Adaptive and Qualitative Changes in Encoding Strategy With Experience: Evidence From the Test-Expectancy Paradigm

Jason R. Finley and Aaron S. Benjamin
University of Illinois at Urbana-Champaign

Three experiments demonstrated learners' abilities to adaptively and qualitatively accommodate their encoding strategies to the demands of an upcoming test. Stimuli were word pairs. In Experiment 1, test expectancy was induced for either cued recall (of targets given cues) or free recall (of targets only) across 4 study–test cycles of the same test format, followed by a final critical cycle featuring either the expected or the unexpected test format. For final tests of both cued and free recall, participants who had expected that test format outperformed those who had not. This disordinal interaction, supported by recognition and self-report data, demonstrated not mere differences in effort based on anticipated test difficulty, but rather qualitative and appropriate differences in encoding strategies based on expected task demands. Participants also came to appropriately modulate metacognitive monitoring (Experiment 2) and study-time allocation (Experiment 3) across study–test cycles. Item and associative recognition performance, as well as self-report data, revealed shifts in encoding strategies across trials; these results were used to characterize and evaluate the different strategies that participants employed for cued versus free recall and to assess the optimality of participants' metacognitive control of encoding strategies. Taken together, these data illustrate a sophisticated form of metacognitive control, in which learners qualitatively shift encoding strategies to match the demands of anticipated tests.

Keywords: encoding strategy, study-time allocation, metacognition, self-regulated learning, test expectancy

Students facing an upcoming test need to make a host of decisions about how to prepare for that test. In order to do so, they should know what material will be covered, when the test will occur, how well they currently know the material, and what steps they need to take to achieve the desired level of mastery. Yet one of the most common questions instructors hear concerns the format of the test: students want to know if the test will be multiple choice, fill-in-the-blank, essay, or some other format. Does this knowledge help students prepare more effectively? Or do they simply use this information to set a criterion for mastery on the basis of the anticipated difficulty of the test format?

There is surprisingly little extant evidence in support of the claim that learners actively change the way in which they learn material as a direct result of test knowledge (Lundeberg & Fox, 1991). Yet it would seem that effective studying requires the ability to tailor one's study behaviors to the foreseeable requirements of the test. If learners do not adjust to the demands of a test,

their studying may be inefficient and ineffective, no matter how hard they work. The current study examined the extent to which learners are able to make adaptive and qualitative changes in the way they learn material after experiencing the demands of an upcoming test format. Such *learning to learn* (Postman, 1964, 1969) requires strategic exercise of metacognitive control over one's memory processes.

Learners can regulate their study experience to enhance learning in a variety of ways. Metamemory research (i.e., research on the metacognition of memory) has focused on control processes such as item selection, study-time allocation, scheduling, and encoding strategy (e.g., Benjamin, 2007; Dunlosky, Serra, & Baker, 2007; Finley, Tullis, & Benjamin, 2010; Serra & Metcalfe, 2009). In the current study, we focused specifically on how learners change their encoding strategies for learning words on the basis of how they expect their memory for those words to be queried. To do so, we employed the test-expectancy paradigm, which compares performance on a particular test format by participants led to expect that format versus the performance of participants led to expect a different format. The findings from prior test-expectancy research (reviewed later) are inconsistent and inconclusive; in a meta-analysis, Lundeberg and Fox (1991) remarked that "we have little clear information on just exactly what students facing a certain kind of test do (that they would not do) if facing another kind of test" (p. 102). The current study—using cued and free recall—provides a means of reconciling and clarifying these prior findings. Furthermore, the study elucidates the characteristics of encoding strategies used by learners, the efficacy of such strategies, and the effectiveness of learners' strategic use of them. First, we will

review the topic of encoding strategies; then we will review prior test-expectancy research.

Encoding Strategy

The way in which learners encode information is critical to how that information is stored in memory (Craik & Lockhart, 1972; Fisher & Craik, 1977). This idea can be traced back to at least the era of verbal learning research; Eagle and Leiter (1964) noted that "the amount and kind of learning that takes place will depend, in large part, upon the kind of learning operations that are carried out upon the stimulus material" (p. 63). Though there is now a wealth of research on the effect of imposing various encoding strategies on learning, only recently have researchers begun to investigate how individual differences in the application of those strategies are informed by experience (cf. Hertzog, Price, & Dunlosky, 2008).

Normative Efficacy of Encoding Strategies

In many studies, researchers have investigated the normative efficacy of various encoding strategies by attempting to control learners' strategies via direct instructions or orienting tasks. The simplest strategy, often used as a baseline for comparison, is rote rehearsal (i.e., overtly or covertly repeating information to one-self). More elaborative strategies that have been shown to be efficacious in certain circumstances include semantic ("deep") encoding of words (Craik & Lockhart, 1972; Craik & Tulving, 1975), organizing words into subjectively meaningful groups (Tulving, 1966), visual imagery (cf. Hertzog & Dunlosky, 2006), and various mnemonics such as the method of loci and the peg word method (cf. Roediger, 1980). Many results can be explained by the framework of transfer-appropriate processing (Morris, Bransford, & Franks, 1977), which holds that encoding strategies are efficacious to the extent that they employ cognitive processes during acquisition that are similar to the processes that will be used at retrieval (cf. Blaxton, 1989; Jacoby, 1983; Roediger, Weldon, & Challis, 1989).

Control of Encoding Strategy

Much less is known about how learners employ encoding strategies when left to their "druthers" and whether they can adaptively adjust their strategies to meet the demands of a future task. That is, little is known about learners' metacognitive control of encoding strategies.

There are two basic types of adjustments that learners can make to their encoding strategies: quantitative changes and qualitative changes.[1] A learner may apply the same encoding strategy to varying degrees based on the anticipated difficulty of an upcoming test—a quantitative change, which could be due to purely motivational factors. Or a learner may apply different encoding strategies based on the anticipated format of an upcoming test—a qualitative change, which cannot be due to merely trying harder. As we will review, there has been ample evidence of the former, but surprisingly little evidence of the latter.

Test Expectancy

The encoding strategies used by learners are difficult to experimentally investigate because unlike item selection, study-time allocation, and scheduling, such processes are not directly observable. The test-expectancy paradigm provides one way to study whether and how effectively learners use different encoding strategies for different tasks. In this paradigm, participants are first led to expect a particular test format, either via instructions or via experience with a series of tests of the same format. They are then given a final test that consists of either the expected format or the alternative format. Final test performance is compared—separately for each final test format—for participants who had expected that format versus participants who had expected the alternate format. If all other forms of metacognitive control (e.g., study-time allocation) are held constant, then performance differences due to the expectancy (also called *mental set*) manipulation reflect differences in the encoding strategies employed by participants during study. Thus, such data allow inferences to be drawn about whether participants tailor their encoding strategies to the demands of a specific expected test format.

The most prominent finding from studies using this paradigm is that expectation of free recall appears to facilitate performance for both free recall and item recognition tests. More specifically, several studies have shown that participants anticipating a free recall test achieve higher performance on tests of both free recall and item recognition than do participants anticipating an item recognition test (Balota & Neely, 1980; Connor, 1977; d'Ydewalle, Swerts, & de Corte, 1983; Foos & Clark, 1983; Hall, Grossman, & Elwood, 1976; Leonard & Whitten, 1983; Maisto, DeWaard, & Miller, 1977; Meyer, 1934, 1936; Neely & Balota, 1981; Schmidt, 1988; Thiede, 1996). It is also clear from studies of intentional versus incidental learning that any knowledge at all of an upcoming test can generally enhance performance (cf. McDaniel, Blischak, & Challis, 1994; Szpunar, McDermott, & Roediger, 2007).

Although it is possible that the above-described pattern of results is due to qualitative differences in encoding strategies for different anticipated test formats, it is also possible that the pattern is due simply to judicious *quantitative* adjustments to encoding strategies based on anticipated test format. For example, learners may anticipate that a recall test will pose greater difficulty than a recognition test, and consequently they may study more (i.e., they may use the same encoding strategies, just to a greater extent). Thus, none of these findings can be safely concluded to reflect *qualitative* changes in encoding strategy as a function of test expectancy. The pattern of data required for such a conclusion is a disordinal (i.e., *cross-over*) interaction, such that, for both final test formats, learners who expected a particular format outperform those who expected the different format. Blaxton (1989, p. 657) argued that such a dissociation in performance on different memory tasks is well explained by "the degree of overlap between mental operations at study and test" (i.e., transfer-appropriate processing). Some studies have explicitly sought to detect such an interaction and have failed to find it (e.g., Hall et al., 1976; Jacoby, 1973; Lewis & Wilding, 1981; Lovelace, 1973; Neely & Balota,

---

[1] Tversky (1973) proposed that encoding strategies may differ in three ways: encoding of more information (quantitative), encoding of different kinds of information (qualitative), and encoding of information organized in a different manner (qualitative). We focus here on the broader distinction between quantitative versus qualitative differences.

1981; Schmidt, 1988; Tversky, 1973, Experiment 1[2]). For example, Neely and Balota (1981) predicted that recall expectancy should yield encoding strategies that were more relational and less item specific than those yielded by recognition expectancy, but they found no evidence of this. These data are curiously inconsistent with students' self-reports that they consider different study methods as best suited for different test formats, such as focusing on details and underlining key terms when preparing for a fill-in-the-blank or true–false test and organizing main points when preparing for an essay test (Terry, 1933, 1934).

Some researchers (e.g., Kulhavy, Dyer, & Silver, 1975; Oakhill & Davies, 1991; von Wright & Meretoja, 1975) have suggested that differences in encoding strategy may not necessarily be reflected in overall *levels* of performance but may appear as different *patterns* of performance. Such differences have been found in intracategory serial position functions (Carey & Lockhart, 1973; but cf. Hall et al., 1976, for a failure to replicate), the shape of serial position functions (d'Ydewalle, 1981; May & Sande, 1982), source memory (Watanabe, 2003), and semantic organization of output in free recall (d'Ydewalle, 1982; Jacoby, 1973). Other suggestive results include those of Leonard and Whitten (1983), who found that learners expecting free recall outperformed learners expecting item recognition on a task in which they sorted items into their original presentation order (Experiment 1) and that related distractors impaired item recognition for learners expecting item recognition but not for learners expecting free recall (Experiment 5; cf. Whitten, 2011). There is even some tentative evidence of different encoding strategies for recognition versus recall from functional neuroimaging (Staresina & Davachi, 2006). While such results may hint at qualitative differences in encoding strategy, it is still difficult to rule out the possibility that learners expecting one test format are simply trying harder than those expecting another. Only the aforementioned disordinal interaction can do so definitively.[3]

There have been only three test-expectancy studies, largely overlooked in the literature, that have shown such a disordinal interaction of expected test format and received test format on final test performance that may be attributed to differences in encoding strategies. Von Wright and Meretoja (1975) and von Wright (1977) showed that participants expecting serial recall outperformed those expecting item recognition, and vice versa. Postman and Jenkins (1948) showed such an interaction between anticipation recall (similar to serial recall) and item recognition tests and between free recall and item recognition tests. These results are the exceptions.

In summary, the majority of experiments from the test-expectancy literature, mostly using free recall versus item recognition, have revealed evidence for only a quantitative difference in encoding strategy between test conditions. Why is this the case? We propose that in order for the key disordinal interaction to obtain, the task demands for the two test formats must be different enough that the same encoding strategies do not suffice for attaining performance goals across both test formats (cf. Sanders & Tzeng, 1975). Free recall and item recognition do not meet this requirement. There is evidence that performance on both of these test formats can benefit from both individual item (i.e., distinctive) processing and associative (i.e., relational) processing (Einstein & Hunt, 1980; Hockley & Cristi, 1996). Thus, even if learners perhaps adopt more associative encoding strategies when expect-

ing free recall and more individual-item encoding strategies when expecting item recognition, such strategies may not differentially benefit performance on the two test formats (cf. Neely & Balota, 1981). Furthermore, if the same encoding strategies are appropriate for both test formats, then expectation of those test formats may not actually elicit qualitative differences in encoding strategies at all. In fact, Hall et al. (1976) found that participants expecting *either* of these test formats self-reported predominant use of associative and imagery strategies and that for both test formats there was a positive correlation between how extensively a participant used either type of strategy (as self-rated on a 1–7 scale) and that participant's test performance. That is, the same encoding strategies were indeed beneficial for free recall and item recognition. Drawing on the theoretical model of Anderson and Bower (1974); Maisto et al. (1977) stated that "testing conditions can be varied so that optimal encoding for recall and recognition overlap to a large extent" (p. 130). Thus, free recall and item recognition may overlap too much in their task demands to prompt qualitative differences in encoding strategy or to reflect such differences if they do occur.

