ADAPTIVE AND QUALITATIVE CHANGES IN ENCODING STRATEGY WITH EXPERIENCE: EVIDENCE FROM THE TEST EXPECTANCY METHOD

BY

JASON R. FINLEY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Advisor:

Associate Professor Aaron S. Benjamin

Abstract

Three experiments demonstrated undergraduate participants' abilities to adaptively and qualitatively accommodate their encoding strategies to the demands of an upcoming test as they gained experience with the test format. Stimuli were lists of word pairs. Experiment 1 induced test expectancy of either cued recall (of targets given cues) or free recall (of targets only) across four study-test cycles of the same test format, then presented participants with 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 pattern 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. The specific ways in which strategies shifted were revealed by final associative and item recognition performance and by self-report data. Participants also came to appropriately modulate metacognitive monitoring (Experiment 2) and study-time allocation (Experiment 3) across study-test cycles. Encoding strategies used for cued versus free recall were characterized and evaluated, and an account was given to reconcile inconsistent prior findings from test expectancy studies.

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

Acknowledgments

This research was supported by funding from the National Institute of Health to ASB (R01 AG026263). I give my thanks to Aaron S. Benjamin for unflappable and indefatigable guidance, to Brian H. Ross for insightful and apropos feedback, and to Laurel M. Methot for love and support.

Table of Contents

Introduction

Effective studying requires the ability to tailor one's study behaviors to the foreseeable requirements of the test. The present research examined the extent to which learners are able to make qualitative and adaptive changes in the way they learn material after experiencing the demands of an upcoming test. Such learning to learn 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 the control processes of: item selection, study-time allocation, scheduling, and encoding strategy (cf. Benjamin, 2008; Dunlosky, Serra, & Baker, 2007; Finley, Tullis, & Benjamin, 2010; Serra & Metcalfe, 2009). The current study focused specifically on how learners change their encoding strategies for learning words based on how they expect their memory for those words to be queried.

Encoding Strategy

Encoding strategy refers to the nature of the processing applied to information that a learner wants to remember. The way in which learners encode information is critical to how that information is stored in memory (Craik & Lockhart, 1972; Fisher & Craik, 1977). This is an idea that 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."

Normative efficacy of encoding strategies. Many studies have investigated the normative efficacy of various encoding strategies by attempting to control learners' strategies via direct instructions, orienting tasks, or materials that are more or less conducive to certain strategies. A rote rehearsal strategy (i.e., overtly or covertly repeating information to oneself) is often used as a baseline comparison for the effectiveness of more elaborative strategies (e.g., generating associations and/or imagery), with the latter almost always producing superior memory performance. Craik and Lockhart (1972) demonstrated that semantic ("deep") encoding of words, such as judging whether each word fit into a category, led to superior subsequent memory compared to more "shallow" encoding, such as making judgments about a word's font. Organizing words into subjectively meaningful groups has been demonstrated as an effective strategy for free recall (Tulving, 1966). Visual imagery has been shown to be effective for encoding paired associates (Hertzog & Dunlosky, 2006), and may be executed in a variety of ways (e.g., forming separate images for a cue and target versus forming a composite or interactive image). Finally, a panoply of mnemonics have been espoused for ages; they vary in their complexity (from acronyms and acrostics to the method of loci and the peg word method), and vary in their effectiveness depending on task demands (Roediger, 1980).

Many of these results can be explained by the concept of transfer-appropriate processing (Morris, Bransford, & Franks, 1977), which holds that effective encoding strategies are those that employ cognitive processes at the time of acquisition that are most similar to those processes used at the time of retrieval. Strong support for this general theoretical claim was provided by experimental results that demonstrated that "weaker" forms of encoding could actually lead to superior memory if the test queried the same aspects of memory as those normatively poorer encoding strategies (Blaxton, 1983; Jacoby, 1983; Roediger, Weldon, & Challis, 1989).

Control of encoding strategy. Thus, much is known about the effectiveness of different encoding strategies under various conditions and with various materials, but much less is known about how learners employ encoding strategies when left to their own devices, and whether they can adaptively adjust their strategies to meet the demands of a future task. That is, we know little about learners' metacognitive control of encoding strategies. In fact, Lundeberg and Fox (1991), in assessment of their meta-analysis on test expectancy studies, 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."

There are two basic types of adjustments that learners can make to their encoding strategies: quantitative and qualitative. A learner may apply the same encoding strategy (e.g., rote rehearsal) to varying degrees based on the anticipated difficulty of an upcoming test—a quantitative change, which could result purely from 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 I review below, 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 method provides one way to study whether and how effectively learners use different encoding strategies for different tasks. In this methodology, participants are led to expect a particular test format (e.g., free recall vs. recognition), 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 their 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 (aka "mental set") manipulation reflect differences in the encoding strategies employed by participants during study. Thus, such data allow us to infer whether participants tailor their encoding strategies to the demands of a specific expected test format.

The most prominent finding from studies using this method is that expectation of free recall appears to facilitate performance for both free recall and recognition tests. More specifically: a number of studies have shown that participants anticipating a free recall test achieve higher performance on tests of both free recall and recognition than do participants anticipating a recognition test (Balota & Neely, 1980; d'Ydewalle, Swerts, & de Corte, 1983; Hall, Grossman, & Elwood, 1976; Maisto, DeWaard, & Miller, 1977; Meyer 1934; Neely & Balota, 1981; Schmidt, 1988; Thiede, 1996).

These findings provide ample evidence that learners can make judicious quantitative adjustments to their encoding strategies based on anticipated test format. Yet none of these findings can be 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 (aka crossover) interaction, such that, for both final test formats, learners who expected that format outperform those who expected the different format. 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; Schmidt, 1988). These data are curiously inconsistent with students' self-reports that they consider different study methods as best suited for different test formats, such focusing on details and underlining key terms when preparing for a fill-in-the-blank or true-false test organizing main points when preparing for an essay test (Terry, 1933, 1934).

There have been only three test expectancy studies, largely overlooked in the literature, that have shown a disordinal interaction of expected test format and received test format that may be attributed to differences in encoding strategies. Von Wright and Meretoja (1975) and von Wright (1977) showed such an interaction with serial recall versus recognition. Postman and Jenkins (1948) showed such an interaction with anticipation recall (similar to serial recall) versus recognition, and with free recall versus recognition. These results, discussed further in the General Discussion, are the exceptions.

Some researchers (e.g., Von Wright & Meretoja, 1975; Kulhavy, Dyer, & Silver, 1975; Oakhill & Davies, 1991) 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 intra-category serial position functions (Carey & Lockhart, 1973; but cf. Hall et al., 1976 for a failure to replicate), overall 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). There is even some tentative evidence of different encoding strategies for recognition versus recall from functional neuroimaging (Staresina & Davachi, 2006).

In summary, the majority of experiments from the test-expectancy literature have revealed evidence for only a quantitative difference in encoding strategy between test conditions. There is, however, some evidence that learners sometimes employ qualitatively different strategies that either do not result in differences in overall performance or that do so only for certain test formats, as reviewed further in the General Discussion.

Current Study

The current study was designed to evaluate learners' abilities to adaptively and qualitatively modify their encoding strategies. In Experiment 1 I employed the test expectancy method using the test formats of cued recall versus free recall, in search of the elusive interaction between expected and received test format indicative of qualitative differences in encoding strategy. In Experiment 2 I 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 I sought to train learners to better exercise strategic metacognitive control by providing them experience with both test formats and allowing them control over study-time allocation.

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. By using two test formats that required production of the same information under qualitatively different task demands, I 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. Using multiple study-test cycles allowed us to observe the development of differential strategy use across experience with the test formats. 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 female) participated for partial fulfillment of course requirements. Data were not recorded for two additional participants due to computer error.

Design. The experiment used a 2 x 2 x 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 [high vs. low]). In addition, the target (right-hand) words of the pairs were counterbalanced within-subjects such that half were high frequency (M_{KF} = 51.9, SD_{KF} = 18.9; Kucera & Francis,

1967) and half were low frequency (M_{KF} = 17.3, SD_{KF} = 5.1). 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 4-8 letter nouns 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).

The word pairs had a mean forward associative strength of .023 (SD = .005), as obtained from the University of South Florida Word Association, Rhyme and Word Fragment Norms (Nelson, McEvoy, & Schreiber, 1998). For each participant, half of the word pairs were randomly selected to remain intact (high associative strength, e.g., *flight-bird*), and the other half were transformed into low associative strength pairs (e.g., *trumpet-planet*) by randomly shuffling the cue words among these pairs such that no target word retained its original cue, and the forward associative strength for all of these pairs was zero. 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 run 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 using 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 (C-C), expected cued recall and received free recall (C-F), expected free recall and received cued recall (F-C), and expected free recall and received free recall (F-F). 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 that they would be studying a series of word pairs that they would later be tested on. 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 inter-stimulus 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 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.

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. 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 which critically 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 described above.

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 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 high and low associative strength) 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 paired with cues from other pairs in the same list). There were no words that had not previously been presented, and cue and target words always appeared on the same side of a pair as previously presented. 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. An equal number of cue words and target words were randomly selected from all five previously studied lists and from word pairs of both high and low associative strengths. 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 calculated using the pooled standard deviation of the groups being compared (Olejnik & Algina, 2000, Box 1 Option B). Effect sizes for ANOVAs are reported as $\hat{\omega}^2_{partial}$ calculated using the formulae provided by Maxwell and Delaney (2004). 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. For comparisons of means with large differences in sample sizes, the Welch-Satterthwaite estimation of degrees of freedom was used.

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 to make is 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 2-way between-subjects ANOVA revealed a reliable disordinal interaction between expected final test format and received final test format, $F(1,96) = 40.28$, $MSE = .035$, $\hat{\omega}^2_{partial} = .28$, $p < .001$, such that on a final cued recall test participants who had expected cued recall ($M = .51$, $SD = .26$) outperformed participants who had expected free recall ($M = .25$, $SD = .19$), $t(48) = 3.90$, $p < .001$, $d = 1.13$, and on a final free recall test, participants who had expected free recall ($M = .27$, $SD = .16$) outperformed participants who had expected cued recall ($M = .06$, $SD = .05$), $t(48) = 6.32$, $p < .001$, $d = 1.83$.

Recall across tests 1-4. Figure 2 shows mean performance across recall tests 1-4 for cued recall versus free recall. Means and standard deviations are presented in Table 1. 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). This effect was confirmed by separate simple linear regressions predicting performance from list number for each participant receiving free recall, $M_b = 0.019$, $SD_b = 0.043$, $t(49) = 3.18$, $p = .003$. Because this improvement was in the face of considerable proactive interference, which often leads to decreases in memory performance across lists (Wickens, Born, & Allen, 1963), it suggests that these subjects 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 3 and Table 2 show mean performance as a function of list number (1-4), test format (cued vs. free), and associative strength (high vs. low). A 3-way mixed ANOVA revealed a reliable 2-way interaction between test format and associative strength, $F(1, 98) = 89.92$, $MSE = .019$, $p < .001$, $\hat{\omega}^2_{partial} = .079$, such that performance was superior for high versus low associative strength word pairs to a much greater degree for cued recall (F(1,

49) = 162.10, MSE = .027, p < .001, $\hat{\omega}^2_{partial}$ = .204) than for free recall (F(1, 49) = 5.62, MSE = .011, p = .022, $\hat{\omega}^2_{partial}$ = .018). There was no reliable 3-way interaction, F(3, 294) = 1.94, MSE = .011, p = .123, $\hat{\omega}^2_{partial}$ < .001, and list number did not interact with associative strength, F(3, 294) = 1.17, MSE = .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.

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 between-subjects ANOVA revealed that this value did not reliably differ across conditions, F(3,96) = 0.68, MSE = 5187.66, p = .568, $\hat{\omega}^2$ < .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 3 shows the frequencies of each strategy for both expectancy conditions. The proportion of participants reporting each strategy was compared 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 (27/50 vs. 9/50, z = 3.75, p < .001). For participants expecting free recall, the most frequently reported strategy was selectively attending to the target words (e.g., "…*towards 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 (35/50 vs. 9/50, $z = 6.59$, $p < .001$). One other strategy approached significance (7/50 vs. 0/50, $z = 2.74$, $p = .006$) 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.

Table 4 shows the frequency data for four common ways in which participants reported that they would have changed their encoding strategies to better prepare for the final test. 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: "*I didnt 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, and/or that they would have made more cue-target associations. An illustrative example response from a participant who expected free recall but received cued recall 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 didnt 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.

Associative recognition. Evidence of the encoding strategies reported by participants is provided by the results of the recognition tests. To best elucidate any differences and changes in encoding strategies induced by receiving different test formats, I analyzed only recognition data from participants who received their expected test format on the final list (i.e., conditions C-C and F-F). Due to computer error, recognition data were not recorded for seven of these participants; thus, $N = 43$ for associative and item recognition analyses ($n_{cued} = 21$, $n_{free} = 22$).

Associative recognition performance by cued-expecting participants ($M_{d'} = 2.18$, $SD_{d'} = 0.84$) was reliably greater than that by free-expecting participants ($M_{d'} = 1.15$, $SD_{d'} = 0.78$), $t(41) = 4.07$, p < .001, d = 1.27. This is consistent with the cued-expecting participants' greater reports of using a cue-target association strategy; because these participants made more efforts to associate cue and target words during encoding, they were better able to recognize the correctly associated pairs.

Figure 4 and Table 5 show associative recognition performance as a function of test expectancy (cued vs. free) and the list number from which the word pairs originated (1-5). Separate simple linear regressions for each participant revealed that performance by free-expecting participants reliably declined across lists of origin, $M_b = -0.20$, $SD_b = 0.28$, $t(21) = -3.28$, $p = .004$, while performance by cued-expecting participants did not reliably change across lists, $M_b = 0.03$, $SD_b = 0.23$, $t(20) = 0.61$, $p = .547$. These results are consistent with the free-expecting participants' greater reports of changing their encoding strategies across lists to ones in which less attention was paid to the connection between cues and targets.

Item recognition. Figure 5 shows item recognition performance as a function of test expectancy (cued vs. free) and item type (cue vs. target). A 2-way mixed ANOVA revealed a reliable disordinal interaction between test expectancy and item type, $F(1,41) = 70.43$, $MSE = .046$, $p < .001$, $\hat{\omega}^2_{partial} = .058$. Cue word recognition performance was greater for cued-expecting participants ($M_{d'} = 2.28$, $SD_{d'} = 1.02$) than for free-expecting participants ($M_{d'} = 0.93$, $SD_{d'} = 0.55$), $t(41) = 5.23$, $p < .001$, $d = 1.66$. Similarly, target word recognition performance was greater for cued-expecting participants ($M_{d'} = 1.76$, $SD_{d'} = 0.86$) than for free-expecting participants ($M_{d'} = 1.18$, $SD_{d'} = 0.52$), $t(41) = 2.61$, $p = .013$, $d = 0.82$. For cued-expecting participants, recognition performance was greater for cue words than for target words, $t(20) = 7.19$, $p < .001$, $d = 0.11$, but for free-expecting participants the opposite was true, $t(21) = -4.34$, $p < .001$, $d = -0.10$.

Cued-expecting participants had seen the cue words twice as many times as the target words (once during presentations and once during the recall tests), and twice as many times as did the free-expecting participants, so their superior performance on these items 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. This is consistent with the free-expecting participants' greater reports of selectively attending to the target words; because they paid less attention to cue words than target words, they were less able to recognize these.

Figure 6 and Table 6 show item recognition performance as a function of test expectancy (cued vs. free), item type (cues vs. targets), and the list number from which the words originated (1-5). Hit rates were used for this analysis because d' could not be computed by list of origin, due to lure words having originated from no previous list by definition. A 3-way mixed ANOVA revealed a reliable 3-way interaction, $F(4,164) = 3.50$, $MSE = .026$, $p = .009$, $\hat{\omega}^2_{partial} = .022$, such that item type and list number did not interact for cued-expecting participants, $F(4,80) = 0.14$, $MSE = .018$, $p = .968$, $\hat{\omega}^2_{partial} < .001$, but did interact for free-expecting participants, $F(4,84) = 5.95$, $MSE = .032$, $p < .001$, $\hat{\omega}^2_{partial} = .085$, such that for these participants there was a reliable negative linear trend across lists for cues, $F(1,21) = 19.51$, $MSE = .036$, $p < .001$, $\hat{\omega}^2_{partial} = .184$, but no reliable linear trend across lists for targets, $F(1,21) = 2.16$, $MSE = .038$, $p = .157$, $\hat{\omega}^2_{partial} = .014$. For cued-expecting participants, list number affected neither hit rate for cues, $F(4,80) = 0.67$, $MSE = .014$, $p = .618$, $\hat{\omega}^2_{partial} < .001$, nor hit rate for targets, $F(4,80) = 0.46$, $MSE = .030$, $p = .763$, $\hat{\omega}^2_{partial} < .001$. Thus, across lists, free-expecting participants showed a steady decline in recognition of cues but not targets, consistent with these participants paying less attention to the cue words as they gained experience with a task for which cues were not important. Cued-expecting participants consistently paid attention to both cue and target words, as both words were important for the task of cued recall.

Summary of results. Taken together, the above results suggest that participants indeed came to strategically employ qualitatively different encoding strategies that were appropriate to the expected test format. 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 a target focus strategy (cf. Underwood, 1963).

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. Hertzog and Dunlosky, 2004). Given the effective differences and changes in encoding strategy observed in Experiment 1, it should also be possible to observe adaptive changes in metacognitive monitoring, as measured by judgments of learning (JOLs). Thus, I 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 irrelevant). To test this prediction I 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 female) participated for partial fulfillment of course requirements.

Design. The experiment used a 2 x 2 mixed design with one between-subjects variable (expected final test format [cued recall vs. free recall]) and one within-subjects variable (word pair associative strength [high vs. low]). In addition, the target (right-hand) words of the pairs were counterbalanced within-subjects such that half were high frequency (M_{KF} = 232.0, SD_{KF} = 157.3) and half were low frequency (M_{KF} = 3.9, SD_{KF} = 2.6). 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), divided into four lists of 32 pairs for each participant. As in Experiment 1, all words were 4-8 letter nouns, 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). Mean forward associative strength of word pairs was .025 (SD = .005). For each participant, associative strength was manipulated and pairs were placed into lists as described in Experiment 1.

Procedure. The procedure was similar to that of Experiment 1, with the major changes being the omission of the final critical 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 inter-stimulus interval.

Questionnaire on encoding strategy. An encoding strategy questionnaire was devised based on the self-report data from Experiment 1 and based 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 1st half, or the 2nd half, or stayed suspicious the whole time. There was no time limit for 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. Again, there was no time limit and no feedback was given.

Results and Discussion

Recall performance. Figure 7 shows mean performance across recall tests 1-4 for cued recall versus free recall. Means and standard deviations are presented in Table 1. 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 performance, 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 8 and Table 2 show mean performance as a function of list number (1-4), test format (cued vs. free), and associative strength (high vs. low). A 3-way mixed ANOVA revealed a reliable 2-way interaction between test format and associative strength, $F(1, 101) = 104.76$, $MSE = .026$, $p < .001$, $\hat{\omega}^2_{partial} = .125$, such that performance was superior for high versus low associative strength word pairs to a much greater degree for cued recall ($F(1, 52) = 181.12$, $MSE = .044$, $p < .001$, $\hat{\omega}^2_{partial} = .347$) than for free recall ($F(1, 49) = 31.20$, $MSE = .006$, $p < .001$, $\hat{\omega}^2_{partial} = .048$). There was no reliable 3-way interaction, $F(3, 303) = 1.22$, $MSE = .010$, $p = .301$, $\hat{\omega}^2_{partial} < .001$, and list number did not interact with associative strength, $F(3, 303) = 1.91$, $MSE = .010$, $p = .127$, $\hat{\omega}^2_{partial} = .002$. Thus, as in Experiment 1, across all lists, associative strength was a very important variable for cued recall but not for free recall.

Metacognitive monitoring. Figure 9 and Table 7 show mean JOLs as a function of list number (1-4), test format (cued vs. free), and associative strength (high vs. low). A 3-way mixed ANOVA revealed a reliable 3-way interaction, $F(3, 303) = 6.38$, $MSE = .046$, $p < .001$, $\hat{\omega}^2_{partial} = .006$, such that, across lists, the JOLs made by free-expecting participants decreasingly differentiated between high and low associative strength pairs ($F(2.4, 117.9) = 40.05$, $MSE = .067$, $\hat{\varepsilon} = .802$, $p < .001$, $\hat{\omega}^2_{partial} = .101$), and did so to a greater degree than did those made by cued-expecting participants ($F(2.5, 128.9) = 14.31$, $MSE = .047$, $\hat{\varepsilon} = .826$, $p < .001$, $\hat{\omega}^2_{partial} = .024$). This pattern was further confirmed by performing separate simple linear regressions predicting difference scores (mean JOLs for high minus low associative strength) 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 very relevant to their task. Just as with their metacognitive control (encoding strategy), their metacognitive monitoring became more attuned to the task.

