Title
Metacognitive judgments and performance

Author
Townsend, Corrine

Publication Date
2011-07-18

UNIVERSITY OF CALIFORNIA

Merced

Metacognitive Judgments and Performance

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Cognitive and Information Sciences

Corinne Townsend

June 20, 2011

1

Table of Contents

ABSTRACT

The question of how students manage and allocate their study time is a complex problem, consisting of decisions regarding switching between material, stopping studying, deciding what to prioritize, how long to study given material, and what learning goals to set. The first project detailed investigates switching decisions, investigating how students choose to switch between lists. Several experiments investigate the effects of self-efficacy on metacognitive judgments and study behaviors such as study time, study strategies, and goal setting and achievement. The third project investigates metacognitive framing, a factor that may influence metacognitive judgments, and potentially, study behaviors. The last project examines the influence of different kinds of study scenarios on metacognitive judgments, specifically, repeated testing, spaced restudy, and massed study. Together, these lines of work show evidence of how metacognitive judgments are influenced, how students choose to switch between materials, and how students enact study strategies to achieve learning goals.

Metacognitive Judgments and Performance

How students choose to make decisions about study time is a complex problem. First, they have to decide what class to work on, and then, what to focus on for that particular class. Judgments have to be made about how much is learned, and if progress is being made. Strategies must be chosen, and at some point a decision must be made about whether to move on to a different strategy or different topic. There are a multitude of different factors that may impact a student's decisions; there are motivational factors such as goals, and the accuracy of different metacognitive judgments that are made. If motivation is low, and the student sets a low performance goal, they will probably do poorly regardless of metacognitive accuracy. On the other hand, if a high goal is set, metacognitive accuracy is important – underconfidence may cause students to persist unnecessarily at the expense of free time or other material, while overconfidence may lead to stopping before the goal is likely to be met.

This literature review will initially focus on material relevant to stopping decisions, which includes judgments of improvement and attempts to reach study goals. After a brief description of various metacognitive judgments thought to be involved in decisions about study time allocation, a new judgment (of improvement) will be presented, explaining why we might expect such a judgment to be used, as well as why it would be beneficial. Next, literature concerning students' performance under assigned study goals will be discussed; specifically, how goals are presumed to operate as stopping rules, and what factors influence adherence to these goals. After discussing judgments of improvement and goal performance, different models of study time allocation will be compared, including a general framework (Nelson & Narens, 1990), discrepancy reduction (Dunlosky & Hertzog, 1998), proximal learning theory (Metcalfe, 2002), hierarchical (Thiede & Dunlosky, 1999), and a social cognitive model (Zimmerman, 1990). The material is organized in this way in order to first describe two possible (metacognitive) means by which students may decide to stop studying material, discussing literature relevant to these two mechanisms, including how adaptive it may be to rely on these judgments and how likely it is that students rely on them. Different models of study time allocation will then be compared, and relevant evidence will be discussed that supports and/or detracts from the various models. A new perspective will be proposed that incorporates certain aspects of the previous models, and includes practical and motivational concerns.

Overview of judgments

Metacognitive judgments are important to the allocation of study time, which in turn impacts performance (Thiede, Anderson, & Therriault, 2003; Kornell & Metcalfe, 2006; Metcalfe & Kornell, 2005; Thiede & Dunlosky, 1999). If students cannot accurately monitor their learning, they will make poor decisions while studying, which will then lower test performance. As an example, they may stop studying when items are still unlearned, or keep studying even though items are already learned well. Conversely, improved metacognitive accuracy is also associated with better test performance: Theide et al. (2003) showed that metacomprehension (comprehension judgments for text material) accuracy was improved when key words (words that students felt were relevant to the meaning of the text) were solicited after a short time delay. This in turn improved students' abilities to effectively focus their efforts when given time to restudy items: they more often chose to restudy texts that were less well learned, which was an effective strategy, evidenced by their superior test performance when compared to students who did not generate the key words. These results suggest that metacognitive accuracy guides learning, and may impact student performance.

Research concerning this relationship between metacognitive accuracy and performance has focused on two similar judgments. Judgments of learning (JOLs) (Kornell & Metcalfe, 2006; Metcalfe & Finn, 2008) are judgments of how much information a person feels that they have learned and will recall on test, and are usually solicited on a percentage scale (for example, a participant may respond that they know 80% of the items). Metacomprehension accuracy (Dunlosky & Lipko, 2007), is a judgment of how well a piece of text is understood, also generally phrased in terms of how well they believe they would do on a test concerning the material. Other judgments that people have studied (though not necessarily in the context of study allocation and performance) include feeling of knowing, which is a measure of how likely you are to recall an item on test for items you can’t currently recall, and ease of learning judgments about the relative ease with which different items can be learned. These judgments have been investigated in many studies (Underwood, 1966, Nelson & Leonesio, 1988, Leonesio & Nelson, 1990; Jameson, Nelson, Leonesio & Narens, 1993), and while on the surface it may seem as if they measure overlapping constructs, it has been found that ease of learning, judgments of learning, and feeling of knowing measures are distinct measures that are not highly correlated (Leonesio & Nelson, 1990). Leonesio and Nelson (1990) conducted a study where participants made ease of learning judgments, learned the items, made judgments of learning, returned in 4 weeks and made feeling of knowing judgments, and had a recognition test. What they found was that the correlations of these three judgments to each other was rather low, and all three were below .19.