Current Study

The current experiments were designed to evaluate and characterize learners' abilities to adaptively and qualitatively modify their encoding strategies. We first sought to establish the elusive disordinal interaction between expected and received test format indicative of qualitative differences in encoding strategy (Experiment 1), and we then sought to better characterize those differences using metacognitive measures (Experiments 2 and 3). In Experiment 1, we used word pairs as learning materials and employed the test-expectancy paradigm with the test formats of cued recall versus target-only free recall. We reasoned that expectation of cued recall should encourage encoding strategies such as cue–target association, while expectation of target-only free recall should encourage encoding strategies such as target–target association and selective attention to the target words. Furthermore, we reasoned that those encoding strategies encouraged by expectation of one test format should not benefit performance on the other test format. In Experiment 2, we investigated adaptive changes in metacognitive monitoring (measured by judgments of learning) across study–test cycles and test formats, because accurate monitoring is necessary to effectively guide control of encoding strategy. In Experiment 3, we further investigated adaptive changes in metacognitive control by providing learners with experience on both test formats and allowing them control over study-time allocation. Key results will be reported in the Results and Discussion sections for each experiment, and analyses of strategy efficacy and strategy usage effectiveness will be reported in the General Discussion.

---

[2] Tversky (1973) only found a disordinal interaction in Experiment 2, in which participants were provided with appropriate encoding strategies in the instructions.

[3] Pashler, McDaniel, Rohrer, and Bjork (2008) made a similar claim in the domain of learning styles and also found a notable absence of such strong evidence. See their article for a more thorough treatment of the interpretive value of disordinal interactions.

Experiment 1

Across four study–test cycles, participants were induced to expect either cued or free recall tests by studying lists of word pairs and receiving the same test format for each list. Tests required recall of target words, either in the presence (cued) or absence (free) of cue words. A final fifth cycle included either the expected or the alternate, unexpected test format. Using two test formats that required production of the same information under qualitatively different task demands, we predicted that participants would adopt qualitatively different encoding strategies and that this would result in a disordinal interaction in final recall performance such that for both final test formats, participants who had expected that format would outperform participants who had expected the other format. The use of multiple study–test cycles allowed us to observe the development of differential encoding strategy use across experience with the test formats. Word-pair associative strength was included as a manipulation that should affect cued recall performance more than free recall performance,[4] thus providing a variable on which we could observe differences in metacognitive monitoring and control in Experiments 2 and 3, respectively. Self-report questions and associative and item recognition tests were given after the final recall test in order to provide more insight on the nature and development of the encoding strategies participants used during the five study–test cycles.

Method

Participants. One hundred undergraduates (47 women) participated in partial fulfillment of course requirements. Data were not recorded for two additional participants due to computer error.

Design. The experiment used a $2 \times 2 \times 2$ mixed design with two between-subjects variables (expected final test format [cued recall vs. free recall] and received final test format [cued recall vs. free recall]) and one within-subjects variable (word-pair associative strength [related vs. unrelated]). In addition, the target (right-hand) words of the pairs were counterbalanced within-subjects such that half were high frequency (mean Hyperspace Analogue to Language [HAL] frequency [$M_{HALfreq}$] = 22,125, $SD_{HALfreq}$ = 29,711; Balota et al., 2007; Lund & Burgess, 1996), and half were low frequency ($M_{HALfreq}$ = 9,413, $SD_{HALfreq}$ = 11,033).[5] Dependent measures were performance on each of five recall tests (either cued recall or free recall), responses to open-ended self-report questions on encoding strategy use, and performance on a final associative recognition test and final item recognition test.

Materials. Materials were 160 English word pairs, divided into five lists of 32 pairs for each participant. All words were nouns composed of between four and eight letters obtained from the Medical Research Council (MRC) Psycholinguistic Database (Coltheart, 1981). Target words were chosen for high imageability (M = 577.3, z = 1.27, SD = 32.0) and high concreteness (M = 576.6, z = 1.16, SD = 33.8). Cue words had mean imageability of 540.0 (z = 0.83, SD = 65.4), mean concreteness of 530.7 (z = 0.77, SD = 84.7), and mean frequency of 18,760 ($SD_{HALfreq}$ = 30,718).

The word pairs had a mean forward associative strength of .027 (SD = .006, mdn = .027, range = .015–.041) and a mean backward associative strength of .027 (SD = .077, mdn = 0, range = 0–.682), as obtained from the University of South Florida Word Association, Rhyme, and Word Fragment Norms (Nelson, McEvoy, & Schreiber, 1998). No cue word had any forward associative strength to other target words within a given list. For each participant, half of the word pairs were randomly selected to remain intact (related, e.g., *flight–bird*), and the other half were transformed into unrelated pairs (e.g., *trumpet–planet*) by randomly reassigning targets to cues such that no target word retained its original cue. Averaged across participants, the mean forward and backward associative strengths of these rearranged pairs were both less than .001. For all word pairs, the cue word was always presented on the left and the target word was always presented on the right. For each participant, word pairs were randomly placed into each of the five presentation lists, with the constraint that the two levels of associative strength were equally represented in each list.

Procedure. Participants were tested individually on computers programmed with Matlab using the Psychophysics Toolbox extensions (Brainard, 1997). All instructions and stimuli were presented visually on the computer screen, and all participant responses were made on the keyboard. Participants were randomly assigned to one of four between-subjects conditions (n = 25 for each group): expected cued recall and received cued recall (CC), expected cued recall and received free recall (CF), expected free recall and received cued recall (FC), and expected free recall and received free recall (FF). The procedure consisted of four expectancy-inducing study–test cycles, a final critical study–test cycle, an open-ended self-report, and two recognition tests.

Expectancy-inducing study–test cycles. Participants first read instructions stating that they would be studying a series of word pairs on which they would later be tested. No details were given regarding test format. Participants were then presented with the first list of 32 word pairs, in a randomized order, one pair at a time for 4 s each, with an interstimulus interval of 0.5 s. They then engaged in an arithmetic distractor task for approximately 45 s. Finally, participants completed a test on the list they had just studied. The test format was either cued recall or free recall, as determined by the expectancy condition to which each participant had been randomly assigned.

In a cued recall test, participants completed a series of 32 trials, one for each of the word pairs they had just studied, in a randomized order. Each test trial showed a cue (left-hand) word and instructed participants to type the corresponding target word or to type a question mark if they could not remember the word. There was no time limit, and no feedback was given.

---

[4] Examining the data from Thomson and Tulving (1970, Experiment 1), it is clear that target-only free recall did *not* reliably differ for targets studied with weakly associated cues (M = 10.7, SD = 2.4) versus strongly associated cues (M = 12.2, SD = 3.6), $t(28)$ = 1.30, p = .145, and furthermore that cued recall *did* reliably differ for targets studied and tested with weak cues (M = 15.7, SD = 4.0) versus strong cues (M = 20.2, SD = 3.4), $t(28)$ = 3.21, $p < .001$.

[5] We originally included target word frequency as a manipulation that should affect performance more for free recall than cued recall, but across lists this variable did not reliably affect performance on either test format. This is perhaps unsurprising given that word frequency effects are not always found in recall (cf. Gregg, 1976). For brevity, we do not further report the effects of this manipulation.

In a free recall test, participants saw a screen with 32 empty boxes in which they were instructed to type only the target (right-hand) words from the list of word pairs they had just studied, in any order. Participants' responses remained onscreen throughout the test, but participants could not go back and edit them. Participants were instructed to press the *Enter* key repeatedly to cycle through all of the remaining empty boxes if they could not remember any more words. There was no time limit, and no feedback was given.

Participants completed this entire study–test cycle a total of four times, with a new list of word pairs for each cycle, and the same test format for all four cycles. That is, a given participant received either four cued recall cycles or four free recall cycles. This was intended to induce the expectancy that they would receive that same format in a final critical study–test cycle.

Final critical study-test cycle. After completing the first four study-test cycles, participants completed a final fifth cycle that featured either the same test format as the previous four (the expected format) or the alternative, unexpected test format, as determined by the final test format condition to which each participant had been randomly assigned. The test formats, cued recall and free recall, were as previously described.

Note that the final list was the same length as the previous four, and presentation was not preceded by any special instructions that might alert participants that this would be the last cycle or that anything about the upcoming test might be different. This is in contrast to some previous test-expectancy experiments (e.g., Balota & Neely, 1980; Neely & Balota, 1981; Thiede, 1996), in which final lists were either much longer than the previous "practice" lists, or participants were instructed that they were about to be presented with the final list, or both. New instructions might conceivably prompt participants to alter their encoding strategies, and Leonard and Whitten (1983) found that some participants spontaneously reported that they had changed their encoding strategy once they realized that the critical list was longer than the previous lists. Thus, the current study did nothing to alert participants that they were practicing for any kind of final critical test.

Self-report on encoding strategy. After completing the fifth recall test, participants responded to two self-report questions. The first question was: "What did you do to try to remember the words for the tests, and did that change as you proceeded through the tests?" The second question varied by condition. For participants who had received an unexpected test format, the second question was: "You received a final test that was different from the previous ones. How did your experience on that test differ from the others, and what might you have done differently to better prepare for that final test?" For participants who had received an expected test format, the second question was: "You received the same type of test throughout the experiment. Looking back, what might you have done differently to better prepare for the final test?" There was no time limit for answering these questions.

Recognition tests. Participants then completed a final associative recognition test followed by a final item recognition test. There had been no prior warning to participants that they would receive such tests.

The associative recognition test consisted of a series of 80 trials in a random order. In each trial, participants saw a word pair, made a yes/no response to indicate whether or not that word pair was in the previously studied lists exactly as shown (i.e., the cue and

target correctly matched), and gave a confidence rating for their answer (1 = *sure*, 2 = *maybe*, 3 = *guess*). Half of the word pairs from each of the five previously studied lists (an equal number of related and unrelated pairs) were randomly selected for this test, with half of these remaining intact (i.e., presented exactly as before) and the other half becoming rearranged lures (i.e., targets randomly reassigned to cues from among all lure pairs). Both the intact pairs and the lure pairs consisted of an equal number of pairs that were previously related versus unrelated. There were no words that had not been previously presented, and words always appeared on the same side of a pair as previously presented (i.e., cues on the left and targets on the right). There was no time limit, and no feedback was given.

The item recognition test consisted of a series of 120 trials in a random order. In each trial, participants saw a single word, made a yes/no response to indicate whether or not that word was in the previously studied lists, and gave a confidence rating for their answer (1 = *sure*, 2 = *maybe*, 3 = *guess*). There were an equal number (40) of lure words, previously studied cue words, and previously studied target words. Lure words were nouns that had not been previously presented and that were similar to the target words in length, imageability, concreteness, and frequency, and were not related to cue or target words (mean forward and backward associative strengths < .001). An equal number of cue words and target words were randomly selected from all five previously studied lists and from both related and unrelated word pairs. No words that had appeared in the associative recognition test were reused in the item recognition test. There was no time limit, and no feedback was given.

Results and Discussion

An alpha level of .05 was used for all tests of statistical significance unless otherwise noted. Effect sizes for comparisons of means are reported as Cohen's *d*, which we calculated using the pooled standard deviation of the groups being compared (Olejnik & Algina, 2000, Box 1, Option B). Effect sizes for analyses of variance (ANOVAs) are reported as $\hat{\omega}^2_{partial}$ calculated with the formulae provided by Maxwell and Delaney (2004, p. 578). Mauchly's test was used to detect violations of sphericity for within-subjects factors in ANOVAs, and in such cases degrees of freedom were adjusted using the Greenhouse–Geisser estimate of ε.

Differences and changes in encoding strategy.

Recall on final critical test. Figure 1 shows mean performance on the final critical recall test as a function of received final test format and expected final test format. The critical comparison was whether, for both final test formats, participants who had expected that format outperformed participants who had expected the other format. This was indeed the case. A two-way between-subjects ANOVA revealed a reliable disordinal interaction between expected final test format and received final test format, $F(1, 96) = 40.28$, mean square error (MSE) = 0.035, $\hat{\omega}^2_{partial}$ = .28, $p < .001$, such that on a final cued recall test, participants who had expected cued recall ($M = 0.51$, $SD = 0.26$) outperformed participants who had expected free recall ($M = 0.25$, $SD = 0.19$), $t(48) = 3.90$, $p < .001$, $d = 1.13$, and on a final free recall test, participants who had expected free recall ($M = 0.27$, $SD = 0.16$)
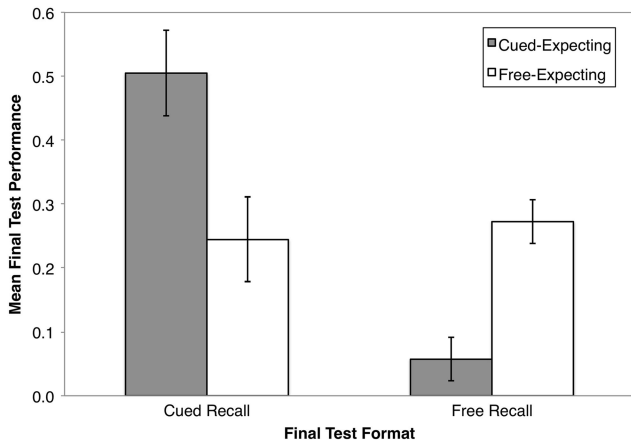
Figure 1. Mean final recall performance as a function of received test format (cued vs. free) and expected test format (cued vs. free) in Experiment 1. Error bars represent the pooled standard errors for comparison of expectancy conditions within each received test format.

outperformed participants who had expected cued recall ($M = 0.06$, $SD = 0.05$), $t(48) = 6.32$, $p < .001$, $d = 1.83$.