Characterizing the encoding strategies used.

Questionnaire on encoding strategy. To confirm the same patterns of strategy use as those suggested by the results of Experiment 1, I consider data from the questionnaire and from the two recognition tests. The mean amount of time spent on the questionnaire was 200.9 s (*SD* = 44.8). This value did not reliably differ between test format conditions, *t*(98) = 1.77, *p* = .080, *d* = 0.36. Questionnaire data were not recorded for four participants; thus *N* = 99 for the below analyses (n_{cued} = 50, n_{free} = 49). Table 8 summarizes participants' responses. Figure 10 shows histograms of participants' usage frequency ratings for four of the eleven encoding strategies as a function of test format (cued vs. free).

Because the measure was ordinal, and because the data were not normally distributed, the two-sample Kolmogorov-Smirnov test (which is non-parametric) was used to compare responses between cued-expecting and free-expecting participants for each of the 11 strategies (listed in Appendix B). Because these analyses were pre-planned, an unadjusted alpha level was used. The response distributions reliably differed as a function of test format for only the four strategies shown in Figure 10. Cued-expecting participants reported more usage of a cue-target association strategy (*D*(99) = .337, *z* = 1.68, *p* = .001), while free-expecting participants reported more usage of target-target association ($D(99) = .247$, $z = 1.23$, $p = .032$), target focus ($D(99) = .336$, $z = 1.66$, $p = .001$), and rote rehearsal ($D(99) = .257$, $z = 1.28$, $p = .020$).

Participants expecting different test formats did not differ in the number of different strategies they reported using (i.e., the count of strategies rated > 1), $M_{cued} = 8.7$, $SD_{cued} = 1.7$, $M_{free} = 8.4$, $SD_{free} = 2.0$, $t(97) = 0.87$, $p = .388$, $d = 0.18$. 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, as measured by the proportion of strategies that were rated > 1 for usage and that were also reported as used more in either the 1st half or the 2nd half of the experiment, $M_{cued} = .37$, $SD_{cued} = .30$, $M_{free} = .63$, $SD_{free} = .27$, $t(97) = 4.42$, $p < .001$, $d = 0.90$. Sign tests revealed that free-expecting participants reported more usage in the 1st half versus the 2nd half of the expectancy-inducing cycles for cue-target association ($p = .001$), and more usage in the 2nd half versus the 1st half for: target focus ($p < .001$), mental imagery ($p = .004$), intra-item narrative ($p = .023$), and inter-item narrative ($p = .041$). Cued-expecting participants reported more usage in the 1st half versus the 2nd half for rote rehearsal ($p = .035$), and more usage in the 2nd half versus the 1st half for personal significance ($p = .019$).

To analyze the self-reports on suspicion about changes in test format, participants were classified as either low-suspicion (reporting no suspicion, or reporting that they stopped suspecting during the first half of the experiment) or high-suspicion (reporting that they stopped suspecting during the second half of the experiment, or reporting that they stayed suspicious the whole time). There were more high-suspicion reports for free recall (41/48) versus cued recall (26/50), $z = 3.59$, $p < .001$. For free recall, low-suspicion participants reported more usage of target-target association than did high-suspicion participants, $t(7.9) = 2.85$, $p = .025$, $d = 1.30$. This result suggests that participants who were more convinced that they would receive free recall were more willing to adopt an encoding strategy that was appropriate for free recall. Usage frequency ratings did not reliably differ by suspicion level for any other encoding strategies for free recall, nor for any encoding strategies for cued recall.

Associative recognition. Recognition data were not recorded for three participants; thus $N = 100$ for associative and item recognition analyses ($n_{cued} = 51$, $n_{free} = 49$). As in Experiment 1, associative recognition performance by cued-expecting participants ($M_{d'} = 2.33$, $SD_{d'} = 0.73$) was reliably greater than that by free-expecting participants ($M_{d'} = 1.78$, $SD_{d'} = 0.78$), $t(98) = 3.58$, $p < .001$, $d = 0.72$. Figure 11 and Table 5 show associative recognition performance as a function of test expectancy (cued vs. free) and the list number from which the word pairs originated (1-4), in Experiment 2. Separate simple linear regressions for each participant revealed that performance by free-expecting participants reliably declined across lists of origin, $M_b = -0.18$, $SD_b = 0.32$, $t(48) = -3.88$, $p < .001$, while performance by cued-expecting participants did not reliably change across lists, $M_b = -0.04$, $SD_b = 0.30$, $t(50) = -1.06$, $p = .293$. These results are consistent with those from Experiment 1, and again indicate cued-expecting participants' greater steady use cue-target association strategies, and free-expecting participants' abandonment of such strategies.

Item recognition. Figure 12 shows item recognition performance (d') as a function of test expectancy (cued vs. free) and item type (cue vs. target). A 2-way mixed ANOVA revealed a reliable disordinal interaction between test expectancy and item type, $F(1,98) = 42.53$, $MSE = .112$, $p < .001$, $\hat{\omega}^2_{partial} = .036$. Cue word recognition performance was greater for cued-expecting participants ($M_{d'} = 2.39$, $SD_{d'} = 0.94$) than for free-expecting participants ($M_{d'} = 1.17$,

$SD_{d'}$ = 0.55), $t(98) = 7.42$, $p < .001$, $d = 1.49$. Similarly, target word recognition performance was greater for cued-expecting participants ($M_{d'}$ = 1.93, $SD_{d'}$ = 0.81) than for free-expecting participants ($M_{d'}$ = 1.33, $SD_{d'}$ = 0.67), $t(98) = 3.99$, $p < .001$, $d = 0.80$. For cued-expecting participants, recognition performance was greater for cue words than for target words, $t(50) = 6.84$, $p < .001$, $d = 0.07$, but for free-expecting participants the opposite was true, $t(48) = -2.35$, $p = .023$, $d = -0.03$.

Figure 13 and Table 6 show item recognition performance (hit rate) as a function of test expectancy (cued vs. free), item type (cues vs. targets), and the list number from which the words originated (1-4). A 3-way mixed ANOVA revealed a reliable 3-way interaction, $F(3,294) = 10.08$, $MSE = .021$, $p < .001$, $\hat{\omega}^2_{partial} = .032$, such that item type and list number did not interact for cued-expecting participants, $F(3,150) = 1.38$, $MSE = .014$, $p = .252$, $\hat{\omega}^2_{partial} = .002$, but did interact for free-expecting participants, $F(3,144) = 11.47$, $MSE = .028$, $p < .001$, $\hat{\omega}^2_{partial} = .080$, such that for these participants there was a reliable negative linear trend across lists for cues, $F(1,48) = 21.86$, $MSE = .044$, $p < .001$, $\hat{\omega}^2_{partial} = .129$, but no reliable linear trend for targets, $F(1,48) = 1.11$, $MSE = .023$, $p = .298$, $\hat{\omega}^2_{partial} < .001$. For cued-expecting participants, list number affected neither hit rate for cues, $F(3,150) = 0.49$, $MSE = .013$, $p = .688$, $\hat{\omega}^2_{partial} < .001$, nor hit rate for targets, $F(3,150) = 1.17$, $MSE = .019$, $p = .322$, $\hat{\omega}^2_{partial} = .002$. These results are again consistent with those from Experiment 1.

Efficacy of encoding strategies. The usage frequency ratings from the questionnaire (to the extent that they are accurate) allow us to evaluate the actual efficacy of the various encoding strategies at improving recall performance across lists, and to compare that effectiveness for cued versus free recall. I first performed separate simple linear regressions predicting recall performance from list number for each participant. The estimated slopes from these regressions represent the amount of increase (positive slopes) or decrease (negative slopes) in performance across lists. Next I 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. 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.

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 these analyses, thus $n_{cued} = 46$ and $n_{free} = 48$. Standard errors were calculated 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 was the pooled standard error of the two individual standard errors involved: $\sqrt{SE^2_{\hat{t}_{b\_1}} + SE^2_{\hat{t}_{b\_2}}}$. Because these analyses were pre-planned, an unadjusted alpha level was used.

Table 9 shows 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. 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 three strategies was not associated with changes in performance across lists for cued recall but was associated with increasing performance across free recall lists: target-target association, inter-item association, and target focus. In all three of these cases, the signs of the correlation coefficients were opposite. Finally, inter-item narrative strategy showed a similar pattern to the previous three strategies, but with the same sign for both test formats. It is also worth noting that greater self-reported use of a rote rehearsal strategy (on which participants differed as a function of test expectancy) was not associated with changes in performance across lists for cued recall or free recall, nor were the correlations reliably different.

To better elucidate the above patterns, median splits were performed to compare performance across lists for participants who reported high versus low usage of each strategy, separately for cued recall and free recall. Because the data on which the split was performed were ordinal, there were many ties. For each cell (e.g., cued-target association: cued recall), data from participants whose usage frequency rating matched the median for that cell were either all placed in the high usage group or all placed in the low usage group, on the basis of whichever grouping would come closest to achieving groups of equal size. In two cells (target-target association: free recall, and inter-item narrative: cued recall), this was not possible and thus data from participants with median ratings were omitted from analyses of those two cells (n = 11, n = 4, respectively).

Figure 14 shows mean recall performance as a function of list number (1-4), test format (cued vs. free), and usage (high vs. low) of the six encoding strategies noted above. Data for all eleven strategies are presented in Table 10. The efficacy of each encoding strategy was analyzed—separately for cued versus free recall—by comparing recall performance slopes (across lists 1-4) for high versus low usage. Cue-target association was beneficial for cued recall, $t(48) = 1.85$, $p = .070$, $d = 0.53$, but detrimental for free recall, $t(47) = -2.30$, $p = .033$, $d = -0.73$. Target-target association was inconsequential for cued recall, $t(48) = -0.21$, $p = .833$, $d = -0.07$, but beneficial for free recall, $t(47) = 2.30$, $p = .026$, $d = 0.68$. Inter-item association was inconsequential for cued recall, $t(48) = -1.10$, $p = .279$, $d = -0.33$, but beneficial for free recall, $t(36) = 2.11$, $p = .042$, $d = 0.70$. Target focus was inconsequential for cued recall, $t(48) = 0.18$, $p = .860$, $d = 0.05$, but beneficial for free recall, $t(47) = 3.94$, $p < .001$, $d = 1.14$. Rote rehearsal was inconsequential for both cued recall, $t(48) = 0.40$, $p = .688$, $d = 0.12$, and free recall, $t(47) = 0.24$, $p = .813$, $d = 0.07$. Inter-item narrative was inconsequential for cued recall, $t(44) = -0.49$, $p = .624$, $d = -0.15$, but beneficial for free recall, $t(47) = 3.14$, $p = .003$, $d = 0.93$.

Effectiveness of metacognitive control. Having considered results suggestive of which encoding strategies were more or less effective for cued versus free recall, we can begin to 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 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 though, 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, free-expecting participants did report using cue-target association less as the experiment progressed, and those who were less suspicious of a change in test format reported more usage of target-target association.

A summary of the differential efficacy and use of encoding strategies is shown in Table 15. Of the five encoding strategies which were differentially effective for cued versus free recall in Experiment 2, participants reported appropriate differences in usage for three of these (cue-target association, target-target association, and target focus) but apparently did not differentially employ the other two (inter-item association and inter-item narrative) and additionally differed on usage for one strategy that was inconsequential for both test formats (rote rehearsal). Free-expecting participants reported more usage of rote rehearsal than did cued-expecting participants, who reported using this strategy even less in the 2nd half of the experiment.

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 effectiveness measure for that strategy (tau-b, as described above), 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. For cued recall, this measure was high (r_{cued} = .71, $t(9)$ = 3.04, p = .014) and for free recall it was low (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* effective than other strategies at improving performance. However, this may be largely driven by these participants' early use of cue-target association, before they knew what the test format would be like. This is supported by correlations conditionalized on participants' reporting greater usage in the 1st half of the experiment (r_{free_1} = -.55, $t(9)$ = -1.98, p = .079) versus the 2nd half of the experiment (r_{free_2} = .007, $t(9)$ = 0.02, p = .983), $t_{diff}(8)$ = 1.42, p = .192.

Taken together, these results suggest that participants came equipped with some degree of relevant metacognitive knowledge 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. Giving participants experience with both test formats may provide them with the opportunity to even further adaptively employ different encoding strategies (cf. Bjork, deWinstanley, & Storm, 2007; deWinstanley & Bjork, 2004); this was done in Experiment 3.

Summary of results. Experiment 2 again showed 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, just as with their metacognitive control, their metacognitive monitoring 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, it should be possible to provide learners with an experience that will 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 I 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, I 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).

It was not feasible to use the critical final test manipulation (as in Experiment 1) for evidence of differences in encoding strategy in a fully factorial within-subjects design, because that would require violating participants' expectations more than once. This would be problematic because participants—many of whom enter the lab with a default suspicion of deception in psychology experiments—are unlikely to fall for the same trick twice. Thus, I 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.

I predicted that participants' recall performance, questionnaire responses, and associative recognition performance would show similar patterns to those observed in Experiments 1 and 2, and furthermore that the within-subjects design would engender greater improvement in recall performance than was observed in the between-subjects designs in Experiments 1 and 2. Finally, I also predicted that study-time allocation would also come to reflect important differences between the task demands of cued versus free recall: differentiating between high and low associative strength for cued recall but not for free recall.

Method

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

Design. The experiment used a 2 x 2 within-subjects design, with independent variables: expected final test format (cued recall vs. free recall), and word pair associative strength (high vs. low). 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. As before, all words were 4-8 letter nouns, 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 55.0 (SD_{KF} = 79.1). Mean forward associative strength of word pairs was .026 (SD = .005). For each participant, associative strength was manipulated and pairs were placed into lists as described in Experiment 1.

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. However, there was no question about when each strategy was used most. The questionnaire did include the same final question regarding suspicion of test format change that was used in Experiment 2. The questionnaire instructions also 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. Figure 15 shows mean performance across recall tests 1-3 for cued recall versus free recall. Means and standard deviations are presented in Table 1. 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 16 and Table 2 show mean performance as a function of list number (1-3), test format (cued vs. free), and associative strength (high vs. low). A 3-way within-subjects ANOVA revealed a reliable 2-way interaction between test format and associative strength, $F(1, 84) = 87.05$, $MSE = .020$, $p < .001$, $\hat{\omega}^2_{partial} = .043$, such that performance was superior for high versus low associative strength 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 ($F(1, 84) = 0.06$, $MSE = .015$, $p = .809$, $\hat{\omega}^2_{partial} < .001$). There was no reliable 3-way interaction, $F(2, 168) = 0.39$, $MSE = .013$, $p = .681$, $\hat{\omega}^2_{partial} < .001$, and list number did not interact with associative strength, $F(2, 168) = 1.12$, $MSE = .014$, $p = .329$,

$\hat{\omega}^2_{partial}$ < .001. Thus, across all lists, associative strength was a very important variable for cued recall but not for free recall.

In order to assess whether recall performance improved more when each participant experienced both test formats, two separate ANCOVAs were used (one for cued recall, and one for free recall) to compare list 3 recall performance in Experiment 3 versus Experiments 1 and 2, while partialing out study time duration and mean recall performance on list 1. Study time duration in each experiment was: 4 s for each word pair in Experiment 1; 4 s plus the JOL response time in Experiment 2 (mean of participant median = 5.91 s, *SD* = 1.04); and determined by participants in Experiment 3 (mean of participant median = 4.58 s, *SD* = 2.39). The JOL response times were not recorded for 19 participants, so study time could only be calculated for 84 participants from Experiment 2. One-way ANOVAs confirmed that performance across lists 1-3 did not reliably differ for the participants excluded from this analysis versus those included, neither for cued recall ($F(1, 51) = 0.14$, $MSE = .090$, $p = .709$) nor free recall ($F(1, 48) = 1.17$, $MSE = .023$, $p = .286$). The length of the lists of word pairs in Experiments 1 and 2 was 32, while the list length in Experiment 3 was 24. Shorter list lengths tend to yield higher proportional performance in free recall (Murdock, 1962), but this potential effect was accounted for by treating each participant's mean performance on list 1 as a covariate. The ANCOVA contrast revealed that list 3 performance was not reliably different for Experiment 3 versus Experiment 1 and 2 for cued recall ($F(1, 173) = 0.29$, $MSE = .029$, $p = .594$, $\hat{\omega}^2_{partial} < .001$) but was reliably greater for free recall ($F(1, 171) = 63.65$, $MSE = .026$, $p < .000$, $\hat{\omega}^2_{partial} = .009$). Across experiments, participants seemed to already do well at effectively studying for cued recall. But for free recall, exposure to the explicit pre-presentation instructions and experience with the alternative test format appeared to help participants adaptively change their encoding strategies.

Study-time allocation. Analyses of study-time allocation were carried out on participants' median study time (in seconds) per cell. Figure 17 shows study-time allocation as a function of list number (1-3) and test format (cued vs. free). A 2-way within-subjects ANOVA revealed a reliable negative linear trend in study-time allocation across lists, $F(1, 84) = 38.06$, $MSE = 9.51$, $p < .001$, $\hat{\omega}^2_{partial} = .077$, and no difference in study-time allocation for cued versus free recall, $F(1, 84) = 0.002$, $MSE = 7.32$, $p = .960$, $\hat{\omega}^2_{partial} < .001$. Participants spent less time studying word pairs across lists, but continued to spend about the same studying for cued recall and free recall.

Figure 18 and Table 11 show study-time allocation as a function of list number (1-3), test format (cued vs. free), and associative strength (high vs. low). A 3-way within-subjects ANOVA revealed a reliable 3-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 low versus high associative strength 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 2-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 high and low associative strength pairs across lists, as evidenced by a reliable 2-way interaction between associative strength and the linear effect of list number, $F(1, 84) = 19.44$, $MSE = 1.68$, $p < .001$, $\hat{\omega}^2_{partial} = .007$.

Characterizing the encoding strategies used.

Questionnaire. To confirm the same patterns of strategy use as those suggested by the results of Experiment 1, I consider data from the questionnaire and from the associative recognition test. The mean amount of time spent on the questionnaire was 195.8 s (*SD* = 41.4). Table 12 summarizes participants' responses. Figure 19 shows histograms of participants' usage frequency ratings for five of the eleven encoding strategies as a function of test format (cued vs. free).

Because the usage frequency measure was ordinal, and because the data were not normally distributed, the Wilcoxon matched-pairs signed-rank test (which is non-parametric) was used to compare participants' responses for cued recall to their responses for free recall for each of the 11 strategies. Because of the small ordinal scale used, there were many ties and potentially many difference scores with a value of zero. To account for ties, any tied difference scores were assigned the mean of the ranks involved in that tie. Furthermore, the test statistic (z) was calculated using the large sample normal approximation with correction for ties as provided by Marascuilo and McSweeney (1977, p. 339). I also employed the correction for continuity (Marascuilo & McSweeney, p. 20). Many sources advise discarding difference scores of zero for this test; however, this inflates Type I error rates, especially when there are many zeros. Thus, I 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 tied differences (i.e., they were all assigned the mean of the ranks involved in their 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 the type of data I had. Data from participants with missing values were excluded from analysis on a test-wise (i.e., per strategy) basis; thus, n varied slightly across tests.

Because these analyses were pre-planned, an unadjusted alpha level was used. The response distributions reliably differed as a function of test format for only the five strategies shown in Figure 19. Participants reported more usage in cued recall versus free recall for the strategy of cue-target association ($n = 83$, $T = 83$, $z = 7.65$, $p < .001$). Participants reported more usage in free recall versus cued recall for the strategies of target-target association ($n = 81$, $T = 647$, $z = 4.81$, $p < .001$), target focus ($n = 80$, $T = 259.5$, $z = 6.61$, $p < .001$), rote rehearsal ($n = 83$, $T = 923.5$, $z = 3.82$, $p < .001$), and inter-item narrative ($n = 83$, $T = 967$, $z = 3.61$, $p < .001$). These results match those from Experiment 2, with the addition of a reliable difference on inter-item narrative. Furthermore, as in Experiment 2, participants did not differ in the number of different strategies they reported using (i.e., the count of strategies rated > 1) for cued recall ($M_{cued} = 7.8$, $SD_{cued} = 2.0$) versus free recall ($M_{free} = 7.8$, $SD_{free} = 2.1$), $t(83) = 0.13$, $p = .899$, $d = 0.01$.