In addition to the judgments mentioned above, there may be other metamemory judgments that are relevant to study time allocation and subsequent performance; specifically, judgments concerning learning rate (Metcalfe & Kornell, 2003; Metcalfe & Kornell, 2005) or in other words, a judgment of improvement. Judgments of improvement would benefit study because students could choose to quit studying an item when their rate of learning drops, and they are making little to no progress. This would be especially useful when there is limited time, because no time would be “wasted” on items where there is not much progress being made. Efforts could be focused on high-rate, quickly learned material to maximize learning.

Judgments of Improvement

Judgments of improvement (JOIs) have not been investigated, but rather repeated JOLs have been assumed to represent the subjective learning curve (Koriat, Sheffer, & Ma'ayan, 2002; Metcalfe & Kornell, 2005). Metcalfe and Kornell (2005) inferred the participants' judged rate of learning by subtracting the stopping JOL from the starting JOL; but this JOL difference score, while likely somewhat related to actual rate of learning, may not reflect the subjective sense of improvement. These JOL difference scores don't truly represent a rate because there is no measure of time, and to infer that these represent the subjective sense of improvement may be an inaccurate assumption because it presumes that people remember their previous JOL states. It is not likely that all previous JOL states would be remembered (or even just one preceding JOL) – especially in a complex learning scenario, so it is important to measure judgments of improvement. If previous assumptions are correct (Metcalfe & Kornell, 2005; Koriat, et al., 2002), then JOIs should simply reflect JOL difference scores. JOIs may indeed be based in part on JOLs, but they may also be influenced by subjective cues, such as a sense of fluency, interest, frustration, and so on.

Metcalfe and Kornell (2005) explained why a judgment of learning rate could be a useful and expected judgment during study time. In the "region of proximal learning" model, students use a judgment of their learning rate as a stopping rule while studying. In this model, students first choose items in order of how easy they are to learn. They should stop study of a particular item when the (perceived) learning rate drops to an unacceptably low value, and move onto the next item, which maximizes the rate of return per unit time spent studying. Similarly, a model of optimal learning (Son & Sethi, 2006) illustrates how the optimal behavior (which results in the maximum amount of learning per unit time spent studying) in most learning scenarios would be to focus on the item with the highest current rate of improvement. This is in stark contrast to the previous prevailing model of study time allocation, discrepancy reduction (Thiede & Dunlosky, 1999; Koriat & Goldsmith, 1996; Carver & Scheier, 1990), in which students chose the most difficult (low JOL) items for restudy; evidence suggested that people chose to study the most difficult items (with the largest discrepancies between the current state of learning and the goal), and spent more time on them. This model is referred to as discrepancy reduction because it is assumed that people work to reduce discrepancies between the current state and the goal state (Carver & Scheier, 1990). These results have been replicated many times (Koriat & Goldsmith, 1996; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988). Recently, however, conditions have been found in which people did not choose to study the most difficult items, and instead started with easier (yet unlearned) items (Metcalfe & Kornell, 2003; Dunlosky & Thiede, 2004; Kornell & Metcalfe, 2006). Whether this is a learned strategy, or an automatic change in preference is unclear; if this is a change in preference, this would lend some support to the proximal learning model, but it still remains to be seen whether or not judgments of improvement are important to stopping decisions.

Uptake Rates and Behavior

Other behavioral research suggests that people are sensitive to uptake rates in their regulation of behavior, for example, information foraging (Pirolli & Card, 1999; Pirolli, 2005) and switching between tasks (Payne, Duggan, & Neth, 2007). This general rule of switching when the uptake rate slows follows from the marginal value theorem in optimal foraging (Charnov, 1976; Stephens & Krebs, 1986), which states that exploration of the current patch should cease when the marginal rate of gain in the patch is equal to the average rate of gain in the environment (The marginal value theorem applies to a situation with diminishing returns). To give an example, if you are gathering berries on a particular bush, and your rate of finding berries drops off to the average rate of obtaining berries, the best decision is to move on to another bush because the current one no longer provides a higher rate of return. This example assumes that complete knowledge of the environment is available, but that is not necessarily required, as demonstrated by Green's assessment rule (Green, 1984), which states that foragers will remain in a patch for a certain amount of time, and this amount of time increases by a constant amount with every 'encounter' of a new item. To illustrate, an animal may (by default) forage in a bush for five minutes, but each time it finds a berry, this amount of time increases by 30 seconds. This ensures that the animal forages longer in more dense areas, and doesn't waste much time in sparse ones.