We interpret this disordinal interaction as evidence that participants expecting the two different test formats employed qualitatively different encoding strategies. However, a possible alternative class of explanations is worth considering: differences at retrieval. For example, poorer performance on the unexpected test formats could be due to a lack of practice with that test format, or even due to reduced motivation in participants receiving an unexpected test format. The retrieval practice possibility was considered by Hall et al. (1976) and by Balota and Neely (1980); the latter found no difference in final recognition or free recall performance by participants who had received practice with that test format versus participants who had not (i.e., balanced vs. unbalanced practice). Though this issue cannot be resolved with just the final test data, many of the remaining analyses bear on this issue, and it will become evident that the disordinal interaction was in fact due to differences in encoding strategies.

Analyses of intrusions are revealing. On a final test of cued recall, there were more intrusions of *target* words from any list by participants expecting free recall ($M = 2.68$, $SD = 2.57$) versus participants expecting cued recall ($M = 0.44$, $SD = 0.80$), $t(48) = 4.07$, $p < .001$, $d = 1.15$. On a final test of free recall, there were more intrusions of *cue* words from any list by participants expecting cued recall ($M = 0.52$, $SD = 0.70$) versus participants expecting free recall ($M = 0.16$, $SD = 0.37$), $t(48) = 2.23$, $p = .030$, $d = 0.63$. These results suggest that participants expecting free recall employed encoding strategies that enabled them to free recall targets but not to correctly place targets with their corresponding cues, whereas participants expecting cued recall employed encoding strategies that enabled them to correctly recall a target for a given cue but not to avoid free recalling cue words.

Recall across Tests 1–4. First, we consider mean performance across Recall Tests 1–4 for cued recall versus free recall. Higher overall performance levels for cued recall, $t(98) = 12.42$, $p < .001$, $d = 2.51$, are expected and not of interest; the tests simply differ in their global difficulty. Of interest is the fact that

participants receiving repeated free recall tests improved their performance across tests, showing a learning-to-learn pattern (Postman, 1964, 1969). We confirmed this effect by separate simple linear regressions predicting performance from list number for each participant receiving free recall, $M_b = 0.019$, $SD_b = 0.043$, $t(49) = 3.18$, $p = .003$. That is, the mean slope (b) of participant performance across lists was reliably positive. Because this improvement was in the face of considerable proactive interference, which often leads to decreases in memory performance across lists (cf. Diaz & Benjamin, 2011; Wixted & Rohrer, 1993), it suggests that these participants were increasingly able to utilize encoding strategies that were suited to the upcoming test. Cued recall performance did not reliably change across lists, $M_b = 0.005$, $SD_b = 0.059$, $t(49) = 0.60$, $p = .553$.

Figure 2 (top panel) shows mean performance as a function of list number (1–4), test format (cued vs. free), and associative strength (related vs. unrelated). A three-way mixed ANOVA revealed a reliable two-way interaction between test format and associative strength, $F(1, 98) = 89.92$, $MSE = 0.019$, $p < .001$, $\hat{\omega}^2_{partial} = .079$, such that performance was superior for related versus unrelated word pairs to a much greater degree for cued recall,
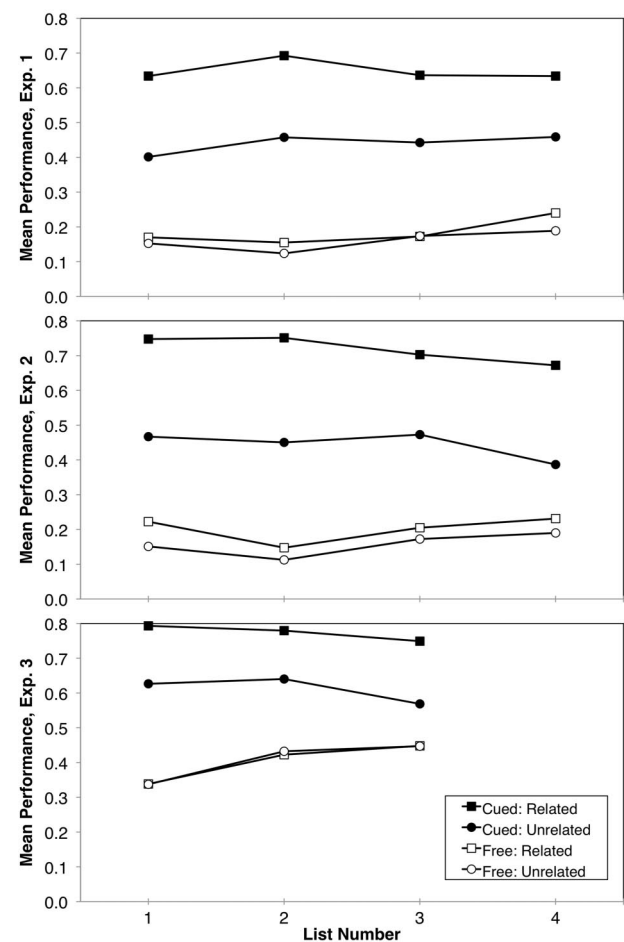

Figure 2. Mean recall performance as a function of list number, test format, and associative strength in Experiments 1, 2, and 3. Exp = Experiment.

$F(1, 49) = 162.10$, $MSE = 0.027$, $p < .001$, $\hat{\omega}^2_{partial} = .204$, than for free recall, $F(1, 49) = 5.62$, $MSE = 0.011$, $p = .022$, $\hat{\omega}^2_{partial} = .018$. There was no reliable three-way interaction, $F(3, 294) = 1.94$, $MSE = 0.011$, $p = .123$, $\hat{\omega}^2_{partial} < .001$, and list number did not interact with associative strength, $F(3, 294) = 1.17$, $MSE = 0.011$, $p = .320$, $\hat{\omega}^2_{partial} < .001$. Thus, as predicted, across all lists, associative strength was a very important variable for cued recall but not for free recall.

Characterizing the encoding strategies used. Analyses of the recognition and self-report data provide insight into the characteristics of the encoding strategies participants used when expecting cued recall versus free recall and provide evidence that the critical disordinal interaction in final recall was indeed due to expectancy-induced differences in encoding strategy. Recognition data were not recorded for some participants; sample sizes are reported in tables.

Associative recognition. Table 1 shows associative recognition performance (d') as a function of test expectancy (cued vs. free recall) and the list number from which the word pairs originated. Overall performance was greater for cued-expecting versus free-expecting participants, $t(41) = 4.07$, $p < .001$, $d = 1.27$. This result suggests that cued-expecting participants attended more to the associations between cues and targets during study than did free-expecting participants, enabling them to better recognize the correctly associated pairs. However, it is also the case that cued recall tests would have provided practice with associative retrieval, perhaps thereby enhancing associative recognition performance. To analyze performance as a function of list of origin, we performed separate simple linear regressions for each participant, and the slopes were then averaged within expectancy condition. Performance by free-expecting participants reliably declined across lists of origin, whereas performance by cued-expecting participants did not reliably change across lists (see Table 1 for inferential statistics). These results suggest that, across lists, free-expecting participants changed their encoding strategies to ones in which less attention was paid to the connection between cues and targets.

Item recognition. Table 2 shows item recognition performance (d') as a function of test expectancy (cued vs. free recall) and item type (cues vs. targets). A two-way mixed ANOVA

revealed a reliable disordinal interaction between test expectancy and item type, $F(1, 41) = 70.43$, $MSE = 0.046$, $p < .001$, $\hat{\omega}^2_{partial} = .058$. For cued-expecting participants, performance was greater for cue words than for target words, but for free-expecting participants the opposite was true. Furthermore, cued-expecting participants outperformed free-expecting participants for both cue words and target words (see Table 2 for inferential statistics).

Cued-expecting participants had seen the cue words twice as often as the target words (once during study and once during the recall tests) and twice as often as did the free-expecting participants, so their superior performance on cue words was expected. The superior target recognition of cued-expecting versus free-expecting participants may be explained by cued recall having afforded more successful retrievals of targets than did free recall (i.e., the testing effect; cf. Roediger & Karpicke, 2006). Of key interest is that free-expecting participants recognized target words better than cue words, the opposite of the pattern for cued-expecting participants. This suggests that free-expecting participants selectively attended to the target words and were thus less able to recognize the cue words.

Table 3 shows item recognition performance (hit rate) as a function of test expectancy (cued vs. free recall), item type (cues vs. targets), and the list number from which the words originated. Hit rates were used rather than d' because it would have been uninformative to compute d' by list of origin given that the lure words originated from no previous list. Separate simple linear regressions were performed for each participant for cue words and for target words and then averaged within expectancy condition. Cue word recognition by free-expecting participants reliably declined across lists of origin; in no other condition did performance reliably change across lists (see Table 3 for inferential statistics). Thus, across lists, free-expecting participants showed a steady decline in recognition of cues but not targets, suggesting that these participants paid less attention to the cue words as they gained experience with a task for which cues were not important. These results also suggest that cued-expecting participants consistently paid attention to both cue and target words, as both words were important for the task of cued recall.

Self-reports on encoding strategy. The mean amount of time spent on the self-report was 158.9 s ($SD = 71.3$). A one-way

Table 1

Means (and Standard Deviations) of Associative Recognition Performance in Experiments 1–3

Expected test format	n	List of origin					Overall	Slope	t	p
		1	2	3	4	5				
Experiment 1										
Cued recall	21	1.70 (0.88)	2.15 (0.72)	2.13 (0.67)	2.00 (0.81)	1.94 (0.98)	2.18 (0.84)	0.03 (0.23)	0.61	.547
Free recall	22	1.55 (0.84)	1.48 (0.79)	0.99 (0.90)	1.03 (1.02)	0.75 (0.97)	1.15 (0.78)	−0.20 (0.28)	−3.28	**.004**
Experiment 2										
Cued recall	51	2.17 (0.69)	2.17 (0.52)	1.96 (0.84)	2.09 (0.79)		2.33 (0.73)	−0.04 (0.30)	−1.06	.293
Free recall	49	2.07 (0.61)	1.62 (0.96)	1.72 (0.82)	1.44 (0.99)		1.78 (0.78)	−0.18 (0.32)	−3.88	**<.001**
Experiment 3										
Cued recall	77[a]	1.76 (0.57)	1.71 (0.68)	1.75 (0.51)			1.74 (0.42)	−0.01 (0.35)	−0.11	.910
Free recall	77[a]	1.34 (0.76)	0.65 (0.84)	0.48 (0.86)			0.82 (0.52)	−0.43 (.58)	−6.48	**<.001**

Note. Experiment 1 data are only from participants who received their expected final test format; performance measure was d'; statistically significant p values are shown in boldface.
[a] Test format was manipulated within-subjects in Experiment 3.

Table 2
Means (and Standard Deviations) of d′ Item Recognition Performance in Experiments 1 and 2

		Item type				
Expected test format	n	Cues	Targets	t	p	d
Experiment 1						
Cued recall	21	2.28 (1.02)	1.76 (0.86)	7.19	**<.001**	0.11
Free recall	22	0.93 (0.55)	1.18 (0.52)	−4.34	**<.001**	−0.10
t		5.23	2.61			
p		**<.001**	**.013**			
d		1.66	0.82			
Experiment 2						
Cued recall	51	2.39 (0.94)	1.93 (0.81)	6.84	**<.001**	0.07
Free recall	49	1.17 (0.55)	1.33 (0.67)	−2.35	**.023**	−0.03
t		7.42	3.99			
p		**<.001**	**<.001**			
d		1.49	0.80			

Note. Experiment 1 data are only from participants who received their expected final test format; statistically significant p values are shown in boldface.

between-subjects ANOVA revealed that this value did not reliably differ across conditions, $F(3, 96) = 0.68$, $MSE = 5187.66$, $p = .568$, $\hat{\omega}^2_{partial} < .001$. Participants' responses to the self-report questions were coded by one of the experimenters using a rubric of binary codes devised from the experimenters' intuitions and from informal observation of the range of participants' responses. Participants' experimental conditions were concealed during coding.

In total, twelve specific strategies were identified and coded (Appendix A). Table 4 shows the self-report frequencies of each strategy for both expectancy conditions. We compared the proportion of participants reporting each strategy for cued recall expectation versus free recall expectation, using a Bonferroni corrected alpha level of .0042 (i.e., .05/12). The only two strategies for which proportions reliably differed across expectancy were also the most frequently reported strategies for each condition. For participants expecting cued recall, the most frequently reported strategy was making cue–target associations (e.g., "I tried to find some connection between the two words that were paired"), and

this was reported with reliably greater frequency than by free-expecting participants. For participants expecting free recall, the most frequently reported strategy was selectively attending to the target words (e.g., "[T]oward the end I just started memorizing the last word and not really paying attention to the first word"), and this was reported with reliably greater frequency than by cued-expecting participants. One other strategy approached significance in being more frequently reported by free-expecting participants: making target–target associations (e.g., "Then I started associating the second word from each pair together"). Finally, more free-expecting than cued-expecting participants reported that they changed strategies across lists (41/50 vs. 17/50), $z = 4.86$, $p < .001$. Thus, participants in both expectancy conditions reported having ultimately used encoding strategies that were appropriate for the test format they expected, and for free-expecting participants, this appeared to require more shifting from initial strategies.