Associative recognition. Recognition data were not recorded for eight participants; thus $N = 77$ for the below analyses. As in Experiments 1 and 2, associative recognition performance for word pairs from cued recall lists ($M_{d'} = 1.74$, $SD_{d'} = 0.42$) was reliably greater than that for word pairs from free recall lists ($M_{d'} = 0.82$, $SD_{d'} = 0.52$), $t(76) = 12.44$, $p < .001$, $d = 1.92$. Figure 20 and Table 5 show associative recognition performance as a function of test format (cued vs. free) and the list number from which the word pairs originated (1-3), in Experiment 3. Separate simple linear regressions for each participant and each test format revealed that performance for word pairs from free recall lists reliably declined across lists of origin,

M_b = -0.43, SD_b = 0.58, $t(76)$ = -6.48, $p < .001$, while performance for word pairs from cued recall lists did not reliably change across lists, M_b = -0.005, SD_b = 0.35, $t(76)$ = -0.11, p = .910. This is the same pattern of results found in Experiment 1 and 2.

Efficacy of encoding strategies. The same analytical approach used in Experiment 2 was employed to evaluate the efficacy of the various encoding strategies at improving recall performance across lists, and to compare that effectiveness for cued versus free recall, this time within-subjects. The standard error used for comparison of dependent tau-b values was: $\sqrt{SE^2_{\hat{\tau}_{b\_1}} + SE^2_{\hat{\tau}_{b\_2}} - 2\operatorname{cov}(\hat{\tau}_{b\_1}, \hat{\tau}_{b\_2})}$. The covariance term was calculated using the formula provided by Cliff and Charlin (1991, equation 20, corrected for the erroneously transposed first matrix), with the consistent variance estimates.

Table 13 shows 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 three of the 11 strategies the tau-b correlation coefficients for cued versus free recall significantly differed, or came close to doing so: target-target association, inter-item association, and inter-item narrative. All three strategies showed negative trends for cued recall and positive trends for free recall, suggesting that they were detrimental for cued recall and beneficial for free recall. It is also worth noting that tau-b correlation coefficients did not reliably differ for cued versus free recall for three strategies on which participants' usage frequency ratings did reliably vary as a function of test format: cue-target association, target focus, and rote rehearsal.

Because of the reduced scale used in Experiment 3 (1-4 vs. 1-7 as used in Experiment 2), it was not feasible to perform median splits on usage frequency ratings. Instead, I first computed, for each participant, the mean of that participant's: cued recall performance slope across lists and free recall performance slope across lists. The median of these values was used to split participants into a "high improver" group (n = 36) and a "low improver" group (n = 36). Data from participants who had any missing values were excluded from analysis.

Figure 21 shows, for six encoding strategies, the mean difference in usage frequency rating for free versus cued recall, for high improvers versus low improvers. Data for all eleven strategies are presented in Table 14. Cue-target association was reported as used more for cued recall versus free recall, and this strategic differentiation of usage was greater for participants who improved more across lists of both formats versus participants who improved less across lists of both, $t(70) = -2.23$, $p = .029$, $d = -0.53$. Target-target association was used more in free recall, and this to a greater degree for high improvers versus low improvers, $t(70) = 2.18$, $p = .033$, $d = 0.52$. High and low improvers did not reliably differ on their reported differential usage of: inter-item association, $t(70) = 0.73$, $p = .467$, $d = 0.18$; target focus, $t(70) = -0.40$, $p = .692$, $d = -0.10$; or rote rehearsal, $t(70) = -1.01$, $p = .316$, $d = -0.24$. Inter-item narrative showed the same pattern as target-target association, $t(70) = 2.27$, $p = .021$, $d = 0.57$. In summary, participants whose recall performance improved the most across lists reported greater strategic usage of cue-target association (used more for cued vs. free recall), target-target association (used more for free vs. cued recall), and inter-item narrative (used more for free vs. cued recall).

The preceding analyses on strategy effectiveness 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 significant point will be addressed further in the General Discussion.

Effectiveness of metacognitive control. A summary of the differential efficacy and use of encoding strategies is shown in Table 15. Of the three encoding strategies which were differentially effective for cued versus free recall in Experiment 3, participants reported appropriate differences in usage for two of these (target-target association and inter-item narrative) but apparently did not differentially employ the other one (inter-item association). Participants reported differences in usage for two more strategies that were found to be differentially effective in Experiment 2 but not in Experiment 3: cue-target association, and target focus. Finally, participants again reported differential usage for one strategy that was inconsequential for both test formats (rote rehearsal, greater reported usage for free recall). Overall, participants' encoding strategy usages appear to be fairly well attuned to the different demands of the two test formats, with the salient exceptions being failure to strategically use inter-item association, and needless differential usage of rote rehearsal.

I again quantified participants' metacognitive control effectiveness by calculating the Pearson correlation between the mean usage frequency rating for each strategy with the strategy effectiveness measure for that strategy (tau-b), separately for cued recall and free recall. For cued recall, the correlation was $r_{cued} = .27$, $t(9) = 0.83$, $p = .428$, and for free recall it was $r_{free} = .148$, $t(9) = 0.45$, $p = .665$. 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, perhaps due in part to the smaller rating scale, 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.

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 word pairs with low associative strength when expecting either test format. As shown in Figure 18, participants continued this pattern of allocation across cued recall study-test cycles, but decreasingly differentiated between high and low associative strength 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 effective encoding strategies and study-time allocation strategies. A related finding is that of deWinstanley and Bjork (1994), who found that when participants were given a chance to experience the differential performance benefits for generated versus read items, they improved their subsequent performance on read items to the level of the generated items; this suggests that participants spontaneously, and adaptively, changed the way that they processed the read items.

General Discussion

Summary of Results

In this study I asked whether learners can adaptively and *qualitatively* modulate their encoding strategies in anticipation of future task demands. In Experiment 1 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 format. The key result was a crossover 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. In Experiment 2 participants furthermore demonstrated concomitant and judicious attunement of metacognitive monitoring, decreasingly differentiating between high and low associative strength word pairs for free recall but not cued recall, as shown in Figure 9. In Experiment 3, which used a within-subjects design, participants demonstrated adaptive changes in metacognitive control of encoding strategy, and of study-time allocation: participants began the experiment spending more time studying word pairs with low versus high associative strength for both test formats, and they decreasingly made this distinction for free recall (for which associative strength was inconsequential), as shown in Figure 18. Furthermore, the explicit instructions and experience with both test formats provided by Experiment 3 enabled participants to adjust their free recall strategies even more adaptively than they had in Experiments 1 and 2. Finally, all three experiments provided insights into the characteristics of the encoding strategies that participants used. In studying for a cued recall test participants relied heavily and consistently on a strategy of cue-target association; in studying for a free recall test, participants abandoned cue-target association in favor of multiple strategies: selective attention to target words (i.e., target focus), making associations across word pairs (target-target association, inter-item association, and inter-item narrative), and rote rehearsal. Participants' metacognitive control of encoding strategies was mostly effective, though not without room for improvement, especially for free recall.

Relation to Prior Research

The present findings are consistent with some prior research. For example, in studies of learning to learn, Postman (1964, 1969) found that several types of recall performance improved across unrelated lists as they acclimated to the task. It is also clear from studies of intentional versus incidental learning that knowledge at all of an upcoming test can change the way participants encode information, though specific knowledge may do so more potently (McDaniel, Blischak, & Challis, 1994). Furthermore, several researchers have advanced views of human memory as a skill that can be improved (cf. Benjamin, 2008; Chase & Ericsson, 1981). Ericsson's work to account for the development of exceptional performance by experts led to the theory that, over years of deliberate practice at domain tasks, experts develop specialized "retrieval structures" (Ericsson & Kintsch, 1995) that enable them to rapidly encode and subsequently retrieve new information in their specific domain in a way that provides both organization and relation to existing knowledge. Such specialized encoding strategies should be learnable by anyone, given enough practice. For example, Ericsson and Chase (1982) worked with an undergraduate, SF, who increased his memory for numbers from a digit span of 7 to upwards of 80, all through the spontaneous development of his own mnemonics over hundreds of hours of lab testing and practice. McDaniel and Kearney (1984) instructing participants to use different encoding strategies (mental imagery, categorization, and sentence construction) led to different patterns of performance for different stimuli and test formats. This, along with many other studies using orienting tasks, demonstrates learners' abilities to execute a variety of encoding strategies. Furthermore, when another group of participants was given no orienting task, they appeared to generally use the most task-appropriate strategy for the stimuli they studied (categorized lists of single words, lists of word pairs, and lists of uncommon words with definitions), prompting McDaniel and Kearney to conclude that "mature learners seem to spontaneously utilize semantic and imaginal strategies and do so task appropriately." Finally, as noted in the Introduction, a little-known handful of prior test expectancy experiments have also shown some evidence of learners adopting qualitatively different encoding strategies (von Wright, 1977; von Wright & Meretoja, 1975; Postman & Jenkins, 1948).

All of these lines of research suggest that human learners are capable of flexible and adaptive metacognitive control of encoding strategies. However, such a view is in contrast to the many test expectancy experiments that have found overall performance patterns that provide only evidence of quantitative differences in encoding strategies (Balota & Neely, 1980; Carey & Lockhart, 1973; Connor, 1977, Experiment 1; d'Ydewalle, 1981, 1982; d'Ydewalle et al., 1983; Foos & Clark, 1983; Hall et al., 1976; Jacoby, 1973; Lewis and Wilding, 1981; Loftus, 1971; Maisto et al., 1977; May & Sande, 1982; Meyer 1934, 1936; Neely & Balota, 1981; Oakhill & Davies, 1991; Schmidt, 1988; Thiede, 1996; Tversky, 1973; Wnek & Read, 1980; see also Lundeberg & Fox, 1991), or no evidence of differences at all (Feldt, 1990; Freund, Brelsford, & Atkinson, 1969; Glass, Clause, & Kreiner, 2007; Kardash & Kroeker, 1989; Kulhavy, Dyer, & Silver, 1975; Lovelace, 1973, Experiments 6-9; McDaniel et al., 1994; Rickards & Friedman, 1978). In summarizing their findings, Hall et al. (1976) concluded that "a view of the learner as a highly active, flexible resourceful strategist … seems to overestimate the degree of control that subjects exercise over the nature of their information processing for memory." In the sections to follow, I explore possible reasons for this conundrum, including the relative value of alternative forms of metacognitive control, prerequisites for effective encoding strategy use, and methodological requirements for detecting qualitative changes and differences in encoding strategies.

Alternatives to Adjusting Encoding Strategies

It may be that, instead of adjusting their encoding strategies, learners generally rely on other forms of metacognitive control, such as item selection, study-time allocation, and scheduling, to modulate their learning in order to meet expected demands of an upcoming test. The literature on these methods of control suggests that learners do indeed use them strategically in the face of varying task demands (cf. Benjamin & Bird, 2006; Finley et al., 2010; Kornell & Metcalfe, 2006; Son, 2004; Son & Metcalfe, 2000). For example, Thiede (1996, Exp. 2), using a test expectancy method in which participants controlled study-time allocation, found that participants expecting a cued recall test studied longer than those expecting a recognition test. It is also worth observing that, although college students often show keen interest in the format of upcoming midterm and final exams, they are more apt to first ask instructors about the *content* of the exams (i.e., "What will be on the test?"), which is a task demand that bears more on item selection and study-time allocation than on encoding strategy. Crooks (1988) concluded that "student expectations of the cognitive level [e.g., surface- vs. deep-processing] and content of tasks probably exert much more influence on their study behavior and achievement than do their expectations of the task format (for given content and cognitive level)."

Compared to spending more time studying, being more selective about what is studied, or simply putting more effort into using even a modestly generally effective encoding strategy, developing and using transfer-appropriate encoding strategies may not be the most cost-effective approach to attaining desired levels of memory performance. According to the conceptual framework proposed by Hertzog and Dunlosky (2004), the demands of such an approach can include: appraising the task, retrieving potential strategies, selecting and executing an appropriate strategy, monitoring learning, and adjusting strategy use accordingly.

Prerequisites for Effective Encoding Strategy Use

Metacognitive monitoring. To accommodate their encoding strategies to future test conditions, learners must be able to accurately monitor their ongoing learning (e.g., as demonstrated in Experiment 2), and also emulate their relevant future cognitive states. Learners may have difficulty assessing the cognitive demands of a future test. For example, if they under-appreciate their own rate of forgetting (Koriat & Bjork, 2005), they may underestimate the initial degree or depth of encoding that they should seek in order to maximize later retrieval. A primary challenge for learners in this situation is the difficulty of discounting their potentially misleading current knowledge state when predicting future performance (Benjamin, Bjork, & Schwartz, 1998; Kelley & Jacoby, 1996). The difficulty of these metacognitive efforts may cause learners to struggle with selecting an appropriate encoding strategy, or with adequately applying such a strategy. Thus, giving learners *experience* with a particular type of learning material and test format across multiple study-test cycles (e.g., as opposed to merely giving instructions about an upcoming test) may be critical in enabling accurate metacognitive monitoring and control.

Metacognitive knowledge. The effectiveness of self-regulated learning depends in part on a learner's metacognitive knowledge (cf. Hertzog & Dunlosky, 2004; Winne, 1995). Von Wright (1975) observed that "...it is by no means obvious that performance should be optimal when the method of testing retention is that anticipated by the subject. Subjects may not know how to encode a material 'efficiently' for a particular type of test and may choose their learning strategies unwisely." In addition to accurate metacognitive monitoring, learners must also be equipped with a repertoire of relevant encoding strategies, or be able to devise new strategies as needed. Free recall is a less constrained task than cued recall, and thus there are a greater number of potentially effective encoding strategies that learners could use. But learners may not have prior knowledge of all such strategies, may fail to retrieve them from memory, or may be unwilling to commit the resources to an effective but difficult strategy. This implies that there should also be more room for improvement in encoding strategy use for free recall versus cued recall, as was observed in Experiments 1-3 of the present study. Thus, again, experience with a leaning task may be critical for enabling development or activation of appropriate knowledge (Delaney & Knowles, 2005; Hertzog, Price, & Dunlosky, 2008).

Goals, motivation, and beliefs. Effective use of encoding strategies furthermore requires that learners have the goal of attaining high performance on a learning task, are motivated enough to pursue that goal, and enabled by the belief that their efforts will be fruitful. When learners' goal is to master learning material, they allocate study-time more strategically than when their goal is a much less difficult one (Son & Metcalfe, 2000; Thiede & Dunlosky, 1999). Given the effort required to custom tailor encoding strategies to expected test format it is likely that learners will not be motivated to go to the trouble if they do not have a goal of high performance. Furthermore, Dweck and colleagues (Dweck, 1986; Elliott & Dweck, 1988) have shown that learners who believe intelligence is a fixed trait are less motivated to put effort into learning than are learners who believe intelligence is an improvable skill. Learners may hold a variety of beliefs about how memory works (Magnussen et al., 2006), and may have anxieties about memory testing that moderate the effects of test expectancy (Minnaert, 2003). Individual differences in goals, motivation, and beliefs are integrated in several accounts of self-regulated learning in general by educational researchers (Biggs, 1985; Butler & Winne, 1995; Pintrich, 2000; Winne, 2001, 2005; Winne & Hadwin, 1998; Zimmerman, 1989, 2002). Two well-established instruments for measuring the ways in which learners study have also arisen from this literature. The Learning and Study Strategies Inventory (LASSI; Weinstein & Palmer, 2002) is based on a model of strategic learning with three components: skill, will, and self-regulation. The Study Process Questionnaire (SPQ; Biggs, Kember, & Leung, 2001) is based on measuring both motives and strategies across three overall approaches to learning: surface, deep, and achieving. Finally, Hertzog and Dunlosky (2004) proposed a conceptual framework that ties together studies on strategic behavior in associative learning tasks. In their framework, as in models from the self-regulated learning literature, learners' epistemologies and performance goals are de facto prerequisites for adaptive encoding strategy use.

Methodological Requirements for Detecting Qualitative Changes and Differences in Encoding Strategy

When the prerequisites above are all satisfied, and when alternative forms of metacognitive control are either unavailable or insufficient, learners may indeed use qualitatively different encoding strategies that are effective for the particular type of test they expect. However, there are several methodological (aka situational) requirements that must be met in order to detect qualitative changes and differences in encoding strategy as a function of test expectancy, particularly in order to detect the distinctive and elusive disordinal interaction between test format expected and test format received. I outline these requirements as follows:

1. Task demands for the two (or more) test types must be different enough that a single encoding strategy does not suffice for attaining performance goals across test types. Conflicting task demands best meet this requirement.

2. Stimuli and method of presentation must sufficiently allow for variability in the ways that items can be encoded.
3. Dependent measures must be sufficiently sensitive and appropriate to detect differences in encoding strategies that are relevant to the task demands.

I will now consider how these methodological requirements help to explain the discrepant findings in studies using test expectancy methods.

Task demands. The first methodological requirement, which was also suggested by Sanders and Tzeng (1975), is that task demands for the two (or more) test types be different enough that a single encoding strategy does not suffice for attaining performance goals across test types. This requirement may play a large role in the widespread failure to find a disordinal interaction between test format expected and test format received for free recall versus recognition. 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 to 7 scale) and that participant's test performance. That is, the same encoding strategies were beneficial for free recall and recognition. Thus, free recall and recognition may overlap too much in their task demands to prompt qualitative differences in encoding strategy. 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." In terms of the framework proposed by Hunt and McDaniel (1993), the task demands of both free recall and recognition call for predominantly distinctive processing.

The methodological requirement of differing task demands may similarly speak to Jacoby's (1973) failure to find a disordinal interaction, despite pitting cued recall against free recall and inducing expectancy via multiple study-test cycles (as in Experiment 1 of the present study). In Jacoby's experiment, the items presented were single words, each of which shared a semantic category with six other words in a given list. The cues given in cued recall were the category names. Thus, each cue was tied to seven different targets. Requiring participants to recall multiple specific targets from a given category may have shifted the most appropriate encoding strategy away from predominantly cue-target relational processing toward more distinctive processing and/or target-target relational processing, both of which would also be appropriate for free recall.

Finally, the requirement of differing task demands may explain, in part, the success of the few studies that have found evidence of qualitative differences in encoding strategies. In the present study, the correlational analyses of encoding strategy effectiveness in Experiments 2 and 3 clearly demonstrated that not only were different strategies beneficial for cued recall (cue-target association) versus free recall (e.g., target focus, target-target association), but furthermore that some strategies that were beneficial for one test format were *detrimental* for the other format. Thus, the task demands of the two test formats, as implemented in the present study, were conflicting.

In the studies by von Wright (1977) and von Wright and Meretoja (1975), the disordinal interaction was found for serial recall versus recognition, but not for free recall versus recognition nor for free recall versus serial recall. Serial recall, while similar in task demands to free recall (cf. Bhatarah, Ward, & Tan, 2008), was likely more different from recognition than free recall was. The specificity of the task demands of serial recall (i.e., outputting items in the same order as they were presented) may have led participants to employ a serial association encoding strategy, which would be beneficial for serial and free recall but not for recognition (which would benefit more from distinctive rather than relational encoding strategies).

In order to explain the lone result showing a disordinal interaction for free recall versus recognition (Postman & Jenkins, 1948), along with many other results, we must turn to the second methodological requirement.

Stimuli and presentation. The second methodological requirement—stimuli and presentation that allow for variability in encoding—was pointed out by Tversky (1973) as an advantage of picture stimuli, which can be encoded visually and/or verbally (see also Peeck, Van Dam, & de Jong, 1978). Balota and Neely (1980) also spoke to this issue in proposing that test expectancy effects are attenuated to the extent that stimuli restrict free-recall-expecting participants from doing more variable encoding than recognition-expecting participants (e.g., when low frequency words are used, providing fewer potential meanings to leverage for encoding; see also May & Sande, 1982). Semantic organization of word lists has also been found to interact with expected test format (Connor, 1977; Neely & Balota, 1981; Schmidt, 1988).