Information foraging (Pirolli & Card, 1999; Pirolli, 2005) is based on research concerning optimal foraging in animals. Pirolli and Card (1999) describe the goal of the forager as maximizing the intake per effort, which is a complex problem: different areas and prey have different values, and there is a cost to pursuing them. In terms of a person foraging for information, they must find relevant, high quality information which exists in different sources (books, internet search queries, journals, etc) that vary in richness or density of information, and differ in how much effort it takes to explore them (in this example, it may be easiest to peruse the internet rather than to hunt for books at the library). The information forager then has to attempt to maximize their intake of information “prey” per time and effort. Pirolli and Card (1999) also note, similarly to animals foraging for food in the wild, our sources of information also seem to exist in a patchy environment. This scenario lends itself very well to predictions from optimal foraging. In their observations of actual information foraging behavior, they found that people did seem to behave in line with predictions from optimal foraging; they chose items based on their perceived utility, and behaved in ways that minimized time cost, such as thinning out piles of already-obtained items.

Task switching behavior is another human behavior that can be analyzed in terms of foraging theory. In a task-switching situation, there may be multiple tasks that need to be finished which vary in their productivity. Payne et al. (2007) examined how people decided to allocate their time and decide when to switch between two different tasks (scrabble and word search puzzles). Unlike information foraging and animal foraging, there was no direct cost to switching, like search or travel costs. In this case, there is a different kind of cost, a period of slower performance right after switching to the other task, which people may or may not have been sensitive to. This situation is not *directly* analogous to foraging, because this is not accumulating items, but completing tasks; there is the additional decision of whether to switch before or after completing a goal, because that may have consequences for performance as well. In this situation, the optimal foraging prediction would be that people would behave in accord with the marginal value theorem, and move to the other task if the rate of gain was no longer greater than the alternative task. What Payne et al. (2007) found was that people were indeed sensitive to reward, and spent more of their time in the easier of the two tasks, but they also tended to switch after completing a subgoal of the task.

Foraging Analogies and Studying

The findings on information foraging and task switching are relevant to study time allocation, and suggest that people may be sensitive to the rate of return, using this information to maximize learning per unit cost. This should be approached with caution, however, because the situation is not entirely the same. When studying, people are not just accumulating *enough* information (they need to understand and remember it), and they are not merely switching between tasks. The goals in studying may even vary among people and among situations: goals could be complete mastery of the material, learning just enough to pass an exam, or allocating a very limited amount of time in order to maximize one's grades. If the goal was complete mastery of the material, then people may not care about rate of return, and may persist at any cost. When learning just enough to pass an exam, people may just study until they reach the point where they think passing will occur, and then stop. When foraging for information the task is different: people simply need to accumulate a satisfactory amount of relevant information, as opposed to committing a specified set of information to memory. Also, people can work in parallel to some extent while foraging for information, with multiple browser windows open, stacks of journals, and so forth, so that it may be easier to gauge the relative utility of items. During study, in contrast, attention is (usually) focused on one item at a time, so they may not be thinking about other items (of which there may be several, making comparisons difficult). It is also true that when studying, people don't have to forage or find information, nor are they required to do specific tasks. They have the goal of learning a specified set of material. Payne et al.'s (2007) findings that subgoal completion was an important factor in task switching also may be applicable to self-paced study.

An important distinction between metacognitive research and the work on information foraging (Pirolli & Card, 1999; Pirolli, 2005) and task switching (Payne et al., 2007) is that people were not asked to explicitly make judgments about items. Participants were not asked to rate items on their utility, or to estimate their rate of task completion; in these tasks, only behavior was observed. This does not tell us about the metacognitive judgments people may, or may not, have been making. Even if people seemed to behave in ways that suggest that they are sensitive to the rate of return, it cannot be assumed that they explicitly use this information. It has been suggested that metacognitive monitoring and control is implicit (Reder & Schunn, 1996). This may sound extreme, yet animals match their response rates on two alternatives to the relative reward schedules (Herrnstein, 1970), and nobody is suggesting that they have explicit knowledge of these rates. Like animals, it is possible that people gain experience that alters their behavior, but the learning and the changes are implicit (Reder & Schunn, 1996).

JOLs and Studying Behavior

Students do (explicitly) produce and use judgments of learning (JOLs) when studying, in order to appropriately decide what to work on, and for how long (Kornell & Bjork, 2007; Kornell & Metcalfe, 2006; Nelson, et al., 1994). When comparing a situation where people select items to restudy, better performance results when people's choices are honored as opposed to dishonored (Kornell & Metcalfe, 2006), showing that people generally choose items which could benefit from restudy, even if not specifically asked to make judgments of learning; this suggests that even when participants are not told to make JOLs, they consider how well learned each item is when making study decisions. Metcalfe and Finn (2008) also demonstrated that JOLs had a direct relationship with what material was selected for restudy; manipulations that influenced JOLs also directly influenced study choices. For example, they presented students with word pairs, some of which were forward associated (high likelihood of the cue word eliciting the target word), backwards associated (high likelihood that the *target* word would elicit the cue word, but not the other way around), or unrelated pairs. Both forward- and backward- associated pairs elicited higher JOLs than the unrelated pairs, and so were less likely to be chosen for restudy. Recall performance was lower for the backwards-associated word pairs, so people were behaving in accord with their 'illusory' JOLs.