Finally, we considered common ways in which participants reported that they would have changed their encoding strategies to

Table 3
Means (and Standard Deviations) of Hit Rate Item Recognition Performance Across Lists of Origin in Experiments 1 and 2

		List of origin							
Test format and item type	n	1	2	3	4	5	Slope	t	p
Experiment 1									
Cued recall	21								
Cue words		.83 (.14)	.89 (.13)	.85 (.18)	.86 (.15)	.84 (.18)	−.002 (.04)	−0.20	.846
Target words		.72 (.21)	.76 (.18)	.72 (.17)	.77 (.19)	.71 (.22)	−.001 (.07)	−0.04	.970
Free recall	22								
Cue words		.72 (.21)	.68 (.18)	.60 (.23)	.55 (.29)	.50 (.18)	−.056 (.06)	−4.42	**<.001**
Target words		.70 (.23)	.60 (.25)	.72 (.18)	.73 (.16)	.73 (.20)	.019 (.06)	1.47	.157
Experiment 2									
Cued recall	51								
Cue words		.85 (.16)	.88 (.14)	.88 (.16)	.87 (.15)		.006 (.06)	0.77	.443
Target words		.79 (.16)	.77 (.18)	.78 (.18)	.74 (.20)		−.012 (.07)	−1.28	.207
Free recall	49								
Cue words		.69 (.22)	.69 (.19)	.59 (.22)	.52 (.26)		−.063 (.09)	−4.68	**<.001**
Target words		.70 (.17)	.61 (.20)	.69 (.20)	.71 (.21)		.010 (.07)	1.05	.298

Note. Experiment 1 data are only from participants who received their expected final test format; statistically significant p values are shown in boldface.

Table 4

Frequencies of Self-Reported Encoding Strategies in Experiment 1

	Expected test format		Cued vs. free	
Encoding strategy	Cued recall	Free recall	z	p
Cue–target association	27	9	3.75	**<.001**
Target–target association	0	7	−2.74	.006
Unspecified association	8	9	−0.27	.790
Target focus	3	35	−6.59	**<.001**
Mental imagery	14	7	1.72	.086
Rote rehearsal	9	18	−2.03	.043
Verbalization	7	3	1.33	.182
Narrative	9	8	0.27	.790
Personal significance	6	6	0.00	>.999
Bizarre	1	2	−0.59	.558
Action	0	2	−1.43	.153
Phonetic	2	2	0.00	>.999

Note. Both test formats: $n = 50$; statistically significant p values are shown in boldface (Bonferroni corrected α level of .0042).

better prepare for the final test. Quantitative changes such as trying harder or paying more attention overall were not coded. The most frequent response from participants who received a final free recall test (whether expected or not) was that they would have focused more on the target words. Participants who both expected and received a final cued recall test reported few changes that they would have made to their encoding strategies. An illustrative example response from a participant who expected cued recall but received free recall was the following:

> I didn't remember much on the last test. My word associated method did absolutely nothing for me. I would have only looked at the second word and just tried to memorize them or associate them with other second words instead.

Participants who had expected a final free recall test but received a final cued recall test reported that they would have attended more to the cue words or that they would have made more cue-target associations. An illustrative example response from such a participant was:

> It was easier to recall, but I had become so used to just looking at the second word that being given the extra stimuli to remember didn't actually help that much. I think that if I had paid more attention to the first words than I would have done better.

Thus, in both of the unexpected conditions, participants reported that they would have made more usage of encoding strategies that were appropriate for that unexpected test format.

Summary of results. Taken together, these results suggest that participants indeed came to strategically employ qualitatively different encoding strategies that were appropriate to the expected test format. For both final test formats, participants who expected that format outperformed those who had not. Furthermore, performance increased across lists for free recall but not for cued recall, and associative strength of word pairs affected performance for cued recall more than free recall. Finally, the recognition and self-report data suggest that participants expecting cued recall maintained encoding strategies of cue–target association, while

participants expecting free recall came to strategically focus more on the target words.

Experiment 2

Tailoring an encoding strategy to the demands of an expected test format requires learners to attune their awareness to those characteristics of the learning material that are relevant to that test format. Thus, accurate metacognitive monitoring is necessary to effectively guide metacognitive control (cf. Finley et al., 2010; Hertzog and Dunlosky, 2004; Tullis & Benjamin, 2011). Given the effective differences and changes in encoding strategy observed in Experiment 1, we should also be able to observe adaptive changes in metacognitive monitoring, as measured by judgments of learning (JOLs). Thus, we predicted that across study–test cycles, JOLs would increasingly diverge such that they would reflect the associative strength of word pairs to a greater degree for participants expecting cued recall (for which associative strength is important) versus participants expecting free recall (for which associative strength is less relevant). To test this prediction, we used a procedure in Experiment 2 that was similar to that in Experiment 1, but with JOLs made for each item during presentation and with only four study–test cycles and no conditions that violated test expectancy (i.e., no unexpected test formats).

Method

Participants. One hundred three undergraduates (60 women) participated for partial fulfillment of course requirements.

Design. The experiment used a 2 × 2 mixed design with one between-subjects variable (expected test format [cued recall vs. free recall]) and one within-subjects variable (word pair associative strength [related vs. unrelated]). In addition, the target (right-hand) words of the pairs were counterbalanced within-subjects such that half were high frequency ($M_{HALfreq} = 82,522$, $SD_{HALfreq} = 101,563$), and half were low frequency ($M_{HALfreq} = 2,868$, $SD_{HALfreq} = 2,234$). Dependent measures were performance on each of four recall tests (either all cued recall or all free recall), responses to a questionnaire on encoding strategy use, and performance on a final associative recognition test and final item recognition test.

Materials. Materials were 128 English word pairs (all but three of which were different from those used in Experiment 1),[6] divided into four lists of 32 pairs for each participant. As in Experiment 1, all words were nouns composed of between four and eight letters, with target words chosen for high imageability ($M = 581.9$, $z = 1.22$, $SD = 30.2$) and high concreteness ($M = 579.1$, $z = 1.18$, $SD = 33.1$). Cue words had mean imageability of 529.5 ($z = 0.74$, $SD = 70.3$), mean concreteness of 523.8 ($z = 0.72$, $SD = 94.0$), and mean frequency of 14,595 ($SD_{HALfreq} = 25,648$). Word pairs had mean forward associative strength of .030 ($SD = .006$, $mdn = .030$, range = .020–.039) and mean backward associative strength of .054 ($SD = .152$, $mdn = 0$, range = 0–.913). For each participant, associative strength was manipu-

---

[6] Materials were changed from those of Experiment 1 in order to make the target word frequency manipulation stronger with the hope that it would affect free recall more than cued recall. However, we still did not obtain such an effect and thus do not report any word frequency analyses.

lated, and pairs were placed into lists as described in Experiment 1. Averaged across participants, the mean forward and backward associative strengths of these rearranged pairs were both $< .001$.

Procedure. The procedure was similar to that of Experiment 1, with the major changes being the omission of the fifth study–test cycle, and the addition of JOLs during the presentation phase of the study–test cycles. Participants were randomly assigned to receive either all cued recall tests ($n = 53$) or all free recall tests ($n = 50$). The procedure consisted of four expectancy-inducing study–test cycles, a questionnaire on encoding strategy use, and two recognition tests.

Expectancy-inducing study–test cycles. The four expectancy-inducing study–test cycles were identical to those described in Experiment 1, with the addition of JOLs following the presentation of each word pair. After a word pair had been shown for 4 s, the following JOL prompt appeared: "How sure are you that you will remember this item on the test?" Participants responded using a scale ranging from 1 (*I am sure I will NOT remember this item*) to 4 (*I am sure I WILL remember this item.*). The presented word pair remained visible during the judgment. There was no time limit for responding, and each trial was followed by a 0.5-s interstimulus interval.

Questionnaire on encoding strategy. We devised an encoding strategy questionnaire based on the self-report data from Experiment 1 and on the learning strategy questionnaire used by Leonard and Whitten (1983, Appendix), which was in turn adapted from Hall et al. (1976). Participants completed the questionnaire on paper following the fourth study–test cycle. For each of 11 specific strategies (listed in Appendix B), participants answered two questions: "How frequently did you engage in the following study strategies during the experiment so far?" to which participants responded on a scale from 1 (*no use*) to 7 (*extensive use*), and "When during the experiment so far did you use this strategy more frequently?" to which participants responded by choosing "1st half," "2nd half," or "same or N/A." Participants could also write in any additional unlisted strategies they had used. Finally, participants indicated whether they thought that the type of test would change over the lists (yes vs. no), and, if yes, they indicated whether they stopped suspecting a change during the first half or the second half, or stayed suspicious the whole time. There was no time limit for the completing the questionnaire.

Recognition tests. Participants then completed a final associative recognition test followed by a final item recognition test. The procedure for these tests was the same as that in Experiment 1, except that there were 64 trials for the associative recognition test and 96 trials for the item recognition test, and no confidence ratings were made. The lure words had the same characteristics as those described for Experiment 1. Again, there was no time limit, and no feedback was given. Also as before, no item appeared on both tests.

Results and Discussion

Recall performance. First, we considered mean performance across Recall Tests 1–4 for cued recall versus free recall. Separate simple linear regressions for each participant revealed that cued recall performance reliably declined across lists, $M_b = -0.025$, $SD_b = 0.066$, $t(52) = -2.68$, $p = .009$, while free recall perfor-

mance, although showing a positive trend, did not reliably change across lists, $M_b = 0.013$, $SD_b = 0.066$, $t(49) = 1.37$, $p = .177$.

Figure 2 (middle panel) shows mean performance as a function of list number (1–4), test format (cued vs. free), and associative strength (related vs. unrelated). A three-way mixed ANOVA revealed a reliable two-way interaction between test format and associative strength, $F(1, 101) = 104.76$, $MSE = 0.026$, $p < .001$, $\hat{\omega}_{partial}^2 = .125$, such that performance was superior for related versus unrelated word pairs to a much greater degree for cued recall, $F(1, 52) = 181.12$, $MSE = 0.044$, $p < .001$, $\hat{\omega}_{partial}^2 = .347$, than for free recall, $F(1, 49) = 31.20$, $MSE = 0.006$, $p < .001$, $\hat{\omega}_{partial}^2 = .048$. There was no reliable three-way interaction, $F(3, 303) = 1.22$, $MSE = 0.010$, $p = .301$, $\hat{\omega}_{partial}^2 < .001$, and list number did not interact with associative strength, $F(3, 303) = 1.91$, $MSE = 0.010$, $p = .127$, $\hat{\omega}_{partial}^2 = .002$. Thus, as in Experiment 1, across all lists, associative strength was a more important variable for cued recall than for free recall.

Metacognitive monitoring. Figure 3 shows mean JOLs as a function of list number (1–4), test format (cued vs. free), and associative strength (related vs. unrelated). A three-way mixed ANOVA revealed a reliable three-way interaction, $F(3, 303) = 6.38$, $MSE = 0.046$, $p < .001$, $\hat{\omega}_{partial}^2 = .006$, such that, across lists, the JOLs made by free-expecting participants decreasingly differentiated between related and unrelated pairs, $F(2.4, 117.9) = 40.05$, $MSE = 0.067$, $\hat{\epsilon} = .802$, $p < .001$, $\hat{\omega}_{partial}^2 = .101$, and did so to a greater degree than did those made by cued-expecting participants, $F(2.5, 128.9) = 14.31$, $MSE = 0.047$, $\hat{\epsilon} = .826$, $p < .001$, $\hat{\omega}_{partial}^2 = .024$. We further confirmed this pattern by performing separate simple linear regressions predicting difference scores (mean JOLs for related minus unrelated) from list number for each participant. The mean JOL difference scores for participants receiving free recall reliably declined across lists, $M = -0.22$, $SD = 0.19$, $t(49) = 8.28$, $p < .001$. Although this was also true for participants receiving cued recall, $M = -0.10$, $SD = 0.16$, $t(52) = 4.84$, $p < .001$, it happened to a reliably lesser extent than for those receiving free recall, $t(101) = 3.34$, $p = .001$, $d = 0.67$. Free-expecting participants' JOLs reflected associative strength less and less over time, which was appropriate given that this characteristic of the word pairs was not relevant to their task. Taken together, these JOL findings complement other studies that have shown that the accuracy of learners' metacognitive monitoring (or metacomprehension) is enhanced when encoding tasks are congruent with test formats (Thomas & McDaniel, 2007a, 2007b) and when test expectancies are congruent with test formats (Thiede, Wiley, & Griffin, 2011).

Characterizing the encoding strategies used.