The stimuli and presentation requirement potentially explains why the inducement of expectancy using instructions alone, or only using only one practice test, often does not result in test expectancy effects: participants have not been given enough experience or opportunity to develop or select differentiated encoding strategies. That experience with a test format is more effective at inducing test expectancy than instructions alone was noted in the meta-analysis by Lundberg and Fox (1991) for multiple choice tests in classroom studies, and was also noted by McDaniel et al. (1994) for laboratory studies that used prose material. Hall et al. (1976), in laboratory studies using word lists, found a small effect of test expectancy using instructions alone (Experiment 2), but greater effects using practice tests (Experiments 1 & 3). Furthermore, in their third experiment Hall et al. found a test expectancy effect when the total time participants were given to study 28 words was longer (180 s) but not when it was shorter (90 s). Balota and Neely (1980) also argued that failures to find test expectancy effects on recognition performance may be due to insufficient practice.

The stimuli and presentation requirement again helps explain the few studies that have found evidence of qualitative differences in encoding strategies. The present study used word pairs as stimuli, in order to accommodate the use of cued recall. Word pairs afford more potential variation in encoding strategy than single words, which have been used as stimuli in most prior test expectancy studies using discrete material. Furthermore, experiments in the present study induced test expectancy over the course of three or four practice study-test cycles, which apparently provided participants with adequate experience to cater their encoding strategies to their expected test format.

The studies by von Wright (1977) and von Wright and Meretoja (1975) used picture stimuli (drawings of objects), which, as noted above, likely provide for more varied encoding than words. Furthermore, although these two studies induced expectancy for test format by instructions alone, they did something which almost no other test expectancy studies have done: used multiple presentations. Items were presented four times for 3 s each in von Wright and Meretoja and two times for 3 s each in von Wright. Von Wright reported that the effects of test expectancy in his experiment were smaller than those found in von Wright and Meretoja, and commented that "this is presumably due to the fact that while a set of fairly elaborate pictures, providing good opportunity for differential encoding, was used in the former study, the pictures in the present experiment were both fewer and simpler." The later study also used fewer presentation repetitions.

The study by Postman and Jenkins (1948) used adjective words as stimuli and induced expectancy by instructions alone, neither of which should have facilitated differential encoding under the present conceptual framework. However, this study also used multiple presentations, with each word read aloud by the experimenter a total of five times. That the use multiple presentations alone could account for the exceptional finding by Postman and Jenkins is supported by the findings of Maisto et al. (1977). They induced expectation of free recall versus recognition via instructions and experience with one practice study-test cycle, and also manipulated the number of times that items were presented: one versus three (between-subjects). They found that, on a final test of free recall, free-recall-expecting participants reliably outperformed recognition-expecting participants *only* when three presentations were used.

Finally, with respect to the stimuli and presentation requirement, it is worth considering the use of prose material (i.e., text passages) in test expectancy studies. Test expectancy effects have been found less consistently with prose than with discrete materials such as word lists (cf. d'Ydewalle et al., 1983; McDaniel et al., 1994; Oakhill & Davies, 1991). There are several possible reasons for this. First, memory performance for prose material may be more heavily influenced by particular characteristics of the text, such as narrative structure (McDaniel et al.). Second, although prose may potentially offer more different ways to encode to-be-remembered information than discrete stimuli would, it also introduces opportunities for participants to adaptively exercise item selection and study-time allocation for subsets of the prose, thus making isolation of encoding strategy effects more difficult. One way to ameliorate this problem is to use a kind of "moving window" method such that single sentences of a passage are presented one at a time, as in McDaniel et al.

Dependent measures. The third and final methodological requirement is that dependent measures be sufficiently sensitive and appropriate to detect differences in encoding strategies that are relevant to the task demands. This requirement is consistent with the efforts of some researchers to seek evidence of encoding strategy differences not in overall levels of test performance (e.g., accuracy) but rather in nuances of performance such as intra-category serial position functions (cf. Carey & Lockhart, 1973; Hall et al., 1976) or semantic organization of output in free recall (cf. D'Ydewalle, 1982; Jacoby 1973). However, to the extent that the task demands differ—or even better, directly conflict—for the test formats used for expectancy (the first methodological requirement), overall final performance on these test formats may well suffice as sensitive measures. This was the case with the few studies that have shown the disordinal interaction between test format expected and test format received (including Experiment 1 of the present study). Otherwise, additional measures may be needed that allow the decomposition of overall performance along dimensions relevant to likely differences in encoding strategy. For example, in the present study the primary result was the disordinal interaction in overall recall performance on the final critical test in Experiment 1; this was bolstered by additional final tests of associative recognition (with performance analyzed as a function of test expectancy and list of origin), and item recognition (with performance analyzed as a function of test expectancy, list of origin, and item type [cues vs. targets]). In order to devise sensitive measures such as these, researchers must already have an idea of what different encoding strategies participants are likely to employ. These may be predicted from theory, from previous research, or from pilot studies. Self-reports from participants may be particularly helpful as well, and can themselves comprise compelling data (cf. Hall et al., 1976; Leonard & Whitten, 1983). Especially where strategy use is concerned, careful use of such qualitative methods may enable key insights that using quantitative methods alone cannot (cf. Dunlosky & Hertzog, 2001; Ericsson & Simon, 1993; Newell, 1973).

A final consideration with respect to the third requirement is that, in many cases, a variety of encoding strategies are likely employed across participants in the same expectancy conditions, and even within participants. This implies that, unless task demands of two test formats are in direct opposition, there may be qualitative differences in group encoding strategy that take the form of different relative *proportions* of various strategies. For example, participants in the cued-expecting conditions in Experiment 1 of the present study appear to have encoded cue-target associations to a greater extent than they selectively attended to the target words (but didn't use either strategy exclusively), while participants in the free-expecting conditions appear to have done the opposite. Such qualitative differences in relative proportion of strategy use may not always be reflected in overall final performance (though in this case, they were). Thus, even if the first methodological requirement is met, there may be need for measures of final performance that are more sensitive than the expected test formats themselves. I believe that a major strength of the current study was the variety of dependent measures used and the convergence of results that they provided.

Future Directions

The points covered in the General Discussion may help guide future studies of the abilities of learners to adaptively cater their encoding strategies to suit expected task demands. The framework I have presented highlights dimensions likely to modulate the amount of observed adaptation in encoding strategy. Alternative forms of metacognitive control, if they are allowed, may overshadow changes or differences in encoding strategy. To effectively use encoding strategies, learners must be equipped with adequate and appropriate metacognitive monitoring skills, metacognitive knowledge, and goals, motivations, and beliefs. Studies using test expectancy in search of qualitative differences in encoding strategies should use test formats with conflicting task demands, should use stimuli and presentation methods that facilitate variations in encoding strategy (including giving participants experience with the task), and should make thoughtful use of multiple dependent measures, including self-reports.

In addition to incorporating the above considerations, future work should do more to systematically characterize and evaluate the variety of encoding strategies that learners may use for given tasks and learning material. For example, Tversksy (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). Efforts should also be made to better integrate empirical studies of encoding strategy with theoretical models and frameworks such as those by Hertzog and Dunlosky (2004) and Winne and Hadwin (1998). Further efforts might be made to model specific encoding strategies as mediating variables between expectancy and performance (Murayama, 2005), or to formally model optimal encoding strategy use as Son and Sethi (2006) have recently done for study-time allocation. Such coupling of continued empirical work with overarching theoretical work should advance our understanding of metacognitive control processes in self-regulated learning.

Conclusion

This study used the test expectancy method to investigate adaptive 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.

Tables

Table 1

Means (and Standard Deviations) of Recall Performance in Experiments 1-3

		List Number			
Test Format	*n*	1	2	3	4
Experiment 1					
Cued Recall	50	.52 (.18)	.58 (.20)	.54 (.23)	.55 (.26)
Free Recall	50	.16 (.09)	.14 (.08)	.17 (.11)	.21 (.11)
Experiment 2					
Cued Recall	53	.61 (.18)	.60 (.17)	.59 (.23)	.53 (.21)
Free Recall	50	.19 (.11)	.13 (.08)	.19 (.18)	.21 (.16)
Experiment 3					
Cued Recall	85[a]	.71 (.20)	.71 (.21)	.66 (.21)	
Free Recall	85[a]	.34 (.24)	.43 (.29)	.45 (.27)	

Note. [a]Test format was manipulated within-subjects in Experiment 3.

Table 2

Means (and Standard Deviations) of Recall Performance by Associative Strength in Experiments 1-3

	List Number			
Test Format and Assoc. Strength	1	2	3	4
Experiment 1				
Cued Recall				
High Assoc.	.63 (.19)	.69 (.19)	.64 (.24)	.63 (.26)
Low Assoc.	.40 (.22)	.46 (.26)	.44 (.25)	.46 (.28)
Free Recall				
High Assoc.	.17 (.10)	.16 (.10)	.17 (.12)	.24 (.14)
Low Assoc.	.15 (.13)	.12 (.10)	.17 (.14)	.19 (.12)
Experiment 2				
Cued Recall				
High Assoc.	.75 (.20)	.75 (.16)	.70 (.24)	.67 (.25)
Low Assoc.	.47 (.23)	.45 (.22)	.47 (.27)	.39 (.23)
Free Recall				
High Assoc.	.22 (.12)	.15 (.10)	.21 (.17)	.23 (.15)
Low Assoc.	.15 (.13)	.11 (.09)	.17 (.20)	.19 (.18)
Experiment 3				
Cued Recall				
High Assoc.	.79 (.19)	.78 (.21)	.75 (.21)	
Low Assoc.	.63 (.23)	.64 (.24)	.57 (.25)	
Free Recall				
High Assoc.	.34 (.26)	.42 (.31)	.45 (.29)	
Low Assoc.	.34 (.25)	.43 (.29)	.45 (.27)	

Table 3

Frequencies of Self-reported Encoding Strategies in Experiment 1

Encoding Strategy	Expected Test Format: Cued Recall	Expected Test Format: Free Recall	Cued vs. Free: *z*	Cued vs. Free: *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. $n = 50$ for both test formats; statistically significant p-values are shown in boldface (Bonferroni corrected alpha level of .0042).

Table 4

Frequencies of Changes to Encoding Strategies that Participants Reported they Would Have Made in Experiment 1

Expected Test Format	Received Test Format	Focus on Targets	Attend More to Cues	Make Cue-Target Associations	Make Target-Target Associations
Cued	Cued	0	0	1	0
Cued	Free	14	0	1	2
Free	Cued	1	10	6	0
Free	Free	14	1	2	1

Note. n = 25 for each condition.

Table 5

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

		List of Origin				
Test Format	*n*	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)
Free Recall	22	1.55 (0.84)	1.48 (0.79)	0.99 (0.90)	1.03 (1.02)	0.75 (0.97)
			Experiment 2			
Cued Recall	51	2.17 (0.69)	2.17 (0.52)	1.96 (0.84)	2.09 (0.79)	
Free Recall	49	2.07 (0.61)	1.62 (0.96)	1.72 (0.82)	1.44 (0.99)	
			Experiment 3			
Cued Recall	77[a]	1.76 (0.57)	1.71 (0.68)	1.75 (0.51)		
Free Recall	77[a]	1.34 (0.76)	0.65 (0.84)	0.48 (0.86)		

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

Table 6

Means (and Standard Deviations) of Item Recognition Performance in Experiments 1-2

Test Format and Item Type	*n*	List of Origin 1	2	3	4	5
		Experiment 1				
Cued Recall	21					
Cues		.83 (.14)	.89 (.13)	.85 (.18)	.86 (.15)	.84 (.18)
Targets		.72 (.21)	.76 (.18)	.72 (.17)	.77 (.19)	.71 (.22)
Free Recall	22					
Cues		.72 (.21)	.68 (.18)	.60 (.23)	.55 (.29)	.50 (.18)
Targets		.70 (.23)	.60 (.25)	.72 (.18)	.73 (.16)	.73 (.20)
		Experiment 2				
Cued Recall	51					
Cues		.85 (.16)	.88 (.14)	.88 (.16)	.87 (.15)	
Targets		.79 (.16)	.77 (.18)	.78 (.18)	.74 (.20)	
Free Recall	49					
Cues		.69 (.22)	.69 (.19)	.59 (.22)	.52 (.26)	
Targets		.70 (.17)	.61 (.20)	.69 (.20)	.71 (.21)	

Note. Experiment 1 data are only from participants who received their expected test format; performance measure was hit rate; Experiment 3 did not include an item recognition test.

Table 7

Means (and Standard Deviations)of Judgments of Learning in Experiment 2

Test Format and Associative Strength	List Number 1	2	3	4
Cued Recall				
High Assoc.	2.93 (0.42)	2.86 (0.45)	2.80 (0.52)	2.72 (0.63)
Low Assoc.	1.90 (0.35)	2.03 (0.42)	2.06 (0.50)	2.01 (0.49)
Free Recall				
High Assoc.	2.96 (0.49)	2.45 (0.55)	2.32 (0.55)	2.17 (0.47)
Low Assoc.	2.01 (0.45)	1.89 (0.43)	1.90 (0.50)	1.90 (0.47)

Note. Response scale was 1 (*I am sure I will NOT remember this item.*) to 4 (*I am sure I WILL remember this item.*); n_{cued} = 53; n_{free} = 50.

Table 8

Encoding Strategy Usage Frequency Ratings in Experiment 2

	Cued Recall Expectation		Free Recall Expectation	
Encoding Strategy	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Cue-target Association	5.60 (1.92)	6.5	4.96 (1.35)	5
Target-target Association	2.32 (1.58)	2	3.06 (2.22)	2
Inter-item Association	2.58 (1.74)	2	2.53 (1.67)	2
Target Focus	3.24 (1.74)	3.5	4.58 (1.88)[b]	5[b]
Mental Imagery	4.98 (1.87)[a]	5[a]	4.59 (2.06)	5
Rote Rehearsal	4.32 (1.87)	4	5.20 (1.48)	5
Verbalization	4.12 (2.35)	4.5	3.84 (2.43)	4
Intra-item Narrative	4.15 (2.03)[b]	4[b]	3.88 (2.36)	5
Inter-item Narrative	3.39 (2.24)[a]	3[a]	2.94 (2.41)	1
Personal Significance	4.86 (1.90)	5.5	4.08 (2.21)	5
Observation	4.00 (1.81)	4	4.43 (1.69)	4

Note. Rating scale was 1 (*no use*) to 7 (*extensive use*); $n_{cued} = 50$; $n_{free} = 49$.
[a]$n = 49$. [b]$n = 48$.

Table 9

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

	Cued Recall		Free Recall		Cued vs. Free	
Encoding Strategy	$\hat{\tau}_b$ *(SD)*	95% CI	$\hat{\tau}_b$ *(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]**
Inter-item 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]
Intra-item Narrative	.20 (.10)	**[.002, .41]**	.23 (.12)	[-.01, .47]	.16	[-.34, .28]
Inter-item 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 used $z_{\alpha/2}$ = 1.96 and standard errors calculated as per Woods (2007) using consistent variance estimates from Cliff & Charlin (1991); statistically significant CIs are shown in boldface.

Table 10

Means (and Standard Deviations) of Recall Performance by Self-rated Encoding Strategy Usage in Experiment 2

Encoding Strategy and Usage Level	*n*	Cued Recall				
		List 1	List 2	List 3	List 4	Slope
Cue-target Association						
High	25	.63 (.20)	.63 (.17)	.63 (.23)	.59 (.22)	-.01 (.07)
Low	25	.60 (.17)	.57 (.16)	.55 (.23)	.45 (.18)	-.05 (.06)
Target-target Association						
High	29	.62 (.15)	.59 (.16)	.58 (.19)	.52 (.18)	-.03 (.06)
Low	21	.61 (.22)	.61 (.19)	.60 (.29)	.53 (.26)	-.03 (.08)
Inter-item Association						
High	22	.59 (.19)	.58 (.18)	.52 (.26)	.47 (.21)	-.04 (.08)
Low	28	.64 (.18)	.62 (.16)	.65 (.20)	.56 (.21)	-.02 (.06)
Target Focus						
High	25	.61 (.17)	.60 (.16)	.56 (.21)	.53 (.20)	-.03 (.05)
Low	25	.62 (.20)	.61 (.18)	.62 (.26)	.51 (.22)	-.03 (.08)
Mental Imagery						
High	23	.61 (.17)	.59 (.17)	.67 (.19)	.58 (.19)	.00 (.06)
Low	26	.62 (.20)	.61 (.17)	.52 (.25)	.48 (.22)	-.05 (.07)
Rote Rehearsal						
High	24	.65 (.18)	.63 (.18)	.65 (.21)	.55 (.20)	-.02 (.05)
Low	26	.58 (.18)	.58 (.16)	.53 (.24)	.49 (.22)	-.03 (.08)
Verbalization						
High	25	.67 (.16)	.63 (.19)	.66 (.20)	.58 (.19)	-.02 (.05)
Low	25	.56 (.19)	.58 (.15)	.52 (.25)	.46 (.22)	-.04 (.08)
Intra-item Narrative						
High	23	.62 (.21)	.62 (.19)	.63 (.29)	.57 (.25)	-.02 (.08)
Low	25	.61 (.16)	.59 (.15)	.55 (.17)	.48 (.17)	-.04 (.05)
Inter-item Narrative						
High	23	.66 (.17)	.61 (.16)	.64 (.24)	.55 (.21)	-.03 (.09)
Low	22	.57 (.21)	.59 (.19)	.55 (.24)	.50 (.22)	-.03 (.04)
Personal Significance						
High	25	.60 (.18)	.61 (.16)	.63 (.22)	.56 (.18)	-.01 (.04)
Low	25	.63 (.18)	.60 (.18)	.55 (.25)	.48 (.23)	-.05 (.08)
Observation						
High	30	.62 (.19)	.59 (.17)	.54 (.24)	.47 (.21)	-.05 (.07)
Low	20	.61 (.17)	.61 (.17)	.66 (.21)	.60 (.19)	.00 (.05)

(Table continues)

Table 10 (continued)

Encoding Strategy and Usage Level	*n*	Free Recall List 1	List 2	List 3	List 4	Slope
Cue-target Association						
High	30	.20 (.11)	.12 (.06)	.14 (.08)	.17 (.10)	-.01 (.05)
Low	19	.18 (.11)	.14 (.10)	.27 (.26)	.27 (.21)	.04 (.08)
Target-target Association						
High	24	.20 (.13)	.13 (.09)	.26 (.23)	.27 (.19)	.03 (.08)
Low	25	.18 (.10)	.13 (.07)	.13 (.08)	.15 (.09)	-.01 (.04)
Inter-item Association						
High	19	.19 (.09)	.14 (.09)	.24 (.21)	.29 (.19)	.04 (.07)
Low	19	.18 (.07)	.13 (.07)	.18 (.20)	.16 (.11)	.00 (.04)
Target Focus						
High	25	.16 (.08)	.13 (.08)	.25 (.23)	.26 (.17)	.04 (.07)
Low	23	.22 (.13)	.14 (.08)	.12 (.08)	.15 (.12)	-.02 (.05)
Mental Imagery						
High	21	.18 (.09)	.16 (.08)	.24 (.20)	.25 (.20)	.03 (.07)
Low	28	.20 (.13)	.11 (.07)	.15 (.16)	.18 (.11)	.00 (.06)
Rote Rehearsal						
High	24	.18 (.09)	.13 (.09)	.18 (.17)	.21 (.13)	.01 (.05)
Low	25	.19 (.13)	.13 (.07)	.20 (.20)	.20 (.18)	.01 (.08)
Verbalization						
High	24	.21 (.10)	.15 (.09)	.20 (.21)	.21 (.15)	.01 (.05)
Low	25	.17 (.12)	.12 (.07)	.19 (.15)	.21 (.17)	.02 (.08)
Intra-item Narrative						
High	25	.18 (.09)	.14 (.08)	.22 (.20)	.24 (.16)	.02 (.05)
Low	24	.20 (.13)	.12 (.07)	.16 (.16)	.18 (.15)	.00 (.08)
Inter-item Narrative						
High	23	.18 (.08)	.13 (.08)	.25 (.24)	.28 (.19)	.04 (.07)
Low	26	.20 (.13)	.13 (.08)	.14 (.09)	.15 (.10)	-.01 (.06)
Personal Significance						
High	26	.17 (.09)	.13 (.09)	.19 (.15)	.23 (.17)	.03 (.06)
Low	23	.22 (.13)	.14 (.07)	.19 (.21)	.18 (.14)	.00 (.07)
Observation						
High	23	.19 (.14)	.12 (.07)	.13 (.08)	.16 (.09)	-.01 (.06)
Low	26	.19 (.09)	.14 (.08)	.24 (.23)	.25 (.19)	.03 (.07)

Table 11

Means (and Standard Deviations) of Study-time Allocation in Experiment 3

Test Format and Associative Strength	List Number 1	2	3
Cued Recall			
High Assoc.	5.33 (3.54)	4.31 (2.68)	3.59 (1.79)
Low Assoc.	6.49 (4.23)	5.39 (3.58)	4.62 (2.77)
Overall	5.77 (3.62)	4.83 (3.06)	4.05 (2.27)
Free Recall			
High Assoc.	5.63 (4.09)	4.97 (4.19)	3.75 (2.27)
Low Assoc.	6.81 (5.07)	5.18 (4.17)	3.70 (2.48)
Overall	6.04 (4.41)	5.00 (4.03)	3.63 (2.16)

Note. Group means were calculated from participant medians; unit of measurement is seconds.