Much research has been conducted concerning the accuracy of these judgments, with the general finding that they are fairly accurate. For example, one study by Nelson and Dunlosky (1991) had a gamma correlation of .93 between JOLs and actual recall when JOLs were delayed for a short period of time after study. Such correlations are promising; however, this is only *relative* accuracy. People generally show biases in their judgments (Koriat, et al., 2002; Finn & Metcalfe, 2008; Meeter & Nelson, 2003), meaning that *absolute* accuracy is impaired. These biases are best illustrated in multi-trial experiments, which exhibit something called the underconfidence with practice effect

(Koriat, et al., 2002; Meeter & Nelson, 2003; Koriat, 1997). On the first trial, people generally show a bias towards overconfidence, giving JOLs that are higher than their actual performance, but on succeeding trials, they give JOLs that are lower than actual performance (Koriat, et al., 2002). Koriat et al. (2002) analyzed 11 experiments in which people learned a list of paired associates that were presented for more than one trial, and made JOLs immediately after studying each pair. They found that for the first presentation, if the JOL was of low magnitude, it was often underconfident, and if the JOL was of high magnitude, it was overconfident; but for presentations 2 through 4, there was consistent and increasing underconfidence, no matter the magnitude of the JOL. They also found that relative accuracy of JOLs (within items in a list) improved with practice. This underconfidence with practice effect occurred under a wide range of circumstances: when participants were given feedback about their responses, with backward and forward associated word pairs, related and unrelated word pairs, hard items and easy items, and aggregate and specific item judgments (Koriat et al., 2002). Koriat et al. (2002) conclude, "The results suggest a dissociation between objective and subjective learning curves such that learners systematically underestimate the benefits from practice on memory performance". The underconfidence with practice effect may translate to inaccurate judgments of improvement; if JOIs are directly related to JOLs, then we might expect that these biases will impair the ability to accurately judge improvement, and that improvement should be increasingly underestimated as well. Also, from trial 1 to 2, judgments of improvement may be extremely underestimated, as JOLs often shift from overconfidence to underconfidence.

In summary, optimal foraging theory (Stephens & Krebs, 1986) and information foraging experiments (Pirolli & Card, 1999; Pirolli, 2005) indicate that behavior should be sensitive to rate of gain (improvement), so that switch times occur at a time when the current item no longer offers a higher rate of gain than the average. Task interleaving data (Payne et al., 2007) suggests that sub-task completion may be important to switch decisions, but rates are important and also inform switch decisions. This literature does not suggest that rates are explicitly assessed or verbalized, while the metamemory literature generally assumes that metacognitive processes are explicit (see Reder & Schunn for an exception). The metamemory literature suggests that study decisions depend on explicit judgments of learning (Kornell & Metcalfe, 2006), and that repeated JOLs may be used to assess improvement (Metcalfe & Kornell, 2005), though this has potential limitations as JOLs are generally affected by increased underconfidence with practice (Koriat et al., 2002).

Previous Experiments

Previous experiments were designed to test the accuracy of judgments of improvement. If improvement was inferred from judgments of learning, we expected that judgments of improvement might follow the same pattern of increasing underconfidence; if other factors such as fluency, interest, or other subjective cues are used to inform JOIs, then a very different pattern may arise, including an increasing sense of confidence with practice. For the experiments it was assumed that JOIs would be explicit judgments, so we tested accuracy of these judgments by soliciting JOIs directly (rather than analyzing switching behaviors between different study tasks). Since we were measuring improvement, the tasks involved memorization over several trials, which allowed for us to collect multiple JOIs. We also assumed that JOIs would be a judgment used during study, so JOIs (and JOLs) were solicited immediately after each study trial, though in one of the experiments we investigated a predictive JOI, as it is possible that a student may predict future learning when making decisions about study allocation.

The following variables were manipulated in the experiments: materials, judgment timing, and scale. The first two experiments used n-grams (Shannon, 1948), while Experiments 3 and 4 used the somewhat more conventional paired-associate learning paradigm. The first three experiments solely used judgments made immediately after each study trial, while the fourth experiment compared judgments made before versus after study (judgments of improvement were made by asking: “how much will you improve?” or “how much did you improve?”). We assumed that for judgments of improvement to be useful, they would be needed during study, so judgments of improvement were requested immediately rather than after a delay. In Experiment 1, improvement judgments were made on the same scale as judgments of learning. In Experiments 2 and 3, to minimize the possibility that the participants were not simply subtracting the second most recent judgment of learning from the most recent judgment of learning, we solicited judgments of improvement on a different rating scale (0-6). To further address this in Experiment 4, judgments were made between subjects, so no participant made both JOIs and JOLs. In Experiment 4 we also investigated potential differences in judgments due to rating scale, so participants were assigned either to percentage judgments or number of word judgments.