Associative and item recognition. As shown in Table 1, overall associative recognition performance was again greater for cued-expecting versus free-expecting participants, $t(98) = 3.58$, $p < .001$, $d = 0.72$, and performance again reliably declined across list of origin for participants expecting free recall but not cued recall. For item recognition, as shown in Table 2, the same interaction between test expectancy and item type was found as that in Experiment 1: cued-expecting participants better recognized cues versus targets, while free-expecting participants better recognized targets versus cues, $F(1, 98) = 42.53$, $MSE = 0.112$, $p < .001$, $\hat{\omega}_{partial}^2 = .036$. Furthermore, as shown in Table 3, the pattern of item recognition performance across lists of origin was the same as that found in Experiment 1: performance reliably
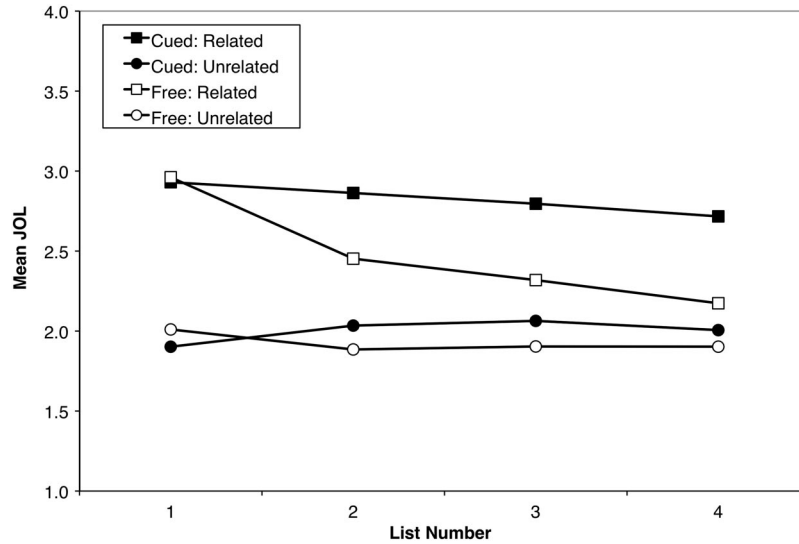
Figure 3. Mean judgments of learning (JOLs) as a function of list number (1–4), test format (cued vs. free), and associative strength (high vs. low) in Experiment 2.

declined across list of origin only in the case of free-expecting participants and cue words.

Questionnaire on encoding strategy. To confirm the same patterns of strategy use as those suggested by the results of Experiment 1, we considered data from the questionnaire. The mean amount of time spent on the questionnaires was 200.9 s ($SD = 44.8$). This value did not reliably differ between test format expectancy conditions, $t(98) = 1.77$, $p = .080$, $d = 0.36$.

Participants' encoding strategy usage frequency ratings are summarized in Table 5. Because the measures were ordinal and because the data were not normally distributed, we made comparisons of ratings for cued recall expectations versus free-recall expectations using the two-sample Kolmogorov–Smirnov test. Because these analyses were preplanned, an unadjusted alpha level was used. Data from participants with missing values were ex-

cluded from analysis on a testwise (i.e., per strategy) basis; thus, *n* varied slightly across tests.

Cued recall expectation yielded greater reported usage of cue–target association, while free recall expectation yielded greater reported usage of target–target association, target focus, and rote rehearsal. The number of different strategies reported (i.e., the count of strategies rated > 1) did not reliably differ for cued versus free recall, $M_{cued} = 8.7$, $SD_{cued} = 1.7$, $M_{free} = 8.4$, $SD_{free} = 2.0$, $t(97) = 0.87$, $p = .388$, $d = 0.18$. This is in contrast to the open-ended self-report data from Experiment 1, in which free-expecting participants spontaneously reported multiple strategies more often than did cued-expecting participants. However, consistent with the data from Experiment 1, free-expecting participants did reliably report more changes in strategy usage than did cued-expecting participants in Experiment 2, as measured by the

Table 5
Encoding Strategy Usage Frequency Ratings in Experiment 2

Encoding strategy	Cued recall expectation		Free recall expectation		Cued vs. free	
	M (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	*z*	*p*
Cue–target association	5.60 (1.92)	6.5	4.96 (1.35)	5.0	1.68	**.001**
Target–target association	2.32 (1.58)	2.0	3.06 (2.22)	2.0	1.23	**.032**
Interitem association	2.58 (1.74)	2.0	2.53 (1.67)	2.0	0.26	.975
Target focus	3.24 (1.74)	3.5	4.58 (1.88)[a]	5.0[a]	1.66	**.001**
Mental imagery	4.98 (1.87)[b]	5.0[b]	4.59 (2.06)	5.0	0.51	.676
Rote rehearsal	4.32 (1.87)	4.0	5.20 (1.48)	5.0	1.28	**.020**
Verbalization	4.12 (2.35)	4.5	3.84 (2.43)	4.0	0.63	.516
Intraitem narrative	4.15 (2.03)[a]	4.0[a]	3.88 (2.36)	5.0	0.79	.276
Interitem narrative	3.39 (2.24)[b]	3.0[b]	2.94 (2.41)	1.0	0.91	.211
Personal significance	4.86 (1.90)	5.5	4.08 (2.21)	5.0	0.96	.141
Observation	4.00 (1.81)	4.0	4.43 (1.69)	4.0	0.57	.570

Note. Rating scale was 1 (*no use*) to 7 (*extensive use*); $n_{cued} = 50$; $n_{free} = 49$. We used the two-sample Kolmogorov–Smirnov test, reporting exact *p* values; statistically significant *p* values are shown in boldface.
[a] $n = 48$. [b] $n = 49$.

proportion of strategies that were rated > 1 for usage and that were also reported as used more in either the first half or the second half of the experiment, $M_{cued} = 0.37$, $SD_{cued} = 0.30$, $M_{free} = 0.63$, $SD_{free} = 0.27$, $t(97) = 4.42$, $p < .001$, $d = 0.90$. Sign tests revealed that free-expecting participants reported more usage in the first half versus the second half of the expectancy-inducing cycles for the strategy of cue–target association ($p = .001$) and more usage in the second half versus the first half for the strategies of target focus ($p < .001$), mental imagery ($p = .004$), intraitem narrative ($p = .023$), and interitem narrative ($p = .041$). Cued-expecting participants reported more usage in the first half versus the second half for the strategy of rote rehearsal ($p = .035$) and more usage in the second half versus the first half for the strategy of personal significance ($p = .019$).

Summary of results. The results of Experiment 2 further bolstered the conclusion that participants used qualitatively different encoding strategies that were appropriate for their expected test format and did so to an increasing extent as they gained experience with the task. Furthermore, participants' metacognitive monitoring—a critical component of self-regulated learning—also became more attuned to the demands of the tasks.

Experiment 3

Experiments 1 and 2 provided evidence of learners' adoption of appropriate and qualitatively different encoding strategies in expectation of two different test formats and also evidence of learners' development of more appropriately attuned metacognitive monitoring. Given these results, we reasoned that it should be possible to provide learners with an experience that would facilitate their learning to better discriminate between the task demands of the two test formats and thus also to more strategically control their study process. Toward this end, in Experiment 3, we employed a within-subjects design in which all participants experienced three cued recall study–test cycles and three free recall study–test cycles and in which participants were accurately informed of the upcoming test format before each study phase. Furthermore, we investigated adaptive changes in control of self-paced study by enabling participants to control study-time allocation (i.e., how long they studied each word pair).

Because we chose to use a fully factorial within-subjects design, it was not feasible to use the critical final test manipulation (as in Experiment 1) for evidence of differences in encoding strategy, as that would require violating participants' expectations more than once. After receiving an unexpected test format for the first time, they would be unlikely to believe any further instructions about upcoming test formats and thus would also be unlikely to continue using encoding strategies specific to one format or the other. Thus, we chose to rely on questionnaire data and associative recognition performance to provide evidence of differences and changes in encoding strategy and to introduce study-time allocation to measure metacognitive control during study (cf. Son & Metcalfe, 2000; Tullis & Benjamin, 2011).

We predicted that participants' recall performance, questionnaire responses, and associative recognition performance would show similar patterns to those observed in Experiments 1 and 2. We also predicted that study-time allocation would come to reflect important differences between the task demands of cued versus free recall: differentiating between related and unrelated pairs for cued recall but not for free recall.

Method

Participants. Eighty-five undergraduates (44 women) participated for partial fulfillment of course requirements.

Design. The experiment used a 2×2 within-subjects design, with independent variables: expected test format (cued recall vs. free recall), and word pair associative strength (related vs. unrelated). Dependent measures were amount of time spent studying each word pair, performance on each of six recall tests (three cued recall and three free recall), responses to a questionnaire on encoding strategy use, and performance on a final associative recognition test.

Materials. Materials were 144 English word pairs, divided into six lists of 24 pairs for each participant. Materials were changed from those of the previous experiments in order to accommodate the greater number of lists. As before, all words were nouns composed of between four and eight letters, with target words chosen for high imageability ($M = 578.5$, $z = 1.19$, $SD = 34.9$) and high concreteness ($M = 572.7$, $z = 1.12$, $SD = 33.4$). Mean target frequency was 21,680 ($SD_{HALfreq} = 34,241$). Cue words had mean imageability of 488.4 ($z = 0.36$, $SD = 73.2$), mean concreteness of 466.6 ($z = 0.24$, $SD = 89.5$), and mean frequency of 21,299 ($SD_{HALfreq} = 37,534$). Word pairs had mean forward associative strength of .032 ($SD = .006$, $mdn = .034$, range $= .020$–$.039$) and mean backward associative strength of .019 ($SD = .050$, $mdn = 0$, range $= 0$–$.384$). For each participant, associative strength was manipulated and pairs were placed into lists as described in Experiment 1. Averaged across participants, the mean forward and backward associative strengths of these rearranged pairs were both $< .001$.

Procedure. The procedure consisted of six expectancy-inducing study–test cycles, a questionnaire on encoding strategy use, and one recognition test.

Expectancy-inducing study–test cycles. Participants first read instructions that they would be studying several lists of word pairs and that they would have unlimited time to study each word pair, but would not be able to return to a pair once they had moved on from it. The instructions also stated that participants would receive either a cued recall or a free recall test on each list after they had finished studying it and before moving on to study the next list. The instructions clearly described both test formats, using an example word pair that did not appear in any of the study lists.

Participants then completed three cued recall study–test cycles (C) and three free recall study–test cycles (F). Participants were randomly assigned to complete these cycles in one of two orders: CFCFCF or FCFCFC. At the start of each cycle, participants read a notification of which list number they were about to study and which test format they would receive for this list, along with a reminder of what that test format required. Participants were then presented with a list of 24 word pairs, in a randomized order, one pair at a time. Each word pair remained on the screen until participants pressed the space bar and was followed by an inter-stimulus interval of 0.5 s. No JOLs were made, and presentation duration was recorded by the computer for each pair. Participants then engaged in an arithmetic distractor task for approximately 45 s. Finally, participants completed a test on the list they had just

studied. The test format that they received always matched the test format that they had been told they would receive for that list. The test formats were as described in Experiment 1, with the exception that there were only 24 trials for cued recall and only 24 empty boxes for free recall. Again, there was no time limit and no feedback was given.

Questionnaire on encoding strategy. Participants completed a paper questionnaire that was similar to that used in Experiment 2. For each of the same 11 encoding strategies (Appendix B), participants rated their usage frequency from 1 (*no use*) to 4 (*extensive use*) for both the cued recall lists and the free recall lists. The questionnaire includes a final question regarding suspicion of test format change that was similar to that in Experiment 2, but participants were asked to indicate on which list they stopped being suspicious (rather than simply indicating which half of the experiment). Finally, the questionnaire instructions reminded participants of the definitions of cued recall and free recall. There was no time limit for the questionnaire.

Recognition test. Participants then completed a final associative recognition test. The procedure for this test was the same as that in Experiment 1, except that there were only 48 trials, and no confidence ratings were made. Again, there was no time limit, and no feedback was given. There was no item recognition test.

Results and Discussion

Recall performance. First we consider mean performance across Recall Tests 1–3 for cued recall versus free recall. Separate simple linear regressions for each participant revealed that cued recall performance reliably declined across lists, $M_b = -0.025$, $SD_b = 0.089$, $t(84) = -2.63$, $p = .010$, while free recall performance reliably increased across lists, $M_b = 0.055$, $SD_b = 0.106$, $t(84) = 4.74$, $p < .001$.

Figure 2 (bottom panel) shows mean performance as a function of list number (1–3), test format (cued vs. free), and associative strength (related vs. unrelated). A three-way within-subjects

ANOVA revealed a reliable two-way interaction between test format and associative strength, $F(1, 84) = 87.05$, $MSE = 0.020$, $p < .001$, $\hat{\omega}^2_{partial} = .043$, such that performance was superior for related versus unrelated word pairs for cued recall, $F(1, 84) = 147.91$, $MSE = .023$, $p < .001$, $\hat{\omega}^2_{partial} = .151$, while performance did not reliably differ as a function of associative strength for free recall, $F(1, 84) = 0.06$, $MSE = 0.015$, $p = .809$, $\hat{\omega}^2_{partial} < .001$. There was no reliable three-way interaction, $F(2, 168) = 0.39$, $MSE = 0.013$, $p = .681$, $< .001$, and list number did not interact with associative strength, $F(2, 168) = 1.12$, $MSE = .014$, $p = .329$, $\hat{\omega}^2_{partial} < .001$. Thus, as in Experiments 1 and 2, across all lists, associative strength was a more important variable for cued recall than for free recall.