Table 12

Encoding Strategy Usage Frequency Ratings in Experiment 3

Encoding Strategy	Cued Recall Expectation		Free Recall Expectation	
	M (SD)	*Mdn*	*M (SD)*	*Mdn*
Cue-target Association	3.67 (0.64)	4	1.58 (0.79)[d]	1[d]
Target-target Association	1.78 (0.92)[d]	2[d]	2.76 (1.21)[d]	3[d]
Inter-item Association	1.65 (0.82)[a]	1[a]	1.99 (1.13)[b]	2[b]
Target Focus	2.43 (0.91)[c]	2.5[c]	3.63 (0.79)[d]	4[d]
Mental Imagery	3.00 (1.10)	3	2.88 (1.18)	3
Rote Rehearsal	2.63 (1.12)	3	3.07 (1.09)	3
Verbalization	2.79 (1.24)	3	2.94 (1.26)	4
Intra-item Narrative	2.75 (1.13)	3	2.61 (1.25)[d]	3[d]
Inter-item Narrative	1.98 (1.13)	1.5	2.62 (1.30)	3
Personal Significance	2.67 (1.12)	3	2.45 (1.14)	2
Observation	2.16 (1.08)[c]	2[c]	2.35 (1.13)[c]	2[c]

Note. Rating scale was 1 (*no use*) to 4 (*extensive use*); $N = 84$.
[a]$n = 80$. [b]$n = 81$. [c]$n = 82$. [d]$n = 83$.

Table 13

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

		Cued Recall		Free Recall		Cued vs. Free	
Encoding Strategy	*N*	$\hat{\tau}_b$ *(SD)*	95% CI	$\hat{\tau}_b$ *(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]**
Inter-item 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]
Intra-item Narrative	83	-.06 (.08)	[-.22, .10]	.03 (.08)	[-.13, .20]	.13	[-.34, .16]
Inter-item 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 used $z_{\alpha/2}$ = 1.96 and standard errors calculated as per Woods (2007) using consistent variance estimates from Cliff & Charlin (1991); statistically significant CIs are shown in boldface.

Table 14

Means (and Standard Deviations) of Encoding Strategy Usage Frequency Ratings by Level of Recall Performance Improvement in Experiment 3

	High Improvers			Low Improvers		
Encoding Strategy	Cued Recall	Free Recall	Free - Cued	Cued Recall	Free Recall	Free - Cued
Cue-target Association	3.83 (0.37)	1.44 (0.68)	-2.39 (0.79)	3.53 (0.80)	1.72 (0.93)	-1.81 (1.33)
Target-target Association	1.58 (0.64)	3.00 (1.08)	1.42 (1.30)	2.03 (1.12)	2.67 (1.22)	0.64 (1.67)
Inter-item Association	1.53 (0.64)	1.94 (1.05)	0.42 (1.06)	1.75 (0.98)	1.97 (1.17)	0.22 (1.16)
Target Focus	2.67 (0.82)	3.75 (0.72)	1.08 (1.11)	2.36 (0.95)	3.56 (0.80)	1.19 (1.22)
Mental Imagery	3.08 (1.09)	2.86 (1.23)	-0.22 (1.23)	2.92 (1.11)	2.92 (1.14)	0.00 (0.62)
Rote Rehearsal	2.53 (1.19)	2.89 (1.12)	0.36 (1.03)	2.75 (1.14)	3.33 (1.03)	0.58 (0.79)
Verbalization	2.53 (1.28)	2.64 (1.34)	0.11 (0.84)	2.94 (1.22)	3.22 (1.16)	0.28 (0.56)
Intra-item Narrative	2.78 (1.16)	2.61 (1.28)	-0.17 (1.64)	2.69 (1.15)	2.58 (1.21)	-0.11 (0.91)
Inter-item Narrative	1.92 (1.11)	2.94 (1.27)	1.03 (1.52)	2.11 (1.17)	2.36 (1.25)	0.25 (1.21)
Personal Significance	2.78 (1.06)	2.50 (1.14)	-0.28 (1.15)	2.50 (1.12)	2.39 (1.16)	-0.11 (0.81)
Observation	2.08 (1.09)	2.22 (1.08)	0.14 (0.82)	2.28 (1.07)	2.53 (1.14)	0.25 (0.72)

Note. Rating scale was 1 (*no use*) to 4 (*extensive use*); $n_{high} = 36$; $n_{low} = 36$.

Table 15

Differential Efficacy and Use of Encoding Strategies in Experiments 1-3

	Exp. 1	Exp. 2		Exp. 3	
Encoding Strategy	Use	Efficacy	Use	Efficacy	Use
Cue-target Association	C	C	C	–	C
Target-target Association	~F	F	F	F	F
Inter-item Association		F	–	~F	–
Target Focus	F	F	F	–	F
Mental Imagery					
Rote Rehearsal		–	F	–	F
Verbalization					
Intra-item Narrative					
Inter-item 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.

Figures

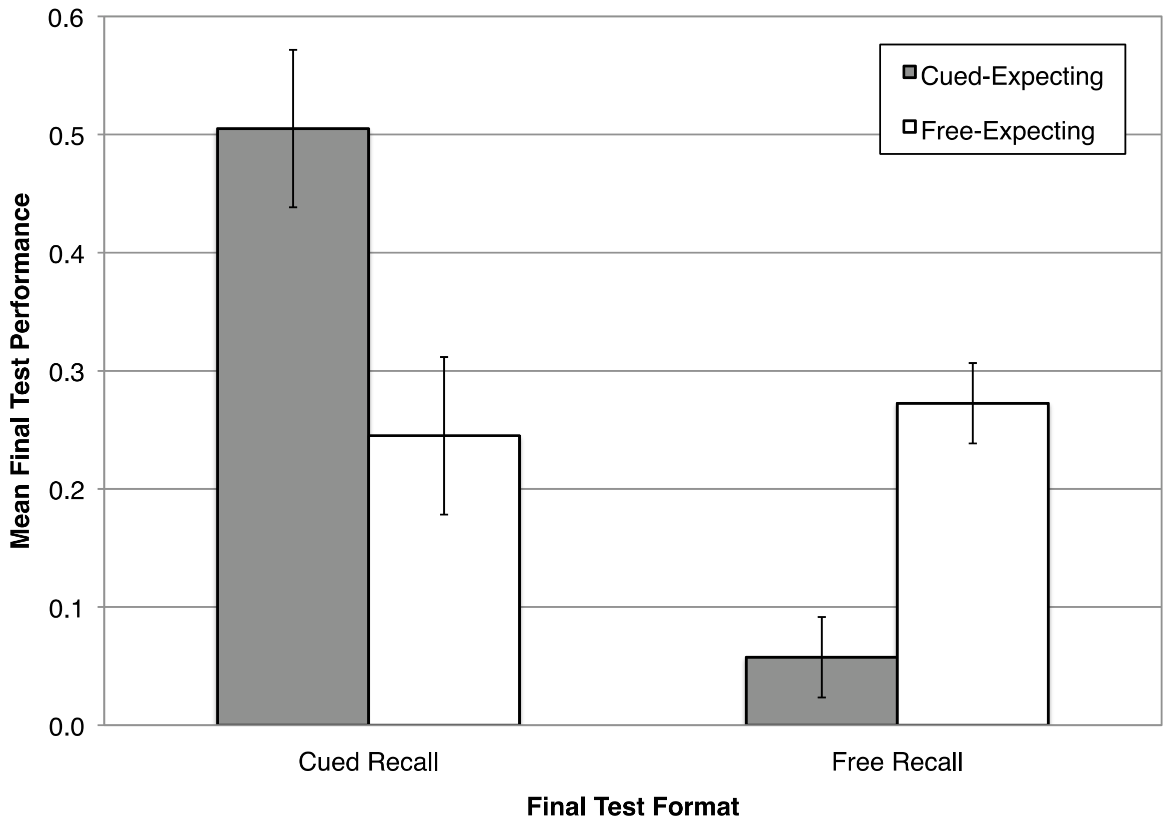

Figure 1. Mean final recall performance as a function of received test format (cued vs. free) and expected test format (cued vs. free) in Experiment 1. Error bars represent the pooled standard errors for comparison of expectancy conditions within each received test format.

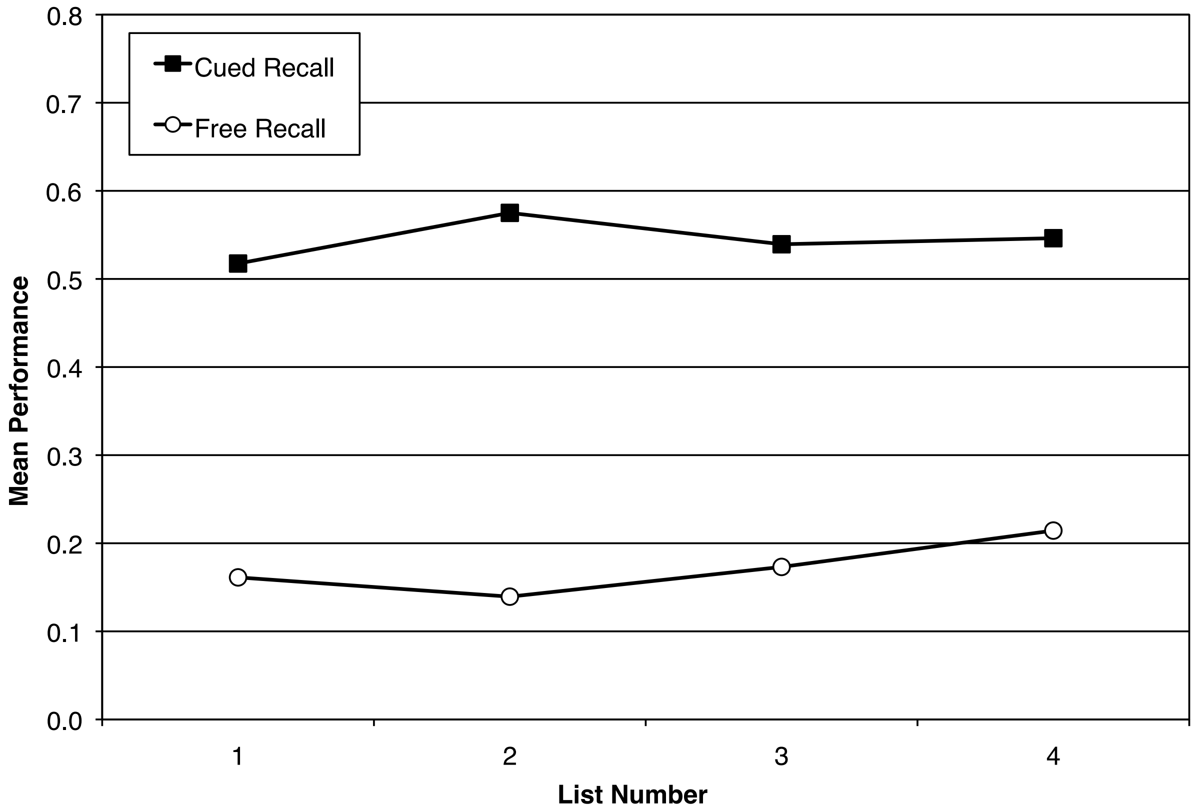

Figure 2. Mean recall performance as a function of list number (1-4) and test format (cued vs. free) in Experiment 1.

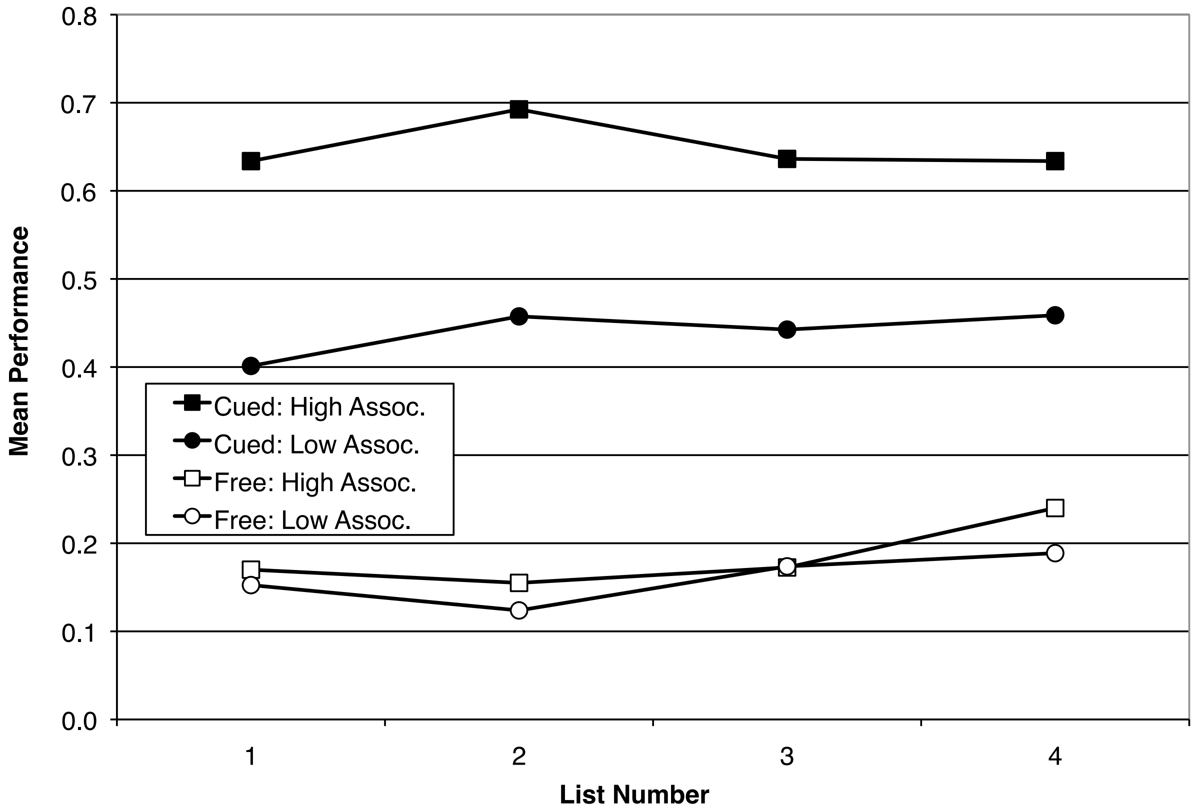

Figure 3. Mean recall performance as a function of list number (1-4), test format (cued vs. free), and associative strength (high vs. low) in Experiment 1.

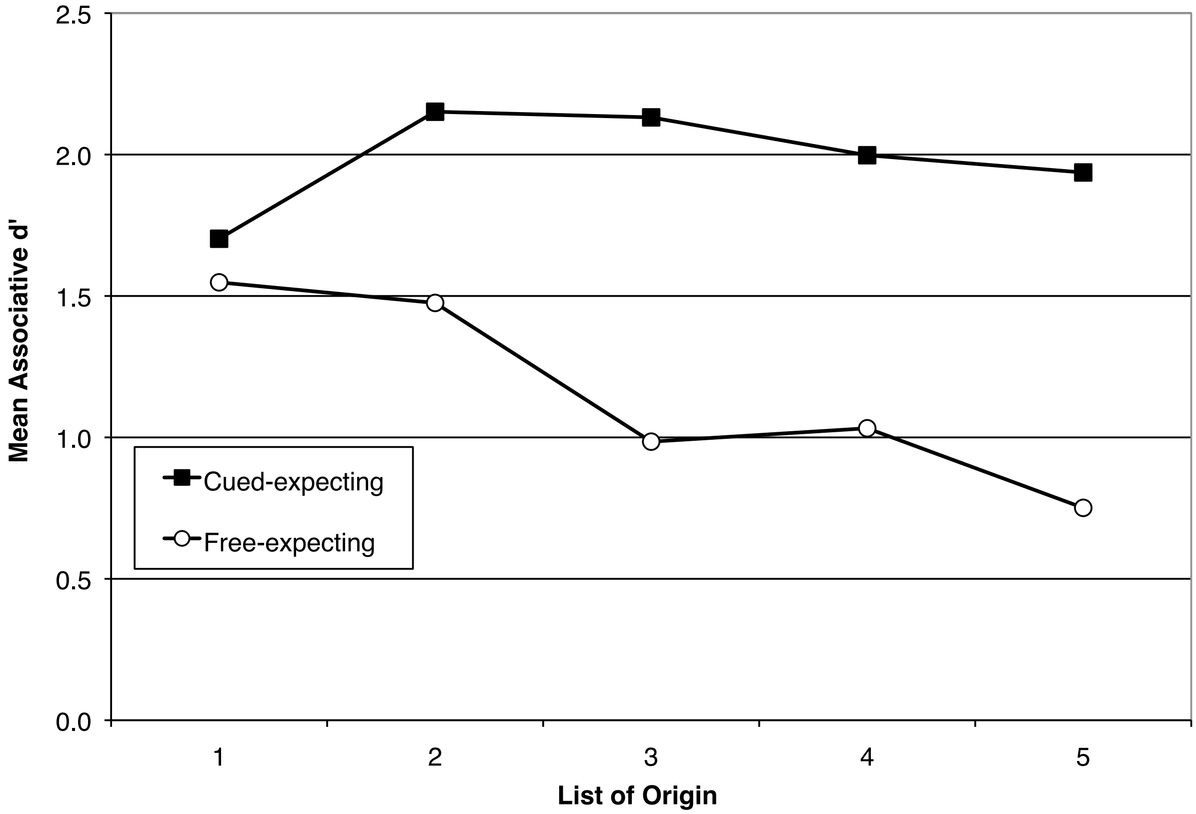

Figure 4. Mean associative recognition performance (d') as a function of test expectancy (cued vs. free) and list of origin of word pairs (1-5) in Experiment 1, for participants receiving their expected test format.

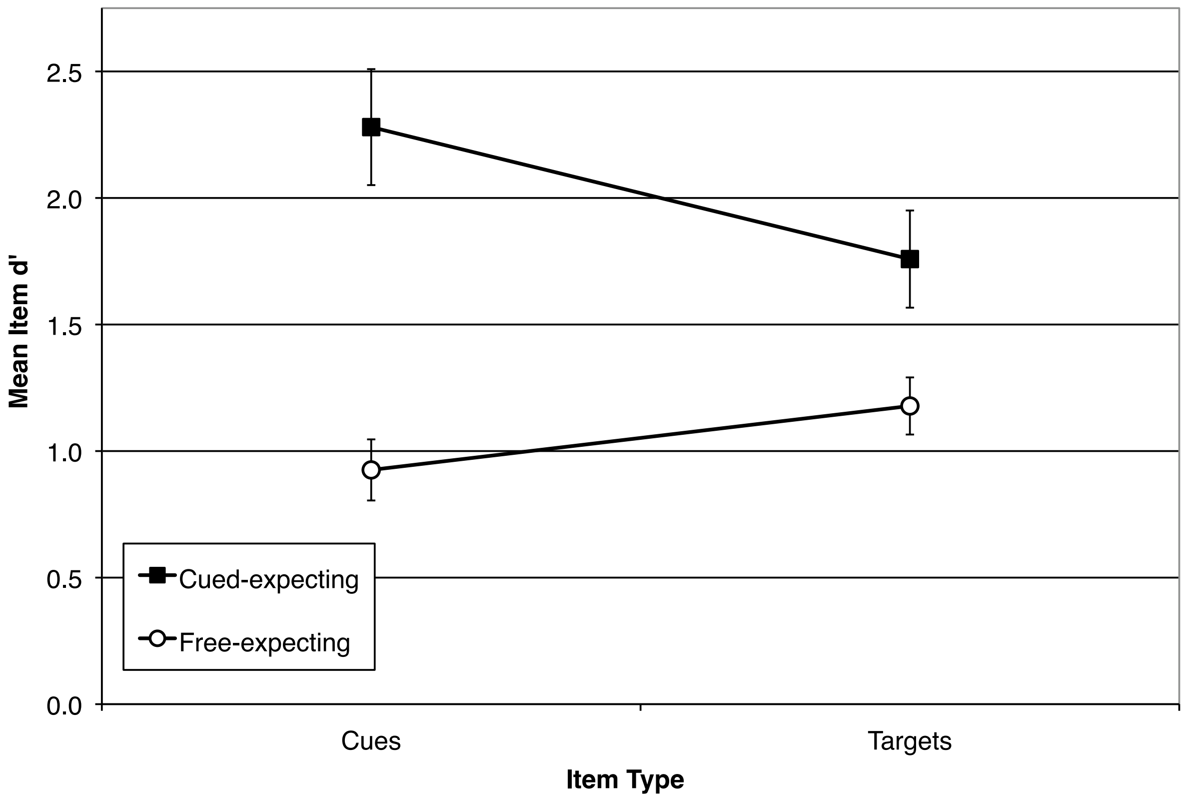

Figure 5. Mean item recognition performance (d’) as a function of test expectancy (cued vs. free) and item type (cues vs. targets) in Experiment 1, for participants receiving their expected test format. Error bars represent standard errors of each cell.