Ultimately, we found no evidence to suggest that participants were accurate above chance at judging their improvement before or after learning trials, with any of the materials used, or with any of the scales used. More troubling, judgments were (on average) higher when improvement was lowest, during the later study trials. A potential criticism of these experiments is an issue with aggregate level judgments, which are often underestimated in comparison to an average of individual item judgments (Mazzoni & Nelson, 1995; Treadwell & Nelson, 1996, Koriat et al., 2002); it might be argued that JOIs could be made accurately, but more on the level of individual items rather than an aggregation of items. However, it seems unlikely that learners accurately can judge the future learnability of different items, as they will drop flashcards even when further improvements are likely (Kornell & Bjork, 2007).

Individual level assessments of improvement appear to be very similar to ease of learning judgments. While individual item JOIs might ask for how much improvement might be made on the different items, ease of learning judgments are estimates of how easy or difficult an item would be to learn. With EOLs, learners may think about how easy items are, or perhaps which would take longer to learn than others. These judgments have been found to be somewhat predictive; previous research has found EOL judgments to be related to the length of time needed to learn items, with correlations varying from .03 (Koriat, 2008) to .90 (Underwood, 1966), though the high correlations that Underwood found were relating individuals’ EOLs to group learning rather than estimates of their own learning (this had a smaller correlation of .48). Kearney and

Zechmeister (1989) asked participants to make EOL rankings of items after some experience studying the items (but no test trial); not only did they find significant correlations between EOL ranks and normative item difficulty, but fast learners had higher correlations than slow learners.

Unfortunately, even if students can make individual JOIs, they still may not able to make accurate aggregate JOIs, so decisions about when to switch study material (such as switching between what chapter to study, or what class to study for) would be sub-optimal if based on aggregate JOIs. A clue seen in a study skills survey we conducted suggests that feelings of learning rate may not actually be the preferred mechanism- more students selected other answers, such as feeling bored or tired (which could indicate either a lack of commitment or studying to fatigue), quizzing oneself and finding performance satisfactory, or feeling that they have learned enough total material. Additionally, most students stated that they set goals for themselves, and that they meet tend to meet these goals.

Study Goals

Goals have been an assumption in most models of study allocation (Nelson & Narens, 1990; Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999; Bandura, 1993), and it appears that these goals refer to performance goals, such as passing a course, or an A or a B grade. Overall, we can assume that learners strive to achieve the goals they set for themselves, and perform higher if they set higher goals, which is reasonable since participants do learn more when given higher goals (as compared to lower goals), even if they don't meet or exceed those goals (Koriat, Ma'ayan, & Nussinson, 2006; Dunlosky & Thiede, 2004; Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999; LaPorte & Nath, 1976).

Past research on study goals in the metamemory literature has examined the effect that goals have on behavior, such as what kinds of items are chosen, and for how long people study. The goal is often referred to as a "norm of study" that learners strive to (or in the case of experiments, are instructed to) meet (Nelson & Leonesio, 1988; LaPorte & Nath, 1976; Dunlosky & Thiede, 1998). One interesting facet of these studies was that learners often did not meet the norm of study; In Nelson and Leonesio (1988), participants were instructed to master a set of trigrams, yet only recalled about half of them. Other studies have also found a failure to meet specified goals when studying paragraphs (LaPorte & Nath, 1976) and paired associates (Thiede & Dunlosky, 1999). Students often claim to set goals for themselves (get an A on this exam, get a B on that quiz, etc), and most students seem to be able to perform satisfactorily. In a survey we conducted of 242 students, 201 stated that they set goals, and another 24 noted that they made goals of simply passing courses; additionally, they stated that they tend to meet their goals: 82 said yes, they meet goals, and 98 stated that they meet them sometimes or usually. Clearly, these students set goals for themselves, and they meet these goals quite often. This situation suggests that there is something missing; a cynical perspective would be that participants are not motivated to perform in experiments like they are in a course, though this does not satisfactorily explain why participants over-perform when given low goals (Thiede & Dunlsoky, 1999; Dunlosky & Thiede, 2004), unless it can be argued that it may take more effort to do otherwise (which may indeed be the case). Relevant factors to goal achievement may include motivation (in the case of high goals), the lack of variety in study activities available in the experiments (specifically, the lack of self testing), or simply poor absolute accuracy of judgments of learning.