Study-time allocation. Analyses of study-time allocation were carried out on participants' median study time (in seconds) per cell. Figure 4 shows study-time allocation as a function of list number (1–3), test format (cued vs. free), and associative strength (related vs. unrelated). A three-way within-subjects ANOVA revealed a reliable three-way interaction, $F(1.6, 137.2) = 4.80$, $MSE = 1.90$, $\hat{\varepsilon} = .817$, $p = .015$, $\hat{\omega}^2_{partial} = .002$. For cued recall, participants consistently spent more time studying unrelated versus related word pairs, as evidenced by a reliable effect of associative strength, $F(1, 84) = 51.79$, $MSE = 2.93$, $p < .001$, $\hat{\omega}^2_{partial} = .037$, and the lack of a two-way interaction between associative strength and list number, $F(1.6, 134.4) = 0.09$, $MSE = 2.13$, $\hat{\varepsilon} = .800$, $p = .873$, $\hat{\omega}^2_{partial} < .001$. For free recall, participants began with the same approach, but decreasingly differentiated between related and unrelated pairs across lists, as evidenced by a reliable two-way interaction between associative strength and the linear effect of list number, $F(1, 84) = 19.44$, $MSE = 1.68$, $p < .001$, $\hat{\omega}^2_{partial} = .007$. It is worth noting that study-time allocation reliably declined across lists for both cued recall ($M_b = -0.86$, $SD_b = 1.54$), $t(84) = -5.10$, $p < .001$, and free recall ($M_b = -1.20$, $SD_b = 2.09$), $t(84) = -5.28$, $p < .001$, which is interesting considering that memory performance actually increased across free recall
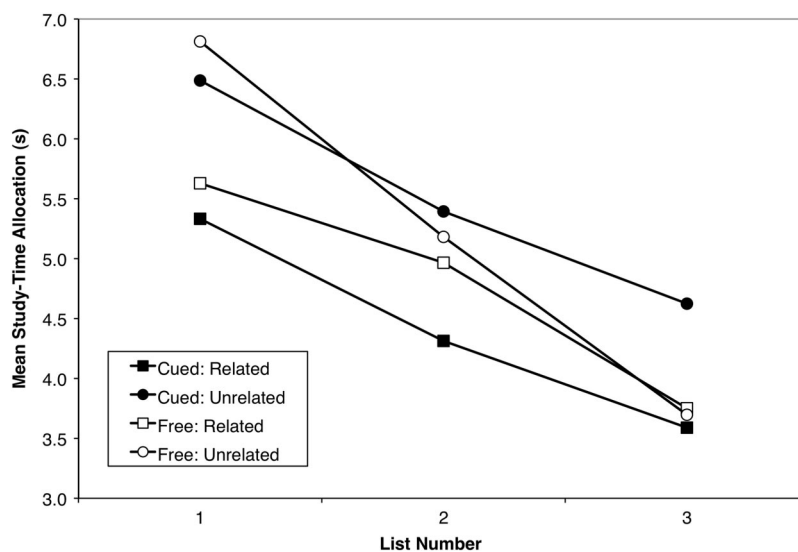

Figure 4. Mean of participant median study-time allocation (in seconds) as a function of list number (1–3), test format (cued vs. free), and associative strength (high vs. low) in Experiment 3.

lists. Participants' earlier encoding efforts for free recall may have been "labor in vain" (Nelson & Leonesio, 1988, p. 676), but as they learned to focus on the target words, their efforts became both more effective and less time-consuming.

Characterizing the encoding strategies used.

Associative recognition. As shown in Table 1, overall associative recognition performance was again greater for cued-expecting versus free-expecting participants, $t(76) = 12.44$, $p < .001$, $d = 1.92$, and performance again reliably declined across list of origin for participants expecting free recall but not cued recall.

Questionnaire on encoding strategy. The mean amount of time spent on the questionnaires was 195.8 s ($SD = 41.4$). Participants' encoding strategy usage frequency ratings are summarized in Table 6. We made planned comparisons of ratings for cued versus free recall expectation using the Wilcoxon matched-pairs signed-rank test.[7] Unadjusted alpha levels were used, and data from participants with missing values were again excluded on a testwise (i.e., per strategy) basis.

The usage ratings show a pattern similar to that of Experiment 2. Cued recall expectation yielded greater reported usage of cue–target association, while free recall expectation yielded greater reported usage of target–target association, target focus, rote rehearsal, and interitem narrative. As in Experiment 2, the number of different strategies reported did not reliably differ for cued versus free recall, $M_{cued} = 7.8$, $SD_{cued} = 2.0$, $M_{free} = 7.8$, $SD_{free} = 2.1$, $t(83) = 0.13$, $p = .899$, $d = 0.01$.

Considered alongside the primary recall results, the recognition, self-report, and questionnaire data from all three experiments paint a consistent picture: participants indeed came to strategically employ qualitatively different encoding strategies based on the test format they expected. It appears that most participants began the experiment using some form of cue–target association strategy and that participants receiving cued recall tests continued to use such a strategy, while participants receiving free recall tests gradually abandoned it in favor of strategies that focused on the targets (cf. Underwood, 1963) and strategies that formed associations across word pairs. The efficacy of the various encoding strategies and the effectiveness of participants' differential usage of them will be analyzed in the General Discussion.

Summary of results. In Experiment 3, individual participants showed qualitative and adaptive differences in encoding strategy and in study-time allocation when they expected two different test formats. Consistent with the results from Experiments 1 and 2, when participants studied for cued recall tests across multiple study-test cycles, they demonstrated sustained use of a cue–target association strategy, and when participants studied for free recall tests across multiple study-test cycles, they abandoned such a strategy in favor of selectively attending to the target word and making associations across pairs. With regard to study time, participants began the experiment by allocating more study time to unrelated word pairs when expecting either test format. As shown in Figure 4, participants continued this pattern of allocation across cued recall study–test cycles but decreasingly differentiated between related and unrelated pairs across free recall study–test cycles. Thus, experience with the nature of a specific test format and the effectiveness of their metacognitive control led learners to increasingly adopt more efficacious encoding strategies and study-time allocation strategies.

General Discussion

Summary of Key Results

In this study, we asked whether learners can adaptively and qualitatively modulate their encoding strategies in anticipation of future task demands. In Experiment 1, the key result was a disordinal interaction (Figure 1) such that, on final tests of both cued recall and free recall, participants who had been led by experience to expect that test format outperformed participants who had been led to expect the other format. Analyses of recognition and self-report data from all three experiments support an encoding strategy interpretation of this interaction. That is, participants demonstrated that they can and do tailor their encoding strategies to fit the demands of the type of test they expect, employing appropriate and qualitatively different strategies for different test formats (e.g., cue–target association for cued recall, and target focus for free recall). In Experiment 2, participants furthermore demonstrated concomitant and judicious attunement of metacognitive monitoring, decreasingly differentiating between related and unrelated word pairs for free recall but not cued recall, as shown in Figure 3. In Experiment 3, in which a within-subjects design was used, participants demonstrated adaptive changes in metacognitive control of encoding strategy and of study-time allocation: participants began the experiment spending more time studying unrelated versus related word pairs for both test formats, and they decreasingly made this distinction for free recall (for which cue–target associative strength was inconsequential), as shown in Figure 4. In the next sections, we will evaluate the actual efficacy of the various self-reported encoding strategies for cued versus free recall, and we will then assess the optimality of participants' strategic differential use of them. We will close by revisiting questions about what types of conditions invite qualitative changes in encoding strategy.

Efficacy of Encoding Strategies

The usage frequency ratings from the questionnaires in Experiments 2 and 3 (to the extent that they are accurate) allowed us to evaluate the actual efficacy of the various encoding strategies at improving recall performance across lists and to compare that efficacy for cued versus free recall. We first performed separate simple linear regressions predicting recall performance from list number for each participant. The estimated slopes from these

---

[7] Because of the small ordinal scale used, there were many ties and potentially many difference scores with a value of zero. Tied difference scores were assigned the mean of the ranks involved in that tie. Furthermore, the Wilcoxon test statistic (z) was calculated using the large sample normal approximation with correction for continuity and correction for ties as provided by Marascuilo and McSweeney (1977, pp. 20, 339.). Many sources advise discarding difference scores of zero for this test; however, doing so inflates Type I error rates. Thus, we retained zeros as described by Marascuilo and McSweeney (p. 334) and Hays (1988, p. 829). If there were an odd number of zeros, one was discarded from analysis. Remaining zeros were ranked along with all other absolute differences and were then treated as any other tie. Finally, half of the zeros were assigned a positive sign, and the other half were assigned a negative sign. This formulation of the Wilcoxon matched-pairs signed-rank test provides the most conservative and accurate comparison test for this type of data.

Table 6
Encoding Strategy Usage Frequency Ratings in Experiment 3

Encoding strategy	Cued recall expectation		Free recall expectation		Cued vs. free	
	M (SD)	Mdn	M (SD)	Mdn	z	p
Cue–target association	3.67 (0.64)	4.0	1.58 (0.79)[a]	1.0[a]	7.65	**<.001**
Target–target association	1.78 (0.92)[a]	2.0[a]	2.76 (1.21)[a]	3.0[a]	4.81	**<.001**
Interitem association	1.65 (0.82)[b]	1.0[b]	1.99 (1.13)[c]	2.0[c]	1.79	.074
Target focus	2.43 (0.91)[d]	2.5[d]	3.63 (0.79)[a]	4.0[a]	6.61	**<.001**
Mental imagery	3.00 (1.10)	3.0	2.88 (1.18)	3.0	0.98	.328
Rote rehearsal	2.63 (1.12)	3.0	3.07 (1.09)	3.0	3.82	**<.001**
Verbalization	2.79 (1.24)	3.0	2.94 (1.26)	4.0	1.82	.069
Intraitem narrative	2.75 (1.13)	3.0	2.61 (1.25)[a]	3.0[a]	0.91	.362
Interitem narrative	1.98 (1.13)	1.5	2.62 (1.30)	3.0	3.61	**<.001**
Personal significance	2.67 (1.12)	3.0	2.45 (1.14)	2.0	1.64	.102
Observation	2.16 (1.08)[d]	2.0[d]	2.35 (1.13)[d]	2.0[d]	1.48	.140

Note. Rating scale was 1 (*no use*) to 4 (*extensive use*); $N = 84$; the Wilcoxon matched-pairs signed-rank test was used (see Footnote 4); statistically significant p-values are shown in boldface.
[a] $n = 83$.   [b] $n = 80$.   [c] $n = 81$.   [d] $n = 82$.

regressions represent the amount of increase (positive slopes) or decrease (negative slopes) in performance across lists. Next we computed Kendall's tau-b correlations between these slopes and the usage frequency ratings for each of the 11 strategies, separately for cued recall and free recall.[8] These correlations indicate the direction and magnitude of the relationship between self-reported use of a particular strategy and the amount that recall performance increased or decreased across lists. Thus, the correlations represent the efficacy of a given encoding strategy for a given test format.

Tables 7 and 8 show estimated tau-b correlation coefficients for cued recall and free recall for all 11 encoding strategies, with 95% confidence intervals for each individual coefficient and for their difference for each strategy. For five of the 11 strategies, the tau-b correlation coefficients significantly differed for cued versus free recall in one or both of the two experiments. Greater self-reported use of a cue–target association strategy was associated with increasing performance across cued recall lists but decreasing performance across free recall lists. Greater self-reported use of four strategies was associated with decreasing or unchanging performance across cued recall lists but increasing performance across free recall lists: target–target association, interitem association, target focus, and interitem narrative.

The preceding analyses of strategy efficacy should be interpreted with some caution because participants were not randomly assigned to use strategies to different extents. Nevertheless, the results from Experiments 2 and 3 are suggestive of which strategies were helpful for cued recall (cue–target association) versus free recall (target focus, and any association across pairs). Furthermore, these strategies appear to be beneficial for one test format and detrimental for the other. This is an important point with respect to methodological requirements for detecting qualitative changes and differences in encoding strategy, as addressed later in the General Discussion and in detail by Finley (2010).

Effectiveness of Metacognitive Control

Having considered results suggestive of which encoding strategies were more or less efficacious for cued versus free recall and

given that use of specific encoding strategies can mediate the effects of test expectancy on performance (cf. Murayama, 2005), we can now evaluate how effectively participants differentially applied encoding strategies to the two test formats. That is, we may assess how optimal their metacognitive control of encoding strategy was.

First, it is evident from Experiment 2 questionnaire data that participants' metacognitive control was not entirely optimal in the free recall condition: even after exposure to the demands of the task in the initial study–test cycle, these participants continued to employ unhelpful strategies to some extent, such as cue–target association. To be fair, it should be noted that participants were not explicitly told in this experiment that they would receive the same test format for each list. Also, the mere nature of the stimuli (i.e., word pairs) may have biased all participants toward associative encoding strategies from the start, and free-expecting participants did report using cue–target association less as the experiment progressed.