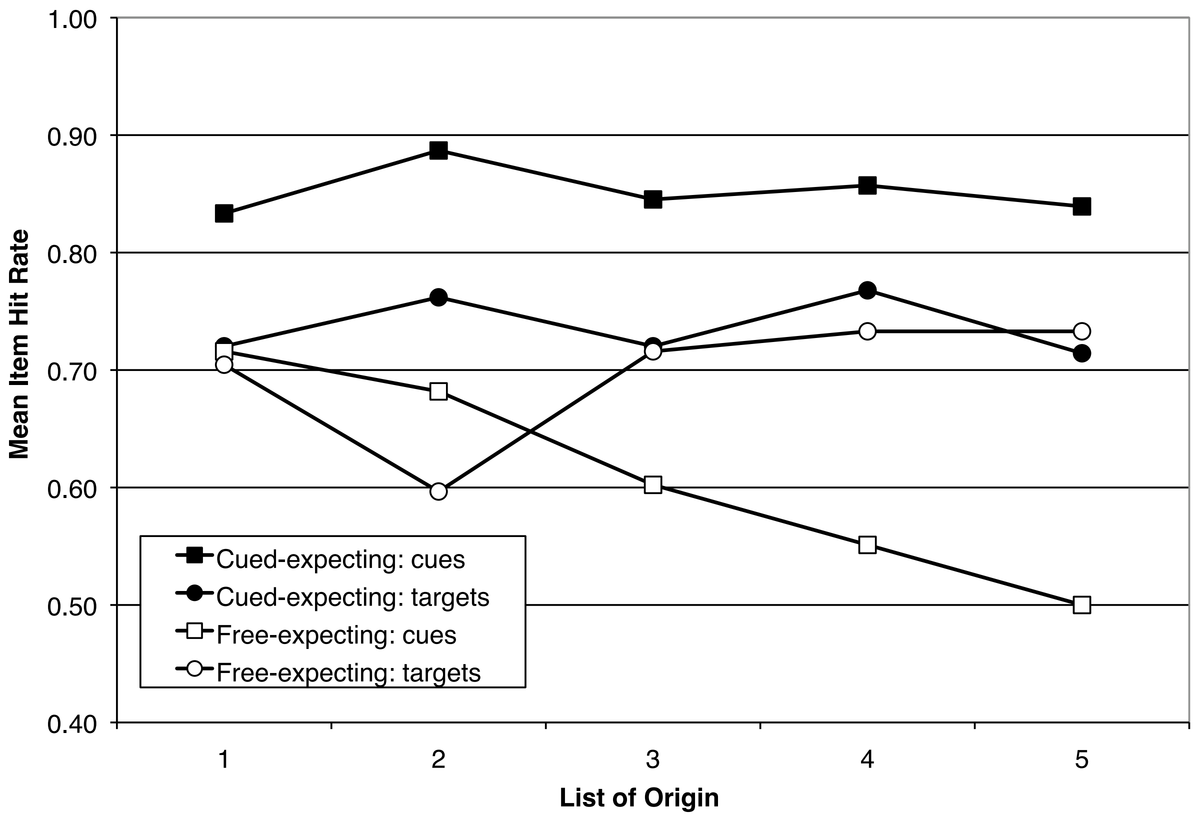

Figure 6. Mean item recognition performance (hit rate) as a function of test expectancy (cued vs. free), item type (cues vs. targets), and list of origin for items (1-5) in Experiment 1, for participants receiving their expected test format.

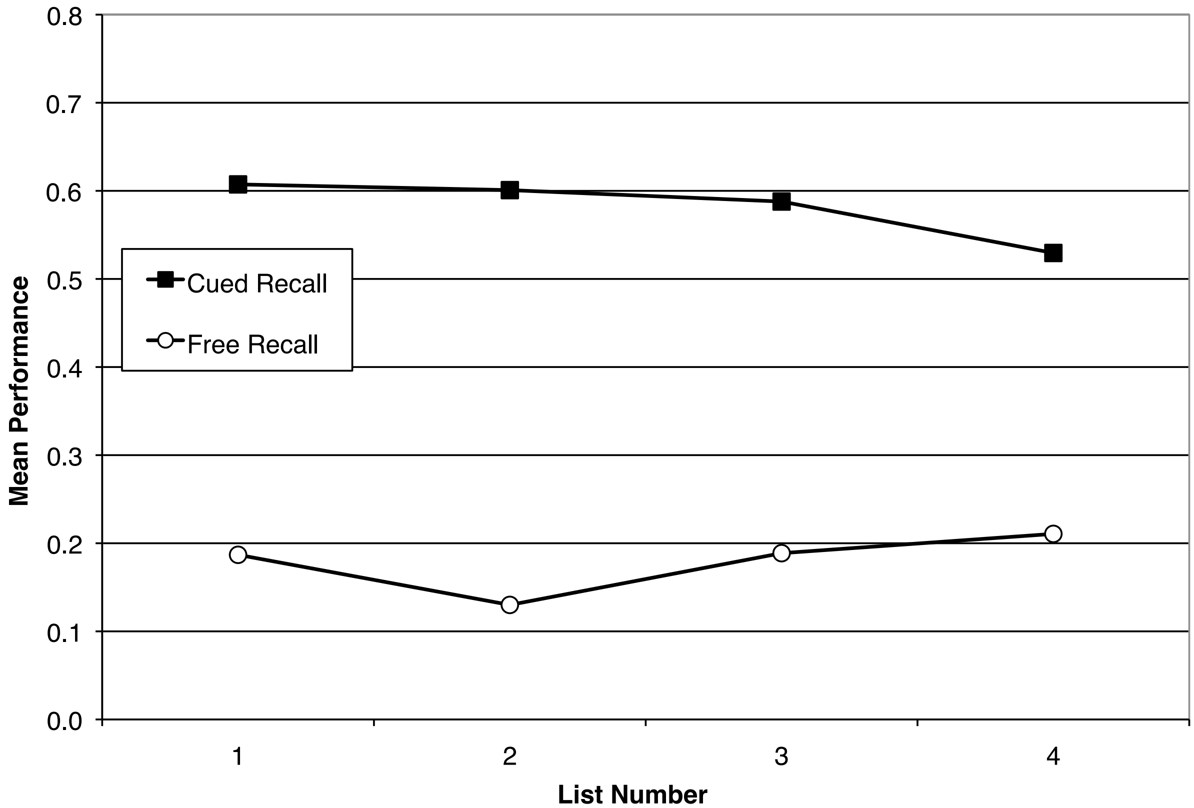

Figure 7. Mean recall performance as a function of list number (1-4) and test format (cued vs. free) in Experiment 2.

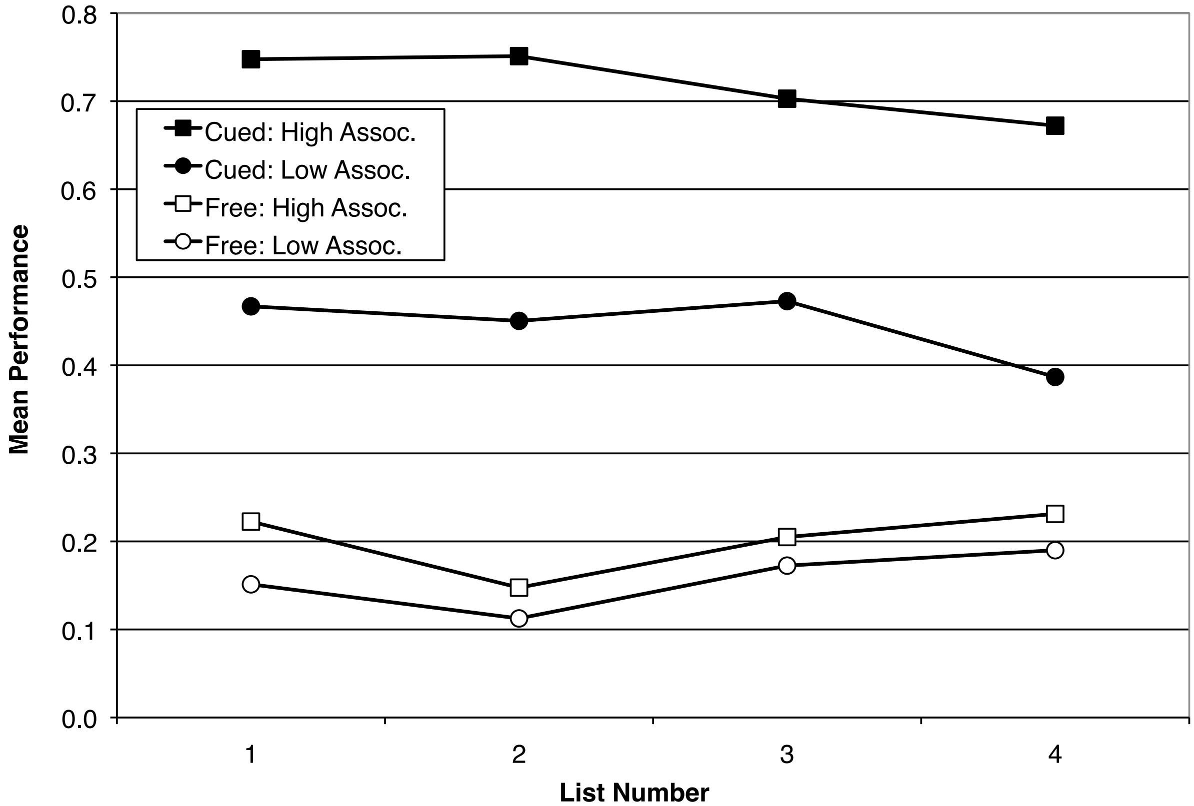

Figure 8. Mean recall performance as a function of list number (1-4), test format (cued vs. free), and associative strength (high vs. low) in Experiment 2.

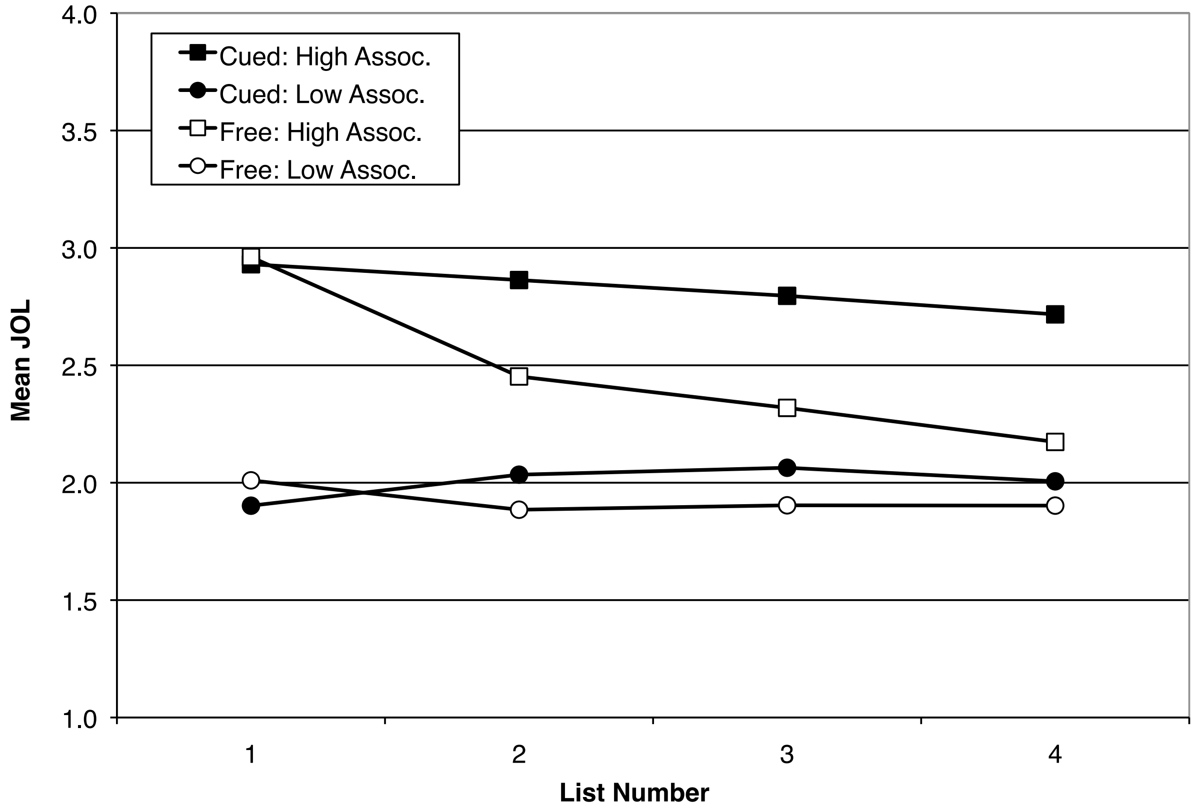

Figure 9. Mean JOLs as a function of list number (1-4), test format (cued vs. free), and associative strength (high vs. low) in Experiment 2.

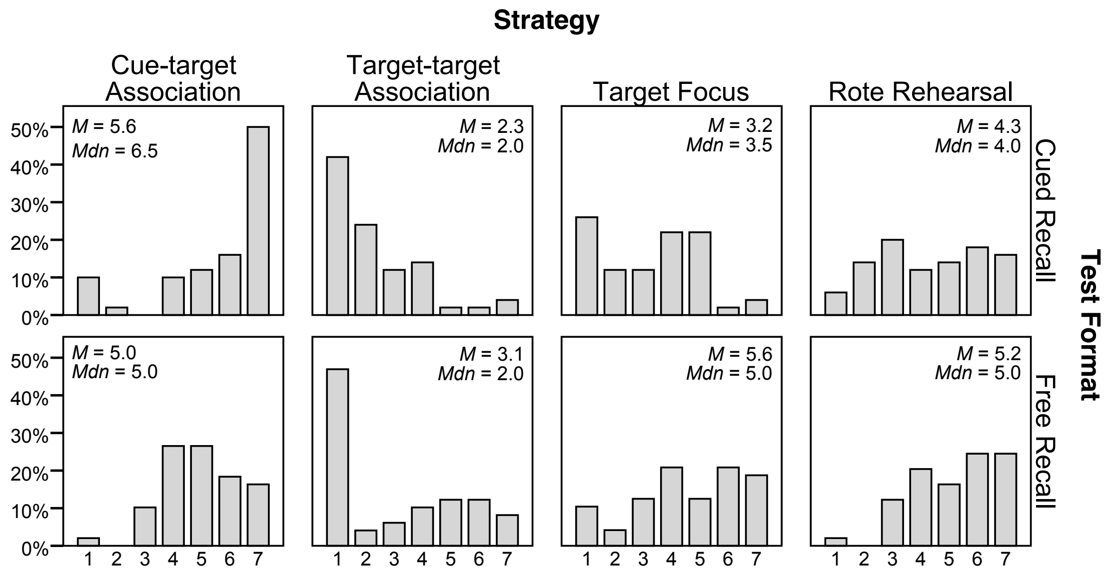

Figure 10. Histograms of usage frequency ratings (1 = *no use*, 7 = *extensive use*) for four encoding strategies as a function of test format (cued vs. free) in Experiment 2.

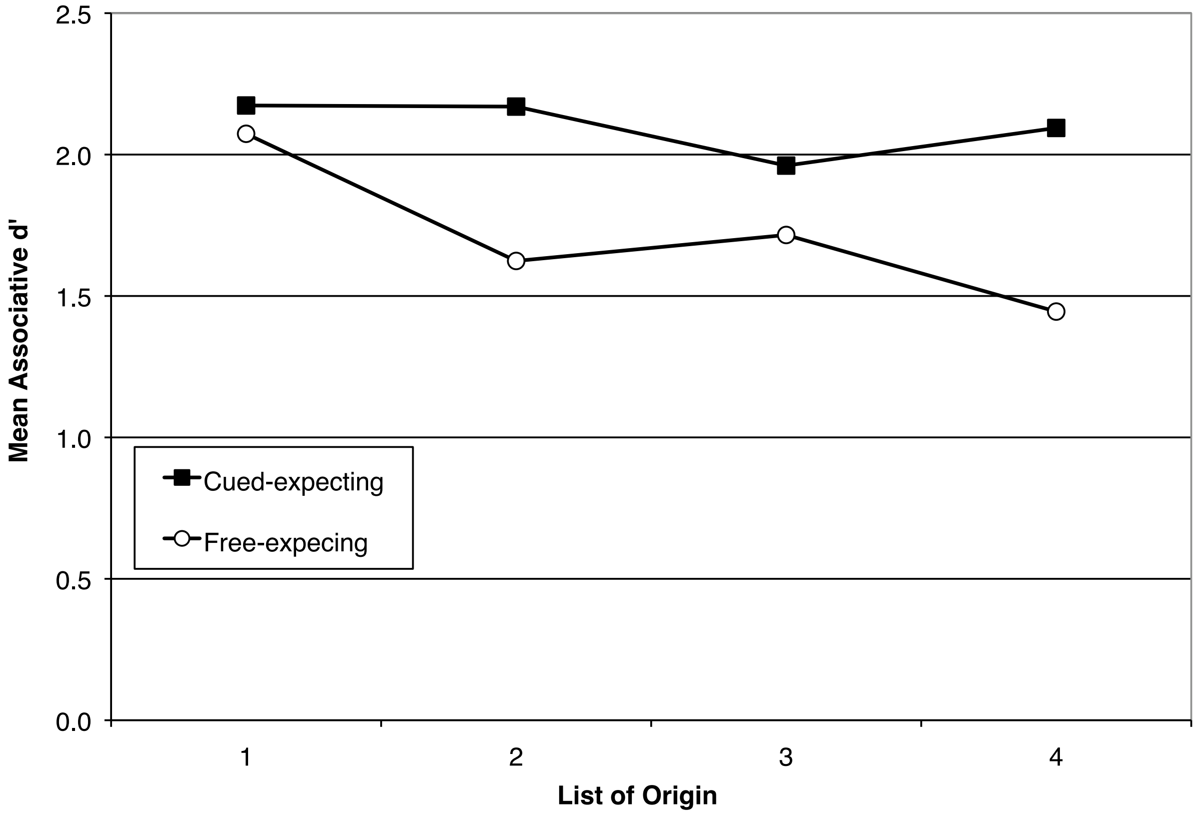

Figure 11. Mean associative recognition performance (d') as a function of test expectancy (cued vs. free) and list of origin of word pairs (1-5) in Experiment 2.

