**) and 2 (**b**). Error bars are standard errors based on the interaction error term of a within-subjects analysis of variance on subject-level mean accuracy scores (based on Loftus & Masson, 1994)

Results and discussion

Training Mirroring the pattern observed in the prior experiment, only 3.6% of missing words were guessed correctly.

Final test The results (see Fig. 4b) converge with those of Experiment 1. There was again robust evidence of a pretesting effect and nearly complete transfer of that effect to the pretested-different condition (see Table 1). There was also a main effect of block, just as in the preceding experiment, but uniquely in this experiment there was a significant Block × Condition interaction in Experiment 2. Inspection of Fig. 4 shows that the interaction was due to a reduced transfer effect in Block 2 versus Blocks 1 and 3. In our experience with multiple experiments in the testing paradigm, such block-wise fluctuations about the general pattern are unlikely to replicate.

General discussion

In two experiments, we demonstrated that the specificity of learning that has been observed over multiple experiments in the cued recall testing paradigm does not manifest in the pretesting paradigm. Indeed, in diametric opposition to the testing paradigm, there was nearly identical performance in the pretested-same and pretested-different conditions (compare Figs. 2 and 4). That finding allows us to reject the possibility that learning specificity for cued recall occurs globally and is an intrinsic consequence of the cued recall attempt. Instead, the results support the alternative hypothesis that learning specificity is unique to the study–test sequence of the testing paradigm and analogous contexts.

The pretesting paradigm provides a close analog to the case of incorrect trials with feedback in the testing paradigm, with the only difference being the initial study phase, which is unique to the testing paradigm. The case of correct retrieval in the testing paradigm is somewhat less analogous to the pretesting paradigm. However, as described earlier, specificity of learning, when predicted by the inclusive-OR account, has occurred across 17 experiments having a wide range of

Table 1 Experiments 1–2 final test analysis of variance (ANOVA) results

Exp.	Contrast type	Factor	df	F	MSE	p	η_p^2
1	Pretested-different vs. studied	Main effect of block	2, 92	216.00	5.41	<.0001	0.82
		Main effect of condition: Pretested-different vs. studied	1, 46	6.65	0.13	.013	0.13
		Interaction	2, 92	1.38	0.016	.26	0.029
2	Pretested-different vs. studied	Main effect of block	2, 90	108.50	1.29	<.0001	0.71
		Main effect of condition: Pretested-different vs. studied	1, 45	33.28	1.01	<.0001	0.43
		Interaction	2, 90	5.64	0.078	.0049	0.11

Note. Exp. = Experiment

training test proportion correct. Thus, our main conclusion—that the specificity of learning occurs not globally, but rather as a consequence of retrieval from episodic memory under conditions of two or more cues—appears to hold for both incorrect and correct retrieval.

A reasonable theoretical conclusion is that learning specificity occurs when there is attempted recall from a specific episodic memory (as on the training test in the testing paradigm), but does not occur when there is attempted recall from semantic memory (as on the training test in the pretesting paradigm). Why would specificity of learning be unique to retrieval from episodic memory? One possibility that is consistent with recent neuroimaging results is that learning through cued recall from an episodic memory invokes a pattern separation process in the posterior hippocampus and anterior medial prefrontal cortex. As concluded by LaRocque et al. (2013) and Schlichting, Mumford, and Preston (2015), that process plays a special role in maintaining separate memories when those memories have the potential to mutually interfere. Consider the studied triplet *gift, wine, rose*. If on a training phase test, two of those elements, say *gift* and *wine*, are presented as cues for the third, then the similarity between the study memory (*gift, rose, wine*) and the recall cues (*gift, ?, wine*) is relatively high. Under those conditions, a separate test memory may be encoded during the initial cued recall test. Because semantic retrieval is not believed to involve the hippocampal system, the pattern separation process would not be expected to operate in the pretesting paradigm.