A summary of the differential efficacy and use of encoding strategies in all three experiments is shown in Table 9. Overall, participants' encoding strategy usages appear to be fairly well attuned to the different demands of the two test formats, with the

---

[8] Kendall's tau-b was used because the usage frequency rating data were ordinal, and there were many ties. Data from participants with missing values for any strategies were excluded entirely from analyses for Experiment 2 and were excluded on a testwise (i.e., per strategy) basis for Experiment 3. We calculated standard errors for tau-b using the formula provided by Woods (2007, square root of Equation 14) with the consistent variance estimates defined by Cliff & Charlin (1991). The standard error used for comparison of independent tau-b values (Experiment 2) was $\sqrt{SE^2_{\hat{\tau}_{b.1}} + SE^2_{\hat{\tau}_{b.2}}}$. The standard error used for comparison of dependent tau-b values (Experiment 3) was $\sqrt{SE^2_{\hat{\tau}_{b.1}} + SE^2_{\hat{\tau}_{b.2}} - 2cov(\hat{\tau}_{b.1}, \hat{\tau}_{b.2})}$. We calculated the covariance term using the formula provided by Cliff and Charlin (1991, Equation 20, corrected for the erroneously transposed first matrix), with the consistent variance estimates. Because these analyses were preplanned, an unadjusted alpha level was used.

Table 7

Correlations Between Self-Reported Strategy Use and Changes In Recall Performance Across Lists in Experiment 2

Encoding strategy	Cued recall		Free recall		Cued vs. free	
	Estimated $\hat{\tau}_b$ (SD)	95% CI	Estimated $\hat{\tau}_b$ (SD)	95% CI	SE	95% CI
Cue–target association	.28 (.11)	**[.06, .50]**	−.20 (.11)	[−.42, .01]	.16	**[.18, .79]**
Target–target association	−.03 (.10)	[−.23, .17]	.39 (.10)	**[.20, .57]**	.14	**[−.69, −.14]**
Interitem association	−.16 (.12)	[−.39, .08]	.23 (.11)	**[.02, .44]**	.16	**[−.70, −.07]**
Target focus	−.03 (.10)	[−.23, .16]	.51 (.08)	**[.35, .67]**	.13	**[−.79, −.29]**
Mental imagery	.25 (.09)	**[.07, .44]**	.04 (.12)	[−.19, .27]	.15	[−.08, .51]
Rote rehearsal	.02 (.12)	[−.21, .26]	.05 (.12)	[−.18, .28]	.17	[−.36, .30]
Verbalization	.10 (.12)	[−.14, .33]	−.05 (.12)	[−.28, .18]	.17	[−.18, .48]
Intraitem narrative	.20 (.10)	**[.002, .41]**	.23 (.12)	[−.01, .47]	.16	[−.34, .28]
Interitem narrative	.02 (.12)	[−.22, .25]	.37 (.10)	**[.17, .57]**	.16	**[−.66, −.05]**
Personal significance	.27 (.09)	**[.10, .45]**	.12 (.10)	[−.08, .33]	.14	[−.12, .42]
Observation	−.26 (.11)	**[−.47, −.05]**	−.20 (.12)	[−.45, .04]	.16	[−.38, .26]

Note. Correlations are estimated Kendall's tau-b; $n_{cued} = 46$, $n_{free} = 48$ (between-subjects). CI = confidence interval; SE = standard error of the difference between correlation coefficients for cued versus free recall; CIs were calculated with $z_{\alpha/2} = 1.96$, and SEs were calculated as per Woods (2007) using consistent variance estimates from Cliff and Charlin (1991). Statistically significant CIs are shown in boldface.

salient exceptions being failure to strategically use interitem association, and needless differential usage of rote rehearsal.

It is possible to quantify participants' metacognitive control effectiveness by calculating the Pearson correlation between the mean usage frequency rating for each strategy with the strategy efficacy measure for that strategy (tau-b, as described earlier), separately for cued recall and free recall. The resulting correlation coefficient represents the degree to which participants reported greater usage of strategies that were more beneficial for that test format. In Experiment 2, this measure was high for cued recall, $r_{cued} = .71$, $t(9) = 3.04$, $p = .014$, and low for free recall, $r_{free} = −.50$, $t(9) = −1.72$, $p = .119$, $z_{diff} = 2.88$, $p = .004$. The negative correlation for free-expecting participants indicates that they reported greater overall usage of encoding strategies that were less efficacious than other strategies at improving performance. However, this result may be largely driven by these participants' early use of cue–target association, before they knew what the test format would be like. This interpretation is supported by correlations conditionalized on participants' reporting greater usage of

such strategies in the first half of the experiment, $r_{free\_1} = −.55$, $t(9) = −1.98$, $p = .079$, versus the second half of the experiment, $r_{free\_2} = .01$, $t(9) = 0.02$, $p = .983$, $t_{diff}(8) = 1.42$, $p = .192$.

In Experiment 3, the correlation for cued recall was $r_{cued} = .27$, $t(9) = 0.83$, $p = .428$, and for free recall it was $r_{free} = .15$, $t(9) = 0.45$, $p = .665$. These correlations did not reliably differ, $z_{diff} = 0.22$, $p = .826$. Although these metacognitive control effectiveness correlations were lower in Experiment 3 than in Experiment 2, they did not in fact reliably differ across experiments for cued recall, $z_{diff} = 1.24$, $p = .216$, nor for free recall, $z_{diff} = 1.39$, $p = .165$. However, the difference in metacognitive control effectiveness correlations for cued versus free recall was marginally reliably lower in Experiment 3 versus Experiment 2, $z = 1.73$, $p = .083$. That is, there was more parity in metacognitive control effectiveness across test formats in Experiment 3 versus Experiment 2. This was likely due to the within-subjects design, which gave participants repeated experience with both test formats.

Taken together, these results suggest that participants came equipped with some degree of relevant metacognitive knowledge

Table 8

Correlations Between Self-Reported Strategy Use and Changes in Recall Performance Across Lists in Experiment 3

Encoding strategy	N	Cued recall		Free recall		Cued vs. free	
		Estimated $\hat{\tau}_b$ (SD)	95% CI	Estimated $\hat{\tau}_b$ (SD)	95% CI	SE	95% CI
Cue–target association	83	−.03 (.09)	[−.21, .15]	−.11 (.09)	[−.29, .07]	.13	[−.17, .33]
Target–target association	82	−.03 (.09)	[−.20, .14]	.22 (.08)	**[.06, .37]**	.12	**[−.49, −.01]**
Interitem association	80	−.12 (.09)	[−.30, .06]	.12 (.08)	[−.05, .28]	.12	[−.48, .01]
Target focus	81	.15 (.09)	[−.03, .33]	.14 (.09)	[−.03, .31]	.13	[−.24, .26]
Mental imagery	84	.03 (.09)	[−.14, .20]	−.001 (.09)	[−.18, .17]	.12	[−.21, .27]
Rote rehearsal	84	−.11 (.08)	[−.27, .05]	−.16 (.08)	**[−.31, −.001]**	.12	[−.19, .29]
Verbalization	84	−.07 (.09)	[−.25, .10]	−.19 (.08)	**[−.35, −.04]**	.13	[−.14, .38]
Intraitem narrative	83	−.06 (.08)	[−.22, .10]	.03 (.08)	[−.13, .20]	.13	[−.34, .16]
Interitem narrative	84	−.13 (.09)	[−.31, .04]	.21 (.09)	**[.04, .38]**	.13	**[−.59, −.09]**
Personal significance	84	.03 (.09)	[−.15, .21]	−.07 (.08)	[−.23, .08]	.12	[−.14, .35]
Observation	81	−.03 (.09)	[−.21, .15]	−.13 (.08)	[−.29, .03]	.13	[−.15, .34]

Note. Correlations are estimated Kendall's tau-b (within-subjects). CI = confidence interval; SE = standard error of the difference between correlation coefficients for cued versus free recall. CIs were calculated using $z_{\alpha/2} = 1.96$ and SEs were calculated as per Woods (2007) using consistent variance estimates from Cliff and Charlin (1991). Statistically significant CIs are shown in boldface.

Table 9
Differential Efficacy and Use of Encoding Strategies in Experiments 1–3

Encoding strategy	Experiment 1 Use	Experiment 2 Efficacy	Experiment 2 Use	Experiment 3 Efficacy	Experiment 3 Use
Cue–target association	C	C	C	—	C
Target–target association	~F	F	F	F	F
Interitem association		F	—	~F	—
Target focus	F	F	F	—	F
Mental imagery					
Rote rehearsal		—	F	—	F
Verbalization					
Intraitem narrative					
Interitem narrative		F	—	F	F
Personal significance					
Observation					

Note. C = reliably greater for cued versus free recall; F = reliably greater for free versus cued recall; ~F = marginally reliably greater for free versus cued recall; empty cell = no reliable difference; dash = no reliable difference when there was a corresponding reliable difference for efficacy or use.

of encoding strategies and were able to employ those strategies with some effectiveness, but that there was still room for improvement, especially for free recall. Further research could determine whether giving participants experience with both test formats, as was done in Experiment 3, may provide them with the opportunity to even further adaptively employ different encoding strategies (cf. Bjork, DeWinstanley, & Storm, 2007; DeWinstanley & Bjork, 2004).

Summary of All Results

The key results from Experiments 1–3 demonstrated that learners can adaptively and qualitatively accommodate their encoding strategies to the demands of an expected test format. Additionally, results from all three experiments provided insights into the characteristics and effectiveness of the encoding strategies that participants used. In studying for a cued recall test, participants relied heavily and consistently on the efficacious strategy of cue–target association; in studying for a free recall test, participants abandoned the inefficacious cue–target association strategy in favor of multiple efficacious strategies: selective attention to target words (i.e., target focus), making associations across word pairs (target–target association, interitem association, and interitem narrative), and rote rehearsal. Participants' metacognitive control of encoding strategies was mostly effective, though not without room for improvement, especially for free recall.

When Do Learners Use Qualitatively Different Encoding Strategies?

In what situations are learners likely to employ qualitatively different and appropriate encoding strategies, and do so in such a way that yields clear evidence? Using effective encoding strategies can be difficult, requiring multiple processes (Hertzog & Dunlosky, 2004). For example, learners must be sufficiently motivated toward a goal of high performance, they must be able to predict the

cognitive demands of a future test, and they must be equipped with a repertoire of relevant encoding strategies or be able to devise new strategies as needed. Thus, learners may often choose to modulate their study using other available forms of metacognitive control, such as item selection, study-time allocation, scheduling, and self-testing (cf. Finley et al., 2010).

In situations that limit other forms of metacognitive control, it must furthermore be the case that the demands of the two tests are different enough that learners cannot effectively use a common encoding strategy for both. We argued in the introduction that free recall and item recognition do not meet this requirement. Indeed, Hall et al. (1976) found that participants expecting either of these test formats self-reported predominant use of associative and imagery strategies and that use of these strategies was positively correlated with performance for both test formats. In contrast, the strategy efficacy analyses in the current study (Experiments 2 and 3) suggest that different encoding strategies were beneficial for cued recall (e.g., cue–target association) versus free recall (e.g., target focus, association across pairs) and furthermore suggest that these strategies were detrimental for the alternative test format. The few other studies in which the key disordinal interaction has been found also primarily used dissimilar test formats: serial recall versus item recognition (von Wright, 1977; von Wright & Meretoja, 1975) and anticipation recall versus item recognition (Postman & Jenkins, 1948).

Even in situations with contrasting task demands, capable learners are only likely to employ qualitatively different encoding strategies if they are given sufficient exposure to and practice with the learning material and if the material is sufficiently complex to afford a variety of strategies. Situations that provide learners with experience over multiple study-test cycles (such as the current study) are more effective at inducing test expectancy than ones that use instructions alone (cf. Lundeberg & Fox, 1991; McDaniel et al., 1994). Furthermore, in the only study to have ever found a disordinal interaction for free recall versus item recognition, Postman and Jenkins (1948) used multiple presentations of stimuli during study, a rare procedure also used by von Wright (1977) and von Wright and Meretoja (1975). Finally, the latter two studies also used line drawings as stimuli, which (like the word pairs used in the current study) would seem to enable more variety in encoding strategies than single word stimuli, thus increasing the possibility that learners will come to use qualitatively different strategies.[9]

Conclusion

Learners can regulate their study experience to enhance learning in a variety of ways (Benjamin, 2007; Dunlosky et al., 2007; Finley et al., 2010; Serra & Metcalfe, 2009), and the effectiveness of their metacognitive control is critical to self-regulated learning.

---

[9] Test-expectancy effects have been found less consistently with prose (i.e., text passages) than with discrete materials such as word lists (cf. d'Ydewalle et al., 1983; McDaniel et al., 1994; Oakhill & Davies, 1991). Although prose enables more diverse encoding strategies, it also enables item selection and study-time allocation of subsets of the text, thus complicating the isolation of encoding strategy effects. McDaniel et al. sought to ameliorate this problem using a moving window method that restricted the amount of text presented at a given time.

Prior research has shown that learners will strategically exercise metacognitive control in response to task properties such as stimulus characteristics (e.g., perceived difficulty; Nelson & Leonesio 1988), time pressure and target performance goals (Son & Metcalfe, 2000; Thiede & Dunlosky, 1999), and expectations of the difficulty of an upcoming test (Thiede, 1996). The present research adds a new component to such findings by demonstrating that qualitative aspects of a test (test format) *can* drive qualitative differences in metacognitive control (encoding strategy). Whether students *will* in fact employ appropriate and qualitatively different encoding strategies and whether doing so would actually benefit them depend on the factors discussed earlier. In educational contexts, students may rely more on other forms of metacognitive control (e.g., item selection, study-time allocation, and scheduling), and various test formats may or may not differ enough for certain encoding strategies to differentially affect performance. The interaction between specific encoding strategies and educationally realistic test formats (e.g., multiple choice vs. essay tests) remains a topic for further investigation. To the extent that test formats do differ in their cognitive demands, it indeed behooves students facing an upcoming test to know the test format, because that information can enable them to tailor their encoding strategies to the demands of the test; without appropriate encoding strategies, their study efforts may be wasted.