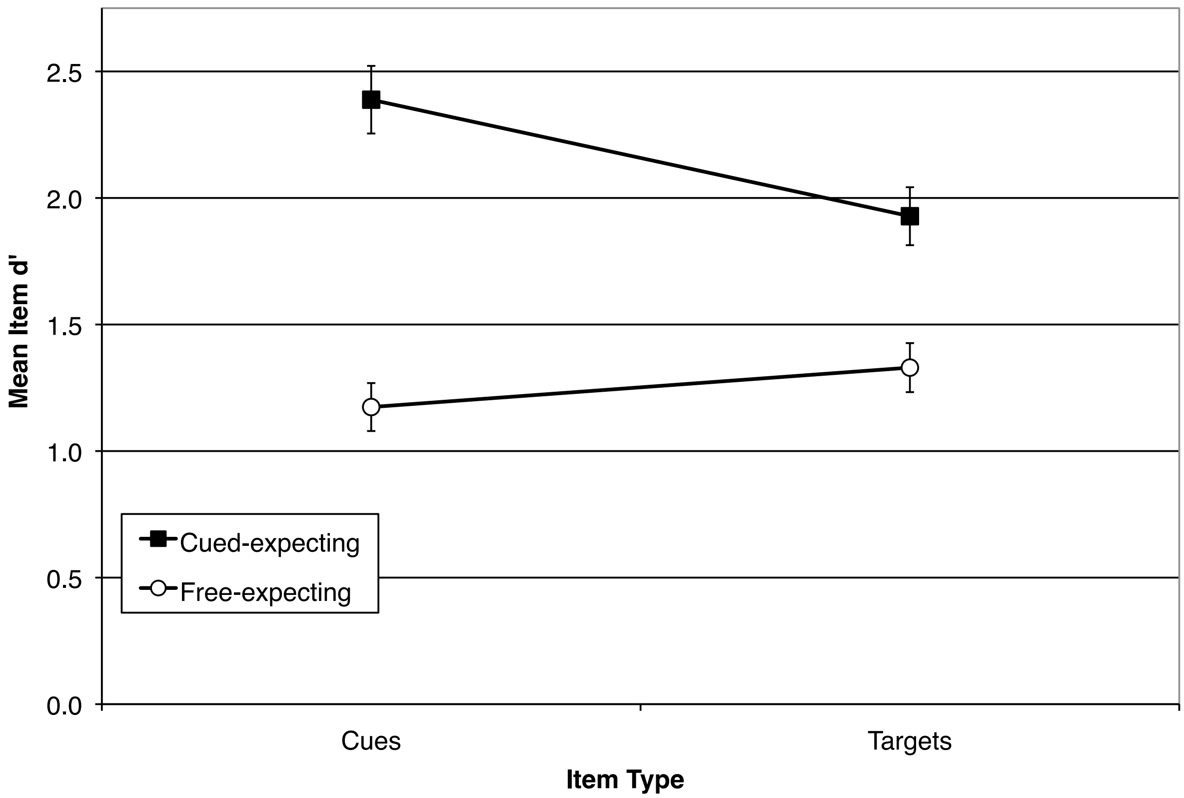

Figure 12. Mean item recognition performance (d') as a function of test expectancy (cued vs. free) and item type (cues vs. targets) in Experiment 2. Error bars represent standard errors of each cell.

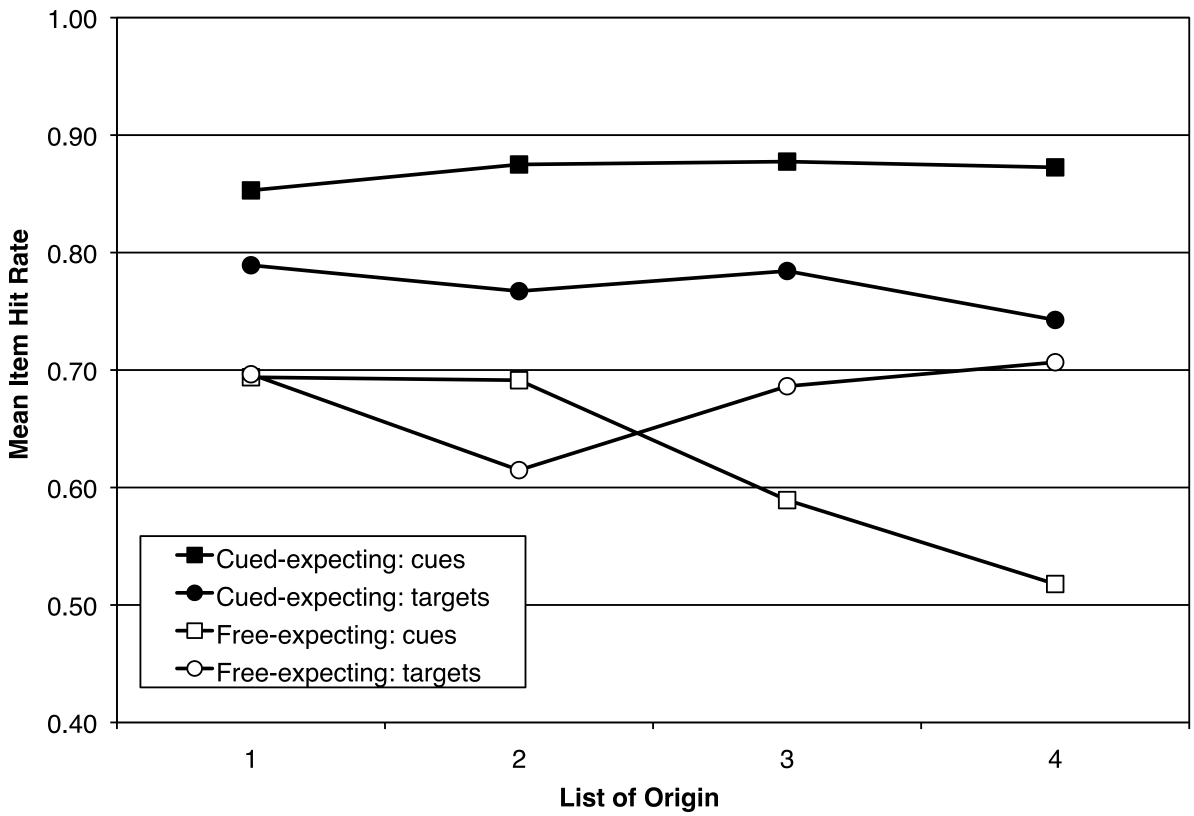

Figure 13. Mean item recognition performance (hit rate) as a function of test expectancy (cued vs. free), item type (cues vs. targets), and list of origin for items (1-5) in Experiment 2.

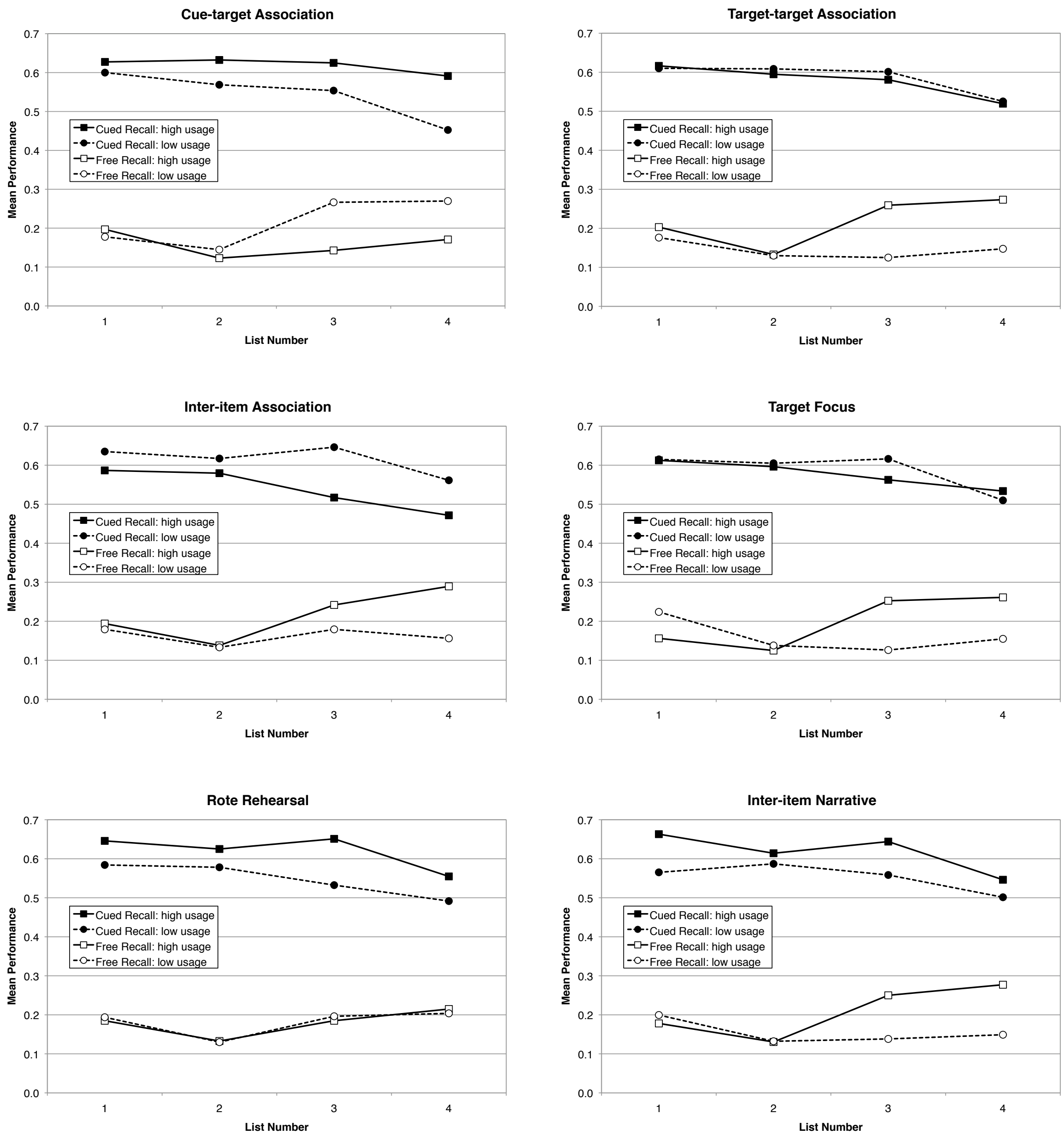

Figure 14. Mean recall performance as a function of list number (1-4), test format (cued vs. free), and usage (high vs. low) of six encoding strategies, in Experiment 2.

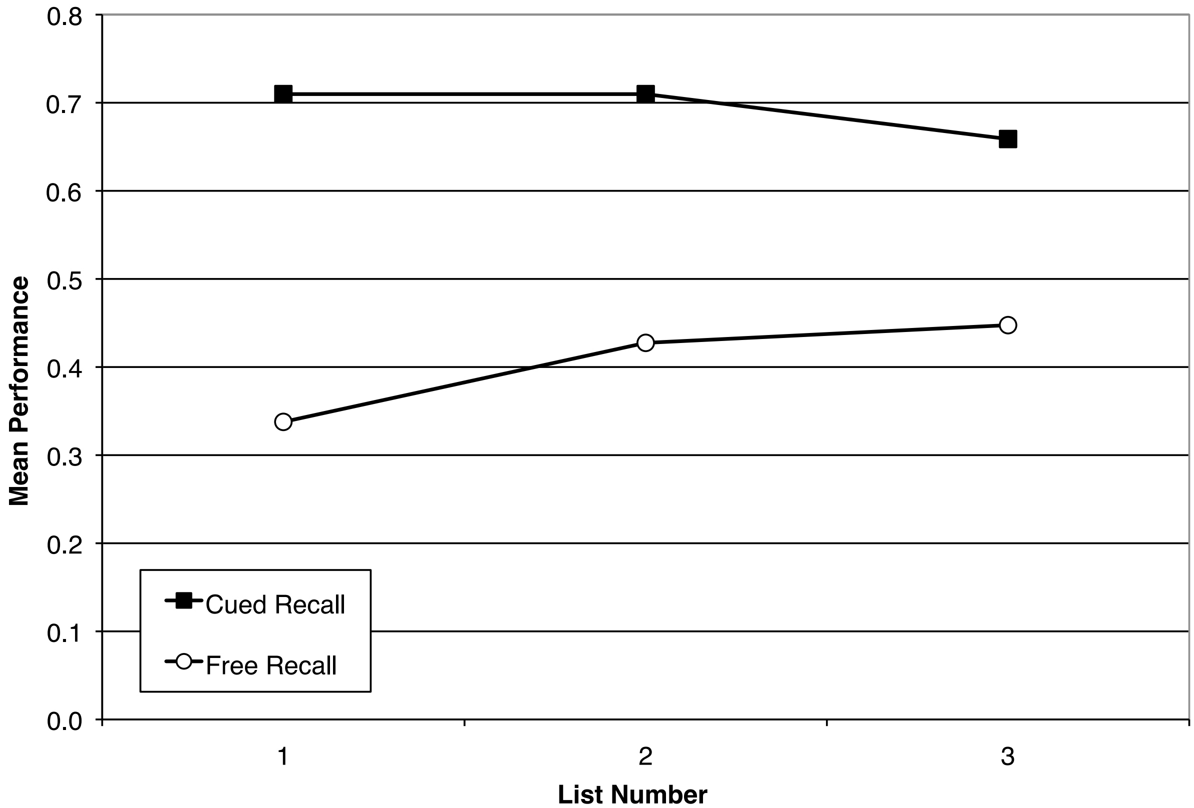

Figure 15. Mean recall performance as a function of list number (1-3) and test format (cued vs. free) in Experiment 3.

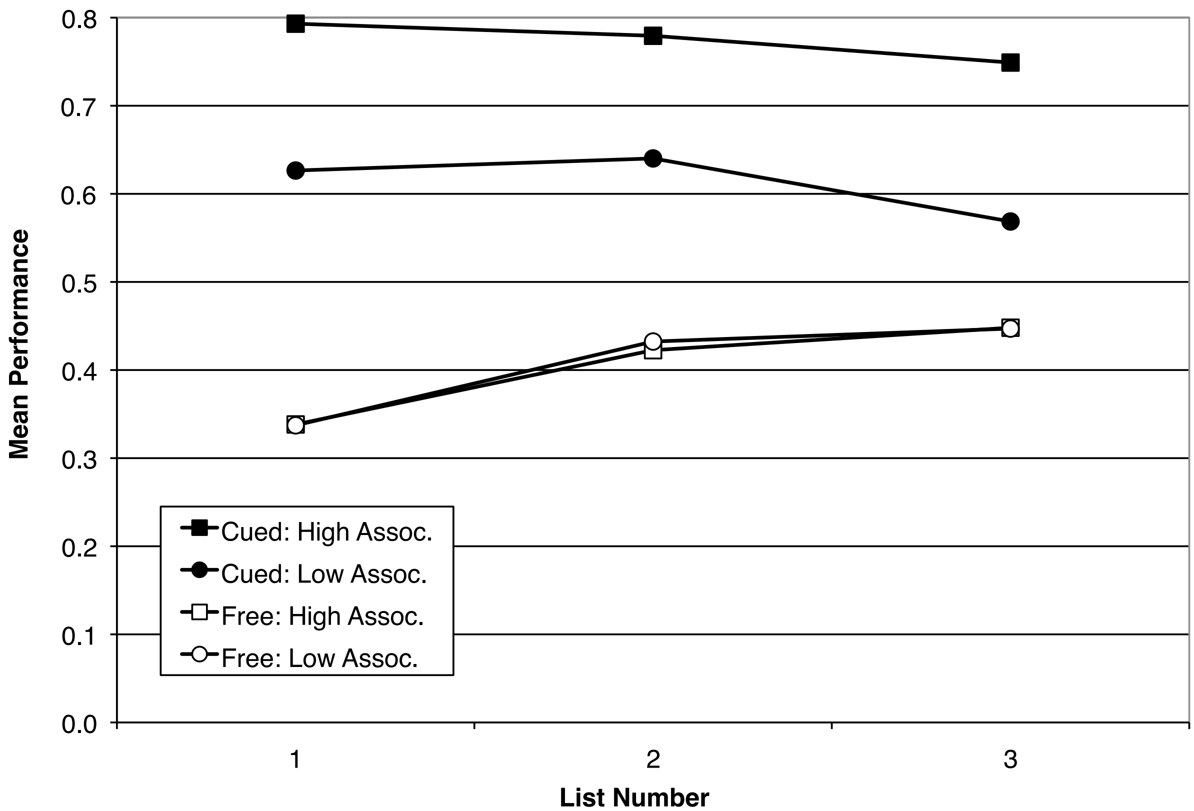

Figure 16. Mean recall performance as a function of list number (1-3), test format (cued vs. free), and associative strength (high vs. low) in Experiment 3.

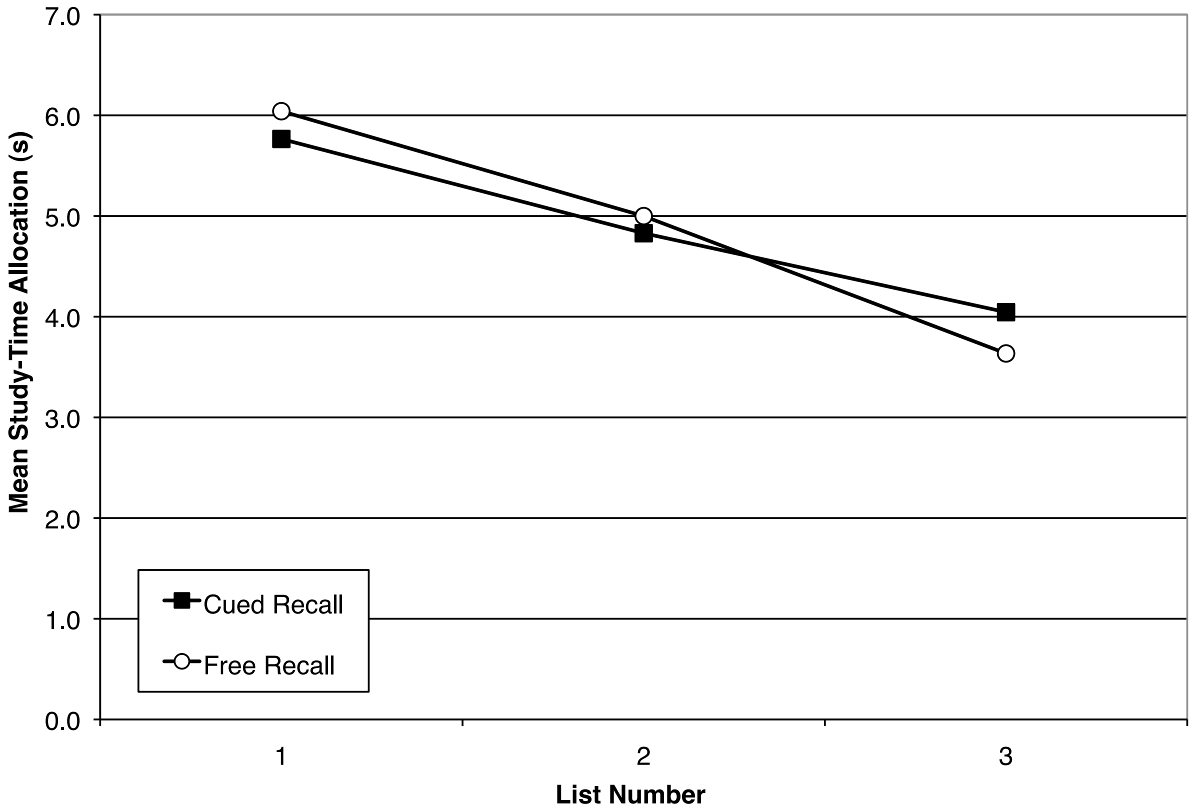

Figure 17. Mean of participant median study-time allocation (in seconds) as a function of list number (1-3) and test format (cued vs. free) in Experiment 3.