Motivation and goals

Motivation may play a key role, at least when it comes to difficult goals. Bandura and Cervone (1983) state that goals without feedback are not motivating; they found that performance only improved over the control condition when both goals and feedback were provided. Also, moderate discrepancies between the current state and the goal were ideally motivating when compared to large discrepancies; this suggests that even if students have feedback, it may be de-motivating if they discover that they are very far from their goal. Aside from feedback, internal factors affect motivation- more specifically, self-efficacy beliefs. Induced high or low feelings of self efficacy have direct effects on performance; simply providing positive or negative verbal feedback (emphasizing being either better or worse than one’s peers) resulted in higher induced self efficacy being associated with completing more problems, and lower induced self efficacy with doing less problems (Bouffard-Bouchard, 1990). Bouffard-Bouchard (1990) also found interesting results when asking participants to report on their goal setting behavior; the high self-efficacy group reported setting higher and more specific goals than the low self-efficacy group. Self efficacy is important to motivation: participants with high self efficacy and high dissatisfaction with current performance increased their performance on an ergometer exercise device much more than other groups, and the low self efficacy, low dissatisfaction group even decreased their performance (Bandura & Cervone, 1983); these results highlight the idea that it is important to believe oneself to be capable of performing the task, and to have a discrepancy between the current state and one's desired goal. If an individual does not think they can perform a task, it may not matter much whether they'd like to perform better (but perhaps will result in anxiety), and likewise if an individual believes themselves capable but is not interested in improvement, no change will occur (they "slack off").

Bandura (1977) introduced the concept of self-efficacy, essentially a belief that one can do something successfully – this could be any action, such as being able to learn calculus or successfully handle snakes. Feelings of self-efficacy are also more than just a general metacognitive assessment; they are informed by past accomplishments, but are also influenced by vicarious experience, verbal persuasion, and emotional arousal, and interpretation of these sources of information (Bandura, 1977). Additionally, self-efficacy may influence the interpretation of failures; those with high self-efficacy ascribed failure to a lack of sufficient effort, while low self-efficacy students assumed that failures are evidence of a lack of ability (Collins, 1982). Such interpretations are likely to influence subsequent behavior; if a student believes that they did not put out enough effort, they may be motivated to work harder to reach their goals, but if they simply believe themselves incapable, they may accept poor performance. Self-efficacy beliefs may influence a wide variety of behaviors, from choice of activities and situations, how much effort to exert, how long to persist, and what goals to strive for (Bandura, 1977). A review by Berry and West (1993) details several studies that show higher self-efficacy to be related to goal setting (higher, more specific goals), choice behavior (to engage in tasks, and choosing more challenging task), longer persistence on task, and more use of metacognitive strategies. Many of these findings have been supported in the domain of learning and studying; higher self-efficacy is related to better high school grades and lower drop out rates (Caprara, Fida, Vecchione, Del Bove, Vecchio, Barbaranelli & Bandura, 2008), and is associated with longer persistence on task (Berry, 1987; Bouffard-Bouchard, 1990), and choosing more difficult items (Sexton & Tuckman, 1991). Higher self-efficacy also seems to be associated with metacognitive skill; high levels of self efficacy in mathematics was associated with solving more problems, reworking more failed problems, and quickly moving on from poor strategies (Collins, 1982).

Metamemory accuracy and goals

Assuming that an individual is attempting to reach a goal, the accuracy of meeting these goals will presumably be affected by metamemory accuracy. The same factors that influence judgments of learning (or JOLs) will likely affect how close a student feels they are to reaching their goals; it is assumed that learners make ongoing JOLs as they study, and through this mechanism they decide when they have learned items well enough (Nelson & Narens, 1990; Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 1998). JOLs have been found to increase in accuracy with practice, in the form of repeated trials, though they become increasingly underconfident (Finn & Metcalfe, 2007; Koriat, Sheffer

& Ma'ayan, 2002; Koriat, Ma'ayan, Sheffer & Bjork, 2006; Meeter & Nelson, 2003) and these changes may depend on experience with test (Finn & Metcalfe, 2007). On the first trial, higher JOLs are overconfident and lower JOLs are underconfident. These shortcomings are also seen when participants are assigned goals to achieve, suggesting that any judgments of learning that are made suffer from the same inaccuracies as trial 1 JOLs; with low goals, more items than necessary are learned, and high goals are not reached (Nelson & Leonesio, 1998; Thiede & Dunlosky, 1999; Dunlosky & Thiede, 2004). These results are suggestive of problems in the absolute accuracy of JOLs. JOL accuracy can, however, improve with practice and generalize to new items if the judgments are said to be theory based, or in other words, informed by learned rules or theories of how learning works (Koriat & Bjork, 2006).

One factor to consider may be that students engage in a few different activities when they study – they may do things like make flashcards, and test themselves. A survey by Kornell and Bjork (2007) supports this claim; students said that they tested themselves while studying. A study by Kornell and Son (in press) found that participants preferred self testing over mere re-presentation, though they didn't do so because it was an effective learning strategy, but rather because, as they answered on a post experiment survey, it helped them monitor their learning. This may mean that at least some students do not rely on their metacognitive judgments when deciding to stop studying, and perhaps they test themselves to confirm if they really have learned enough. Additionally, when studying for a course, students may be also be relying on past study and exam experiences for information (studying this way (or this hard) got me a "B" on the last test). Both self-testing and experience with the task may result in better goal achievement than what has been typically seen in experimental situations. The behavior of participants with high and low goals may be described by different models of study time allocation, which are detailed in the following section.