In summary, this study used the test-expectancy paradigm to investigate adaptive and qualitative changes in encoding strategy in response to experiencing the demands of an upcoming test format. Recall, recognition, and self-report results demonstrated learners' abilities to adaptively and qualitatively modify their encoding strategies (Experiment 1), metacognitive monitoring (Experiment 2), and study-time allocation (Experiment 3) on the basis of the test format they expected (cued recall vs. free recall). In short, learners showed that they can work smarter, not just harder.

References

Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition, 2,* 406–412. doi:10.3758/BF03196896

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 576–587. doi:10.1037/0278-7393.6.5.576

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39,* 445–459. doi:10.3758/BF03193014

Benjamin, A. S. (2007). Memory is more than just remembering: Strategic control of encoding, accessing memory, and making decisions. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Vol. 48. Skill and strategy in memory use* (pp. 175–223). London, England: Academic Press. doi:10.1016/S0079-7421(07)48005-7

Bjork, E. L., DeWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: Can experiencing the outcome of different encoding strategies enhance subsequent encoding? *Psychonomic Bulletin & Review, 14,* 207–211. doi:10.3758/BF03194053

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 657–668. doi:10.1037/0278-7393.15.4.657

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Carey, S. T., & Lockhart, R. S. (1973). Encoding differences in recognition and recall. *Memory & Cognition, 1,* 297–300. doi:10.3758/BF03198112

Cliff, N., & Charlin, V. (1991). Variances and covariances of Kendall's tau and their estimation. *Multivariate Behavioral Research, 26,* 693–707. doi:10.1207/s15327906mbr2604_6

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33,* 497–505. doi:10.1080/14640748108400805

Connor, J. M. (1977). Effects of organization and expectancy on recall and recognition. *Memory & Cognition, 5,* 315–318. doi:10.3758/BF03197576

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684. doi:10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. doi:10.1037/0096-3445.104.3.268

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32,* 945–955. doi:10.3758/BF03196872

Diaz, M., & Benjamin, A. S. (2011). The effects of proactive interference (PI) and release from PI on judgments of learning. *Memory & Cognition, 39,* 196–203. doi:10.3758/s13421-010-0010-y

Dunlosky, J., Serra, M., & Baker, J. M. C. (2007). Metamemory. In F, Durso, R. Nickerson, S. Dumais, S. Lewandowsky, & T. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 137–159). New York, NY: Wiley.

d'Ydewalle, G. (1981). Test-expectancy effects in free recall and recognition. *Journal of General Psychology, 105,* 173–195. doi:10.1080/00221309.1981.9921071

d'Ydewalle, G. (1982). Psychophysical methods to unravel test-expectancy effects. *Studia Psychologica, 24,* 177–191.

d'Ydewalle, G., Swerts, A., & de Corte, E. (1983). Study time and test performance as a function of test expectations. *Contemporary Educational Psychology, 8,* 55–67. doi:10.1016/0361-476X(83)90034-6

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology, 68,* 58–63. doi:10.1037/h0044655

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 588–598. doi:10.1037/0278-7393.6.5.588

Finley, J. R. (2010). *Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test expectancy method* (Unpublished master's thesis). University of Illinois at Urbana-Champaign. Retrieved from http://www.ideals.illinois.edu

Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2010). Metacognitive control of learning and remembering. In M. S. Khine & I. Saleh (Eds.), *New science of learning: Cognition, computers and collaboration in education.* New York, NY: Springer.

Fisher, R. P., & Craik, F. I. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory, 3*(6), 701–711. doi:10.1037/0278-7393.3.6.701

Foos, P. W., & Clark, M. C. (1983). Learning from text: Effects of input order and expected test. *Human Learning, 2,* 177–185.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–216). London, England: Wiley.

Hall, J. W., Grossman, L. R., & Elwood, K. D. (1976). Differences in encoding for free recall vs. recognition. *Memory & Cognition, 4,* 507–513. doi:10.3758/BF03213211

Hays, W. L. (1988). *Statistics* (4th ed.). Fort Worth, TX: Holt, Rinehart, and Winston.

Hertzog, C., & Dunlosky, J. (2004). Aging, metacognition, and cognitive

control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 45. Advances in research and theory* (pp. 215–251). San Diego, CA: Academic Press.

Hertzog, C., & Dunlosky, J. (2006). *Using visual imagery as a mnemonic for verbal associative learning: Developmental and individual differences.* Amsterdam, the Netherlands: Benjamins.

Hertzog, C., Price, J., & Dunlosky, J. (2008). How is knowledge generated about memory encoding strategy effectiveness? *Learning and Individual Differences, 18,* 430–445. doi:10.1016/j.lindif.2007.12.002

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24,* 202–216. doi:10.3758/BF03200881

Jacoby, L. L. (1973). Test appropriate strategies in retention of categorized lists. *Journal of Verbal Learning and Verbal Behavior, 12,* 675–682. doi:10.1016/S0022-5371(73)80047-7

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508. doi:10.1016/S0022-5371(83)90301-8

Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of notetaking and test-expectancy on the learning of text material. *Journal of Educational Research, 68,* 363–365.

Leonard, J. M., & Whitten, W. B., II. (1983). Information stored when expecting recall or recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 440–455. doi:10.1037/0278-7393.9.3.440

Lewis, K., & Wilding, J. M. (1981). Influences of test expectations on memory-processing strategies. *Current Psychology, 1,* 61–74. doi:10.1007/BF02684427

Lovelace, E. A. (1973). *Effects of anticipated form of testing on learning* (Final report, Project 2-C-019, Grant OEG-3–72-0033). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203–208. doi:10.3758/BF03204766

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test-expectancy generalize to classroom outcomes? *Review of Educational Research, 61,* 94–106.

Maisto, S. A., DeWaard, R. J., & Miller, M. E. (1977). Encoding processes for recall and recognition: The effect of instructions and auxiliary task performance. *Bulletin of the Psychonomic Society, 9,* 127–130.

Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences.* Monterey, CA: Brooks/Cole.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ: Erlbaum.

May, R. B., & Sande, G. N. (1982). Encoding expectancies and word frequency in recall and recognition. *American Journal of Psychology, 95,* 485–495. doi:10.2307/1422137

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary Educational Psychology, 19,* 230–248. doi:10.1006/ceps.1994.1019

Meyer, G. (1934). An experimental study of the old and new types of examination: I. the effect of the examination set on memory. *Journal of Educational Psychology, 25,* 641–661. doi:10.1037/h0073102

Meyer, G. (1936). The effect of recall and recognition on the examination set in classroom situations. *Journal of Educational Psychology, 27,* 81–99. doi:10.1037/h0059333

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. doi:10.1016/S0022-5371(77)80016-9

Murayama, K. (2005). Exploring the mechanism of test-expectancy effects on strategy change. *Japanese Journal of Educational Psychology, 53,* 172–184.

Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & Cognition, 9,* 283–300. doi:10.3758/BF03196962

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://www.usf.edu/FreeAssociation/

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. doi:10.1037/0278-7393.14.4.676

Oakhill, J., & Davies, A. (1991). The effects of test-expectancy on quality of note taking and recall of text at different times of day. *British Journal of Psychology, 82,* 179–189. doi:10.1111/j.2044-8295.1991.tb02392.x

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241–286. doi:10.1006/ceps.2000.1040

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9,* 105–119.

Postman, L. (1964). Studies of learning to learn: II. Changes in transfer as a function of practice. *Journal of Verbal Learning and Verbal Behavior, 3*(5), 437–447. doi:10.1016/S0022-5371(64)80014-1

Postman, L. (1969). Experimental analysis of learning to learn. In J. T. Spence & G. H. Bower (Eds.), *The psychology of learning and motivation: Vol. 3. Advances in research and theory.* New York, NY: Academic.

Postman, L., & Jenkins, W. O. (1948). An experimental analysis of set in rote learning: The interaction of learning instruction and retention performance. *Journal of Experimental Psychology, 38,* 683–689. doi:10.1037/h0057311

Roediger, H. L., III. (1980). The effectiveness of four mnemonics in ordering recall. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 558–567. doi:10.1037/0278-7393.6.5.558

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. doi:10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3–39). Hillsdale, NJ: Erlbaum.

Sanders, N. M., & Tzeng, O. (1975). Type-of-test-expectancy effects on learning of word lists and prose passages. *Acta Psychologica Taiwanica, 17,* 1–11.

Schmidt, S. R. (1988). Test-expectancy and individual-item versus relational processing. *American Journal of Psychology, 101,* 59–71. doi:10.2307/1422793

Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition and education* (pp. 278–298). New York, NY: Routledge.

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. doi:10.1037/0278-7393.26.1.204

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience, 26,* 9162–9172. doi:10.1523/JNEUROSCI.2877-06.2006

Szpunar, K. K., McDermott, K. D., & Roediger, H. L., III. (2007). Expectation of a final cumulative test enhances long-term retention. *Memory & Cognition, 35,* 1007–1013. doi:10.3758/BF03193473

Terry, P. W. (1933). How students review for objective and essay tests. *The Elementary School Journal, 33,* 592–603. doi:10.1086/456929

Terry, P. W. (1934). How students study for three types of objective tests. *Journal of Educational Research, 27,* 333–343.

Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49,* 901–918.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024–1037. doi:10.1037/0278-7393.25.4.1024

Thiede, K. W., Wiley, J., & Griffin, T. D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology, 81,* 264–273. doi:10.1348/135910710X510494

Thomas, A. K., & McDaniel, M. A. (2007a). Metacomprehension for educationally relevant materials: Dramatic effects of encoding–retrieval interactions. *Psychonomic Bulletin & Review, 14,* 212–218. doi:10.3758/BF03194054

Thomas, A. K., & McDaniel, M. A. (2007b). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition, 35,* 668–678. doi:10.3758/BF03193305

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology, 86,* 255–262. doi:10.1037/h0029997

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64,* 109–118. doi:10.1016/j.jml.2010.11.002

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 5,* 193–197. doi:10.1016/S0022-5371(66)80016-6

Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology, 5,* 275–287. doi:10.1016/0010-0285(73)90037-6

Underwood, B. J. (1963). *Stimulus selection in verbal learning.* New York, NY: McGraw-Hill.

von Wright, J. (1977). On the development of encoding in anticipation of various tests of retention. *Scandinavian Journal of Psychology, 18,* 116–120. doi:10.1111/j.1467-9450.1977.tb00264.x

von Wright, J., & Meretoja, M. (1975). Encoding in anticipation of various tests of retention. *Scandinavian Journal of Psychology, 16,* 108–112. doi:10.1111/j.1467-9450.1975.tb00170.x

Watanabe, H. (2003). Effects of encoding style, expectation of retrieval mode, and retrieval style on memory for action phrases. *Perceptual and Motor Skills, 96,* 707–727.

Whitten, W. B., II. (2011). Learning from and for tests. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork* (pp. 217–234). New York, NY: Psychology Press.

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1024–1039. doi:10.1037/0278-7393.19.5.1024

Woods, C. M. (2007). Confidence intervals for gamma-family measures of ordinal association. *Psychological Methods, 12,* 185–204. doi:10.1037/1082-989X.12.2.185

Appendix A

Encoding Strategy Categories Identified in Experiment 1

Encoding strategy	Characteristic response
Cue–target association	I tried to find some connection between the two words that were paired.
Target–target association	I started associating the second word from each pair together.
Unspecified association	I just tried to associate the words.
Target focus	Toward the end, I just started memorizing the last word and not really paying attention to the first word.
Mental imagery	I tried to visualize a picture for each of the words.
Rote rehearsal	I attempted to repeat the words over in my head.
Verbalization	I was trying to just say the words out loud to remember them.
Narrative	I tried to remember the words based on events and a story that I would make up.
Personal significance	I tried to match the words with something or someone I know.
Bizarre	I always try to remember the words in completely outlandish situations.
Action	I tried to act out both words.
Phonetic	I also tried to remember words that began with the same letter.

Appendix B

Encoding Strategies Listed in the Questionnaire in Experiments 2 and 3

Strategy label	Full text used in questionnaire
Cue–target association	Made associations between the left-hand word and right-hand word in a pair.
Target–target association	Made associations between the right-hand words across multiple pairs.
Interitem association	Made associations between multiple pairs across a list.
Target focus	Focused more on the right-hand words.
Mental imagery	Used mental imagery (formed a picture in your head).
Rote rehearsal	Repeated individual words or pairs over and over.
Verbalization	Spoke words out loud or under your breath.
Intraitem narrative	Used a single pair or word in a sentence, phrase, or story.
Interitem narrative	Used groups of pairs or words across a list in a sentence, phrase, or story.
Personal significance	Related words to something personally significant.
Observation	Just read or looked at the words.

Note. Adapted from Hall, Grossman, and Elwood (1976) and Leonard and Whitten (1983). Strategy labels are for reference and were not used in the questionnaire.