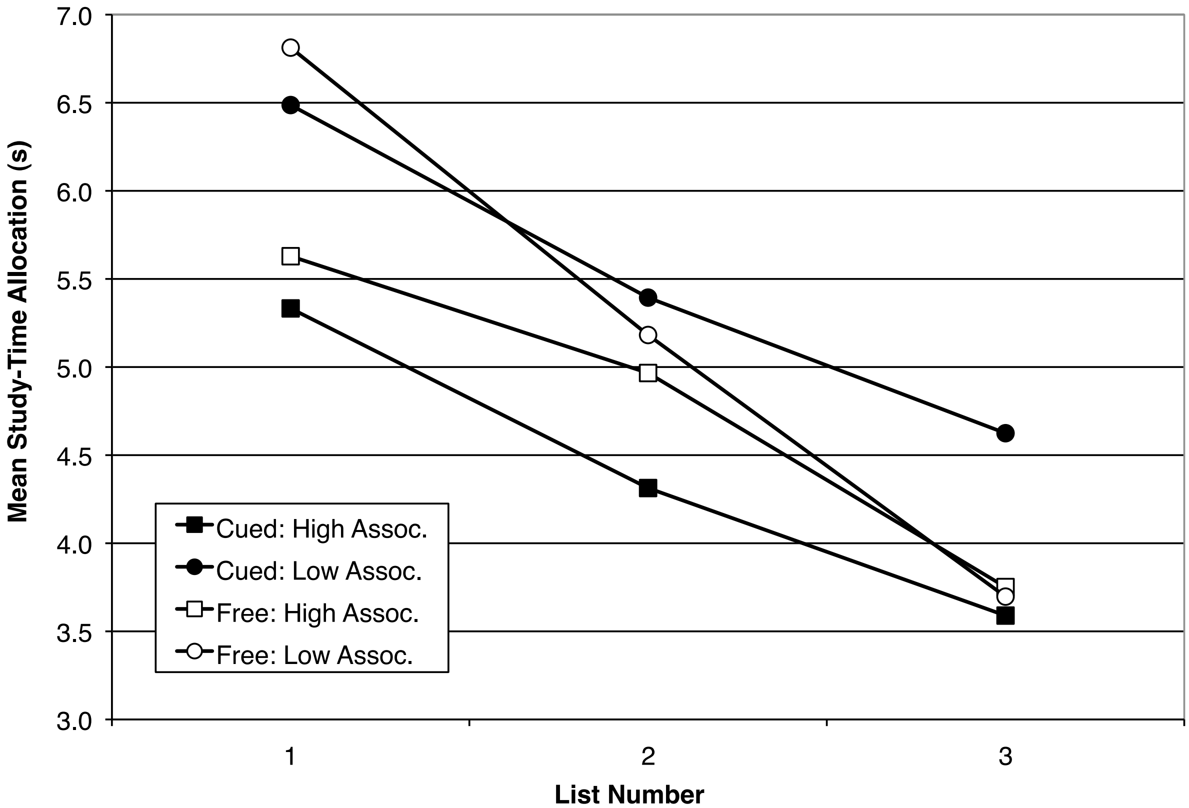

Figure 18. Mean of participant median study-time allocation (in seconds) as a function of list number (1-3), test format (cued vs. free), and associative strength (high vs. low) in Experiment 3.

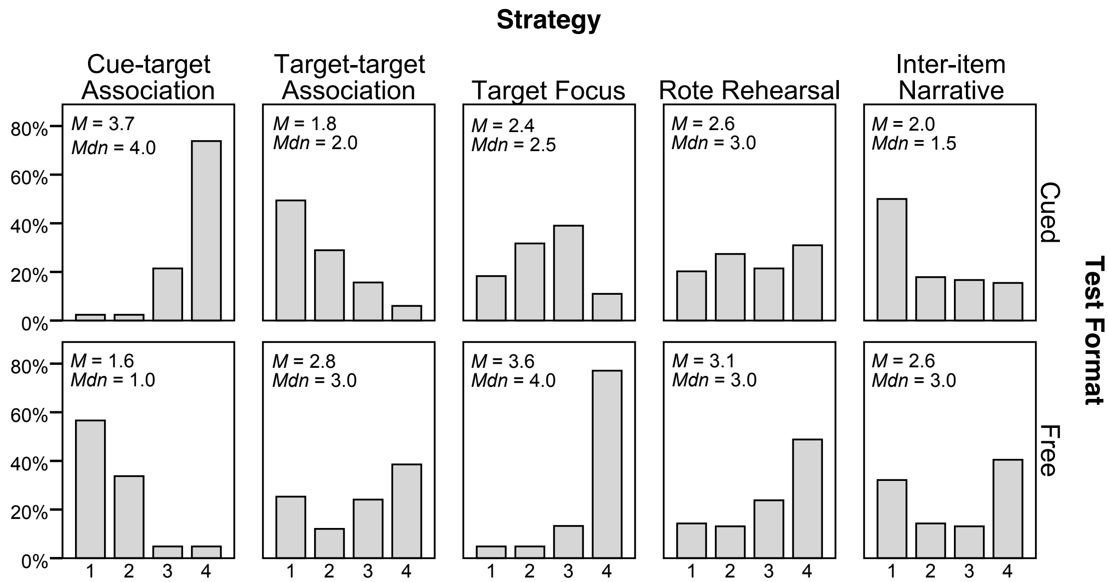

Figure 19. Histograms of usage frequency ratings (1 = *no use*, 4 = *extensive use*) for five encoding strategies as a function of test format (cued vs. free) in Experiment 3.

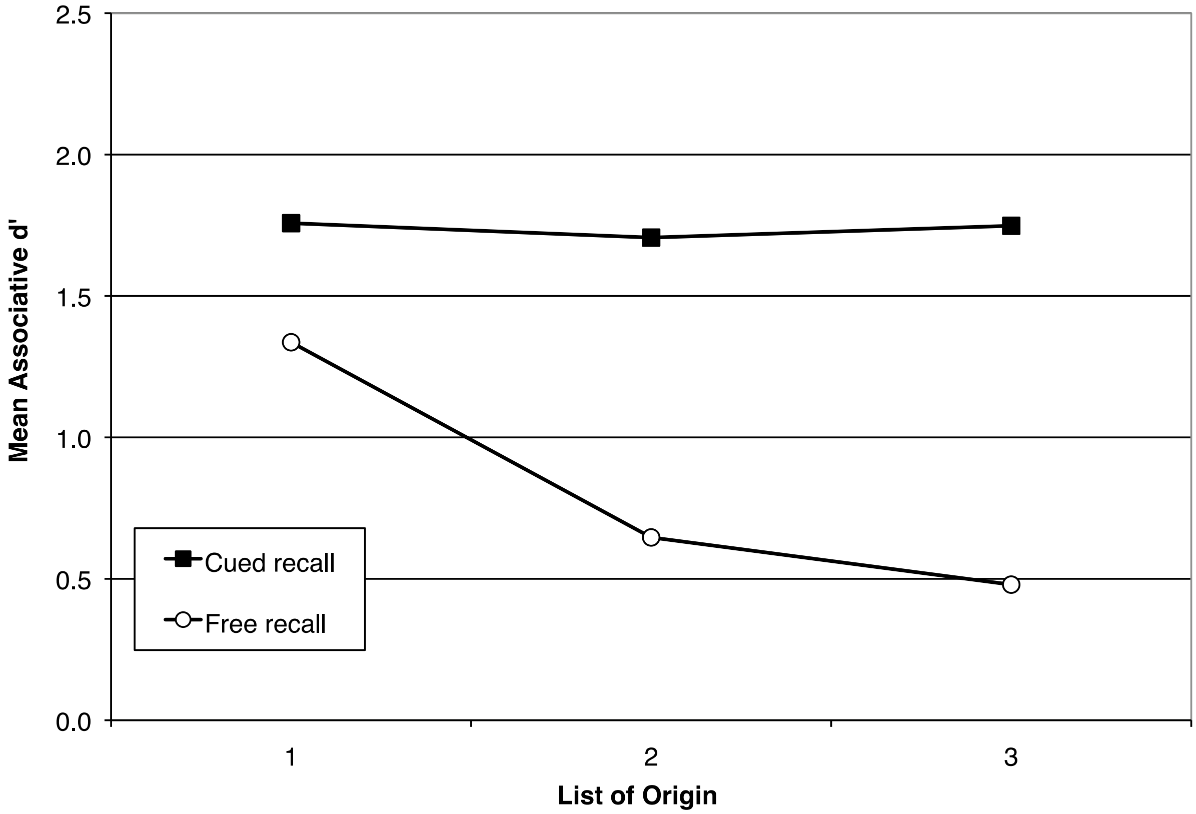

Figure 20. Mean associative recognition performance (d') as a function of test format (cued vs. free) and list of origin of word pairs (1-3) in Experiment 3.

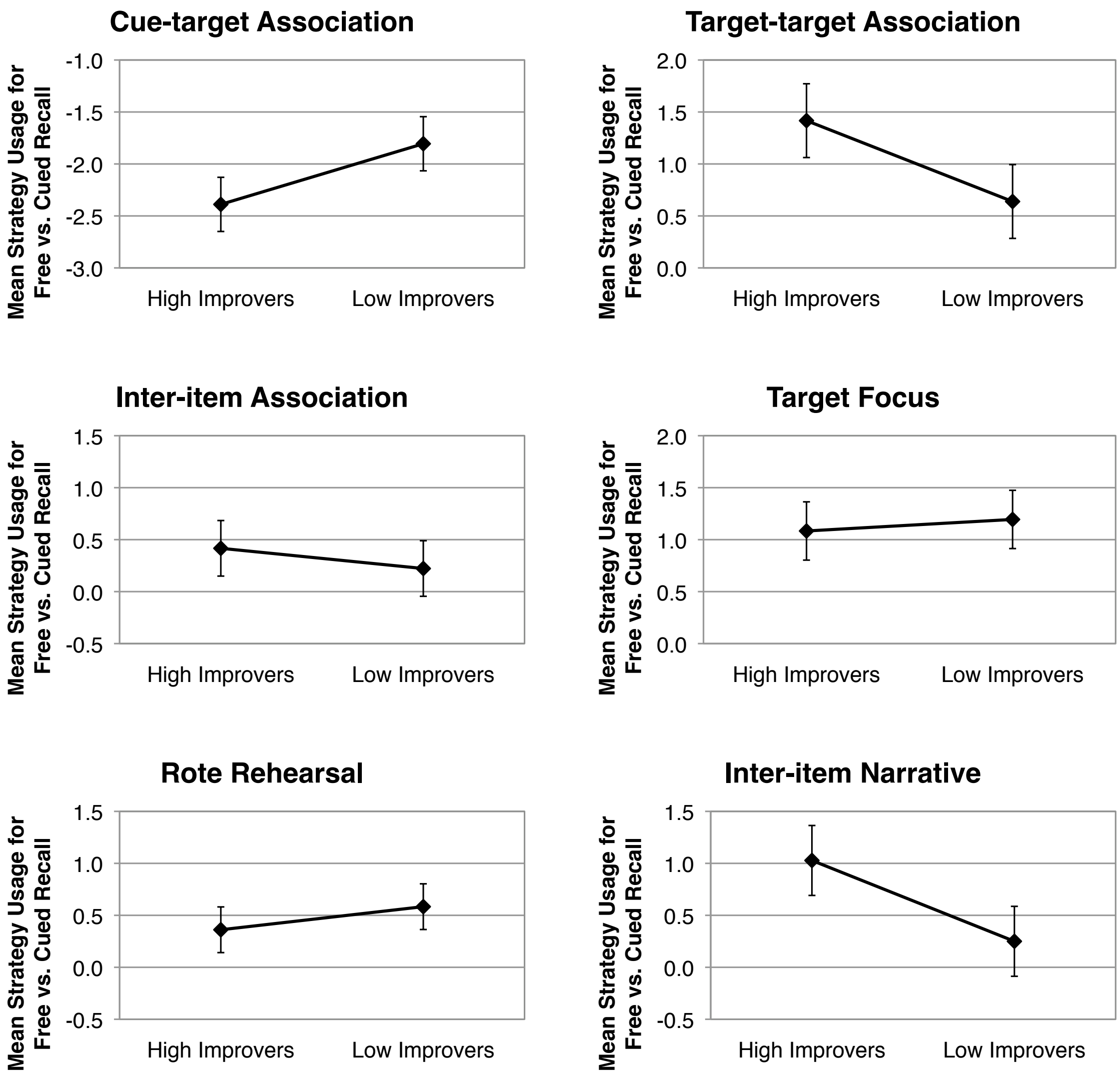

Figure 21. Mean difference in usage frequency rating for free versus cued recall, for high improvers versus low improvers, for six encoding strategies, in Experiment 3. Error bars represent the pooled standard error for comparison of improvement groups.

References

Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory & Cognition, 2*(3), 406-412.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory, 6*(5), 576-587.

Benjamin, A. S. (2008). 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: Skill and Strategy in Memory Use* (Vol. 48; pp.175-223). London: Academic Press.

Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language, 55*(1), 126-137.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55-68.

Bhatarah, P., Ward, G., & Tan, L. (2008). Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. *Memory & Cognition, 36*(1), 20-34.

Biggs, J. B. (1985). The role of metalearning in study processes. *British Journal of Educational Psychology, 55*(3), 185-212.

Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*(1), 133-149.

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*(2), 207-211.

Blaxton, T. A., 1989. Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition 15*, 657-668.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*(4), 433-436.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Carey, S. T., & Lockhart, R. S. (1973). Encoding differences in recognition and recall. *Memory & cognition, 1*(3), 297-300.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Erlbaum.

Cliff, N., & Charlin, V. (1991). Variances and covariances of Kendall's tau and their estimation. *Multivariate Behavioral Research, 26*(4), 693–707.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33*(4), 497-505.

Connor, J. M. (1977). Effects of organization and expectancy on recall and recognition. *Memory & cognition, 5*(3), 315-318.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*(6), 671-684.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing in free recall. *Journal of Memory and Language, 52*, 120-130.

deWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & cognition, 32*(6), 945-955.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & cognition, 29*(2), 247-253.

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.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

d'Ydewalle, G. (1981). Test expectancy effects in free recall and recognition. *Journal of General Psychology, 105*(Pt 2), 173-195.

d'Ydewalle, G. (1982). Psychophysical methods to unravel test expectancy effects. *Studia Psychologica, 24*(3-4), 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*(1), 55-67.

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of experimental psychology, 68*(1), 58-63.

Elliott, E. S., & Dweck, C. S. (1988) Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54,* 5-12.

Ericsson, K. A., & Chase, W. G. (1982). Exceptional memory. *American Scientist, 70*(6), 607-615.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review, 102*(2), 211-245.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (rev. ed.)*. Cambridge, MA, US: The MIT Press.

Feldt, R. C. (1990). Test expectancy and performance on factual and higher-level questions. *Contemporary educational psychology, 15*(3), 212-223.

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

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

Freund, R. D., Brelsford, J. W., Jr., & Atkinson, R. C. (1969). Recognition vs. recall: Storage or retrieval differences? *Quarterly Journal of Experimental Psychology, 21*(3), 214-224.

Glass, L. A., Caluse, C. B., & Kreiner, D. S. (2007). Effect of test-expectancy and word bank availability on test performance. *College Student Journal, 41*(2), 342-351.

Hall, J. W., Grossman, L. R., & Elwood, K. D. (1976). Differences in encoding for free recall vs. recognition. *Memory & cognition, 4*(5), 507-513.

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: Advances in research and theory* (45), 215-251. San Diego, CA, US: Elsevier Academic Press.

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

Hertzog, C., Price, J., & Dunlosky, J. (2008). How is knowledge generated about memory encoding strategy effectiveness? *Learning and Individual Differences, 18*(4), 430-445.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*(4), 421-445.

Jacoby, L. L. (1973). Test appropriate strategies in retention of categorized lists. *Journal of Verbal Learning & Verbal Behavior, 12*(6), 675-682.

Jacoby, L. L., 1983. Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal behavior 22*, 485-508.

Kardash, C. M., & Kroeker, T. L. (1989). Effects of time of review and test expectancy on learning from text. *Contemporary Educational Psychology, 14*, 323-335.

Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language. Special Issue: Illusions of memory, 35*(2), 157-175.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 187-194.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 609-622.

Kucera, H., & Francs, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

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*(10), 363-365.

Lewis, K., & Wilding, J. M. (1981). Influences of test expectations on memory-processing strategies. *Current Psychological Research, 1*(1), 61-74.

Leonard, J. M., & Whitten, W. B. (1983). Information stored when expecting recall or recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(3), 440-455.

Loftus, G. R. (1971). Comparison of recognition and recall in a continuous memory task. *Journal of experimental psychology, 91*(2), 220-226.

Lovelace, E. A. (1973). *Effects of anticipated form of testing on learning* (Final report, Project 2-C-019, Grant OEG-3-72-0033). U.S. Office of Education. (ERIC Document Reproduction Service No. ED085420)

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research, 61*(1), 94-106.

Magnussen, S., Andersson, J., Cornoldi, C., De Beni, R., Endestad, T., Goodman, G. S., et al. (2006). What people believe about memory. *Memory, 14*(5), 595-613.

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*(2), 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, New Jersey: Lawrence Erlbaum Associates.

May, R. B., & Sande, G. N. (1982). Encoding expectancies and word frequency in recall and recognition. *American Journal of Psychology, 95*(3), 485-495.

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary educational psychology, 19*(2), 230-248.

McDaniel, M. A., & Kearney, E. M. (1984). Optimal learning strategies and their spontaneous use: The importance of task-appropriate processing. *Memory & Cognition, 12*(4), 361-373.

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*(9), 641-661.

Meyer, G. (1936). The effect of recall and recognition on the examination set in classroom situations. *Journal of educational psychology, 27*(2), 81-99.

Minnaert, A. E. (2003). The moderator effect of test anxiety in the relationship between test expectancy and the retention of prose. *Psychological reports, 93*(3), 961-971.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*(5), 519-533.

Murayama, K. (2005). Exploring the mechanism of test-expectancy effects on strategy change. *Japanese Journal of Educational Psychology, 53*(2), 172-184.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482-488.

Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & cognition, 9*(3), 283-300.

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

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

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*(2), 179-189.

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

Peeck, J., Van Dam, G., & de Jong, J. (1978). Test expectancy and encoding of pictures and words. *Perceptual and motor skills, 46*(1), 249-250.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation*. (pp. 451-502). San Diego, CA, US: Academic Press.

Postman, L. (1964). Studies of learning to learn: II. Changes in transfer as a function of practice. *Journal of Verbal Learning & Verbal Behavior, 3*(5), 437-447.

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: Academic Press.

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*(6), 683-689.

Rickards, J. P., & Friedman, F. (1978). The encoding versus the external storage hypothesis in note taking. *Contemporary Educational Psychology, 3*(2), 136-143.

Roediger, H. L., III. (1980). The effectiveness of four mnemonics in ordering recall. *Journal of Experimental Psychology: Human Learning & Memory, 6*(5), 558-567.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. Chapter 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*(17), 1-11.

Schmidt, S. R. (1988). Test expectancy and individual-item versus relational processing. *American Journal of Psychology, 101*(1), 59-71.

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. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 601-604.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221.

Son, L. K., & Sethi, R. (2006). Metacognitive control and optimal learning. *Cognitive Science, 30*(4), 759-774.

Staresina, B. P., Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience, 26*(36), 9162-9172.

Terry, P. W. (1933). How students review for objective and essay tests. *The Elementary School Journal, 33*(8), 592-603.

Terry, P. W. (1934). How students study for three types of objective tests. *Journal of Educational Research, 27*(5), 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*(4), 901-918.

Thiede, K. T., & 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, 24*, 1024-1037.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning & Verbal Behavior, 5*(2), 193-197.

Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive psychology, 5*(3), 275-287.

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

von Wright, J. (1977). On the development of encoding in anticipation of various tests of retention. *Scandinavian journal of psychology, 18*(2), 116-120.

von Wright, J., & Meretoja, M. (1975). Encoding in anticipation of various tests of retention. *Scandinavian journal of psychology, 16*(2), 108-112.

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.

Weinstein, C. E., & Palmer, D. R. (2002). *Learning and study strategies inventory (LASSI): User's manual* (2nd ed.). Clearwater, FL: H & H Publishing.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning & Verbal Behavior, 2*(5-6), 440-445.

Winne, P. H. (1995). Self-regulation is ubiquitous but its forms vary with knowledge. *Educational Psychologist. Special Issue: Current issues in research on self-regulated learning: A discussion with commentaries, 30*(4), 223-228.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives (2nd ed.).* (pp. 153-189). Mahwah, NJ, US: Erlbaum.

Winne, P. H. (2005). A perspective on state-of-the-art research on self-regulated learning. *Instructional Science, 33*(5-6), 559-565.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice*. (pp. 277-304). Mahwah, NJ, US: Erlbaum.

Wnek, I., & Read, J. D. (1980). Recall and recognition encoding differences for low- and high-imagery words. *Perceptual and Motor Skills, 50*, 391-394.

Woods, C. M. (2007). Confidence intervals for gamma-family measures of ordinal association. *Psychological Methods, 12*(2), 185–204.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology, 81*(3), 329-339.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64-72.

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	...towards 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 outloud 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 completly 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 Questionnaire in Experiments 2 and 3.

Strategy Label	Full Text Used in Questionnaire
Cue-target association	Made associations between the left-hand and right-hand word in a pair.
Target-target association	Made associations between the right-hand words across multiple pairs.
Inter-item 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.
Intra-item narrative	Used a single pair or word in a sentence, phrase, or story.
Inter-item 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.