Models of Study Time Allocation

The problem of how to allocate study time is a problem central to the lives of students, affecting how well they learn and perform in the classroom (and possibly future endeavors). It is important to understand how students allocate their time- whether allocation is generally optimal or sub – optimal, whether successful students are more optimal – or conversely, if they are merely willing to spend more of their time. Qualities of optimal allocation may include studying in such a way as to maximize learning while minimizing time, and obtaining the best grades possible from the time that one is willing (or able) to spend studying. To this end, a number of decisions must be made: what to study next, when to stop the current activity and change strategies, and when to stop studying.

Models or frameworks of study allocation tend to have these central ideas and assumptions in common: that learners have a goal (or, “norm”), that learners make metacognitive judgments, and that they use these metacognitive judgments to guide their behavior. Nelson and Narens’ general framework (1990) describes a system where there are two levels of cognitive processes: an object level and a meta level, and that these two levels interact through monitoring and control processes. Monitoring processes are metamemory judgments that may assess the current state of learning, and control processes are decisions to initiate, continue, or terminate actions. Nelson and Narens (1990) group specific monitoring and control processes into different stages of the learning process. During the acquisition stage, the student sets a goal for how well each item should be learned (which may be influenced by their theory of retention (Maki & Berry, 1984), and beliefs about how quickly material is forgotten (Koriat, Bjork, Sheffer & Bar, 2004)), and sets a study plan to meet the desired goal. Before items are learned, ease of learning (EOL) judgments predict the relative ease or difficulty of the material-Nelson and Narens also note that EOL judgments can be made about what strategies may make learning easier. EOL judgments are correlated with learning rates (correlations around .9; Underwood, 1966), and with study time allocation (correlations around -.3) (Nelson & Leonesio, 1988), however research does not specifically show that learners will spontaneously, explicitly, make EOL judgments. Judgments of learning (JOL) are made during and/or after learning, and predict test performance with correlations as high as .91 (Nelson & Dunlosky, 1991). Feelings of knowing (FOK) are made during or after learning as well, but they make learning predictions about items that are currently not recallable; these judgments are not as accurate as JOLs (Leonesio & Nelson, 1990).

During the study session, students make decisions about whether to continue or move on to another item, and this decision may depend on if their JOL matches the study goal that was set (Nelson & Narens, 1990); there currently is not much evidence to support this claim, as we have found in pilot experiments that students do not meet or exceed study goals that are given, unless those goals are low, and similar results were also found in other experiments where participants were given goals (Nelson & Leonesio, 1988; LaPorte & Nath, 1976; Theide & Dunlosky, 1999). It could be argued, however, that students' JOLs are simply inaccurate and very over-confident, which tends to be the case unless participants experience a test trial (Koriat, Sheffer, & Ma'ayan, 2002). In Nelson and Leonesio's (1988) experiment, participants given a goal of 100% only reached 49% correct performance; Thiede & Dunlosky (1999) found that when given a goal of 24 of 30 items the mean correct recall was 10.5 words. Both of these experiments only had one study trial, and, additionally, Thiede & Dunlosky had participants choose a subset of the items for restudy. These factors may have affected goal achievement negatively, without having all the items to study or the opportunity to restudy after a test trial (students may rely on testing to get a better idea of their performance). After initial study, during the retention phase, Nelson and Narens (1990) propose that students make decisions similarly to the way they are made during acquisition, as well as according to their theory of forgetting. They may allocate time to items based on the discrepancy between the current JOL and the goal, and if an item can't be recalled, based on its FOK.

The discrepancy reduction model of study allocation addresses the question of how decisions are made about what items to study, and whether to continue or end study of a particular item. The central idea of this model is that there is a discrepancy between the current state of the item, and the goal for that item (for example, my goal is 100% recall on the exam, but I give this item an 80% JOL – so, I need to study further). Items should be prioritized in terms of how far they are from the goal state, which is indeed the case in many studies; Son and Metcalfe (2000) found that in 35 of 46 published experimental conditions, more difficult materials were prioritized. According to this model, students will continue studying an item if they have not reached their goal, and once they reach their goal they will stop (Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 1998). In order to meet these demands, items with lower judgments of learning should have longer study times than items with high JOLs, and also items with higher study goals should be studied longer than those with lower goals (Dunlosky & Thiede, 1998; LaPorte & Nath, 1976). For example, when accuracy is emphasized, study time is longer compared to speed – emphasized instructions for learning (Nelson & Leonesio, 1988; Dunlosky & Thiede, 1998); study time is longer for items awarded more points (Dunlosky & Thiede, 1998), and for those items more likely to be on test (Dunlosky & Thiede, 1998). Study allocation was not optimal in these studies, as students didn't focus enough on the prioritized items, nor could they meet specific goals that were given to them (LaPorte & Nath, 1976; Thiede & Dunlosky, 1999). This may be due to inaccurate metacognitions; goals cannot be met if learners incorrectly judge that they have met their goal, but sometimes it may be the case that goals are not attainable. This leads to the labor in vain effect, which refers to a situation in which further time spent studying doesn't lead to any gains in performance (Nelson & Leonesio, 1988).