It is also plausible that a separate test memory for recall from two or more cues, being a record of sequential cued recall events (i.e., cue presentation, followed by answer production and then feedback), has the inclusive-OR access property that we have inferred from behavioral data (Rickard & Pan, 2019). This proposal of separate study and test memories for the same item may be at odds with intuition. Instead, one might expect that those two memories would combine and reinforce even in the tested-different condition. However, a pattern separation process, as a general purpose mechanism, may separate memories with overlapping elements under all circumstances, irrespective of what psychologists might expect. Further, pattern separation may be adaptive under the current circumstances, as it may achieve two goals: (a) protection of the originally encoded study memory, such that it can serve as the basis for retrieval for previously unretrieved responses, and (b) facilitation of future retrieval of the previously tested response.

The pattern separation account might also explain our prior results showing that transfer to the tested-different condition does occur in the testing paradigm when only one cue is presented on the training test. That result has been obtained for both paired associates and triplets. In Experiment 3 of Rickard and Pan (2019), for example, a single word from each tested triplet was presented on the training test and the remaining two

words were to be recalled (e.g., *gift, ?, ?* was presented). On the final test, two words from each triplet were presented as cues and the third word was to be recalled. In addition to the restudy condition, there were two tested-different conditions on the final test: (a) the two prior responses (i.e., the responses during the training phase) cue was presented and the prior cue was to be recalled (e.g., *?, rose, wine*); and (b) a prior response and the prior cue was presented, with the prior response to be recalled (e.g., *gift, rose, ?*). Final test performance in both of those tested-different conditions was substantially and significantly better than in the restudy control condition.

For both pairs and triplets in those experiments, there is objectively less similarity between the presented cues and the elements of study memory than in the case wherein two cues for a triplet are presented on the training test (i.e., in terms of proportion of cue overlap). It may thus be that the pattern separation mechanism does not operate in those cases. Alternatively, the pattern separation process may always operate following cued recall from episodic memory, but the test memory may only have learning specificity when two or more cues are presented. A single familiar word cue, having well-learned lexical and semantic representations, may, for reasons not yet identified, yield a separate test memory that is flexibly accessible, just as study memory is, in turn supporting transfer. Resolution of that issue remains for future work.

From the applied perspective, our results both generalize the pretesting effect to new materials and, more critically, demonstrate that pretesting can yield more flexible, or transferrable, learning than does a study–test sequence. Thus, evidence is accruing that pretesting can be used not only as an assessment of student knowledge prior to instruction but also as a potent learning method. That possibility invites further research using educational materials.

References

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory & Cognition, 40*(4), 505–513.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290.

Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory, 19*(3), 290–304.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40*(4), 514–527.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*(5), 493–501.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989.

LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *Journal of Neuroscience*, *33*(13), 5466–5474.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476–490.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Pan, S. C., & Rickard, T. C. (2017). Does retrieval practice enhance learning and transfer relative to restudy for term-definition facts? *Journal of Experimental Psychology: Applied*, *23*(3), 278.

Pan, S. C., Gopal, A., & Rickard, T. C. (2015). Testing with feedback yields potent, but piecewise, learning of history and biology facts. *Journal of Educational Psychology 107*(4), 563–575. doi:https://doi.org/10.1037/edu0000074

Pan, S. C., Wong, C. M., Potter, Z. E., Mejia, J., & Rickard, T. C. (2016). Does test-enhanced learning transfer for triple associates? *Memory & Cognition 44*(1), 24–36.

Pan, S. C., Hutter, S., D'Andrea, D., Unwalla, D., & Rickard, T. C. (2018). In search of transfer following cued recall practice: The case of biology concepts. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3506

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243.

Rickard, T. C., & Pan, S. C. (2017). A dual memory theory of the retrieval practice effect. *Psychonomic Bulletin & Review, 25*(3), 847–869. doi:https://doi.org/10.3758/s13423-017-1298-4

Rickard, T. C., & Pan, S. C. (2019). *Test-enhanced learning of pairs and triplets: When and why does transfer occur?* Manuscript under review available at Open Science Framework via: https://bit.ly/2FbFM2k

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151.