Dunlosky and Thiede (1998) propose an alternative mechanism for terminating study- rather than solely depend on the perception that goals have been met, learners could perceive a lack of progress for some amount of time. This alternative is also the key element to Metcalfe's (2002) proximal learning theory. In the proximal learning theory, rate of learning is key to decisions about what to study first, and when to continue or terminate study (rather than absolute discrepancies between current state and the desired state). Son and Metcalfe (2000) discovered a scenario in which people did not choose to prioritize the difficult to learn items: when given long passages, or when under time pressure. Further experiments confirmed these results and referred to this as a shift to easier materials effect (Son & Metcalfe, 2000; Dunlosky & Thiede, 2004). A key finding of Metcalfe (2002) was that prioritized items (of Spanish vocabulary) were different for Spanish experts versus Spanish novices – the experts prioritized medium vocabulary items while novices started with easy vocabulary items; this is explained by the simple fact that experts already knew the easy items. This supports the idea that learners chose items with a higher rate of learning, which for the experts meant studying the easiest items they did not know already. Decisions about whether to terminate study are made according to the current rate of learning; if it is low relative to gains that could potentially be made elsewhere, the current item is terminated and the learner moves on to a more productive item. Currently, there is not substantial evidence that supports this claim, but research in information foraging (Pirolli & Card, 1999) and task switching (Payne, Duggan & Neth, 2007) suggests that it may be possible to switch tasks according to rates of return.

Thiede and Dunlosky (1999) describe a general hierarchical model of study time allocation. In this model, there is a super-ordinate level of pre-study planning, involved with more general decisions about how to meet their goals that seek to maximize the efficiency of study time. The thought that students exert the minimal effort required to achieve their scholastic goals makes intuitive sense, and it is supported by the finding that under time pressure, learners shift to studying easier items (Dunlosky & Thiede, 2004; Metcalfe, 2002; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999). The subordinate level then regulates the study of individual items. The model also explains the divergent findings of the shift to easier materials effect, and the more common finding of spending more time with low JOL items: according to Thiede and Dunlosky (1999), when a learner has a mastery goal, low JOL items will be prioritized, and with low goals, high JOL (easy) items will be prioritized so the least effort is expended. Under mastery goal conditions, expectations are identical to the discrepancy reduction model- the only difference occurs under low goal conditions, in which learners seek to minimize their effort to achieve the desired goal. This is opposite to the predictions of the discrepancy reduction model. As noted above, there is evidence to support the prioritization of easier items under low goals; however, there are conflicting results that suggest a boundary condition. Easier items are selected for study only when the items were presented as an array, not when shown individually; participants chose more difficult items, and chose more than required to meet the goal (Thiede & Dunlosky, 1999). This finding may be explained by taking into account the limitations of working memory – it is simply too difficult to keep the goal in mind, and all of the items in mind, and choose accordingly; it is easier to revert to a “default” plan of discrepancy reduction (Thiede & Dunlosky, 1999). Still, a problem with this model is the issue of achieving goals that are set – even when presented with all the items simultaneously, learners seem to over-achieve when given low goals, and fail to learn enough when given more difficult goals (Thiede &

Dunlsoky, 1999, LaPorte & Nath, 1976; Nelson & Leonesio, 1988).

The models mentioned above have many things in common – they focus on how learners choose what items to study, how long to study, and when to give up. They discuss behavior under different kinds of goals, yet they do not discuss how those goals are set, or why high goals are not reached. A social cognitive model of self-regulated learning may bear on the problem of goal achievement and goal setting (Bandura, 1991; Zimmerman, 1990). The social cognitive model describes a system in which self efficacy plays an important role as a mediator in goal setting and persistence (Bandura, 1993); essentially, self-efficacy is thought to be important for motivation, and affects the use of self–regulatory skills and avoiding distractions (Pintrich & DeGroot, 1990; Zimmerman & Bandura, 1994). Self-efficacy is defined as "an individual's sense of competence and confidence related to performance in a given domain" (Berry & West, 1993, p 351), and is influenced by past experience, vicarious experience, verbal feedback, and physiological reactions (e.g. physical arousal, such as heart rate and sweating) (Zimmerman, 2000). Self-efficacy has been found to influence goals that are set, with higher self-efficacy associated with the setting of higher goals, and both self-efficacy and goals affect academic performance positively (Zimmerman, Bandura & Martinez-Pons, 1992; Zimmerman & Bandura, 1994; Zimmerman, 2000). Self-efficacy and goal setting explained 29% of the variance in writing course grades in one experiment (Zimmerman & Bandura, 1994), induced high or low self-efficacy affected task persistence in a verbal concept formation task (Bouffard-Bouchard, 1990), and self-efficacy was correlated with reported use of cognitive strategies and self regulation (Pintrich & DeGroot, 1990).

Clearly, these models are not mutually exclusive, they overlap somewhat with each other and there is evidence to support all (partly because they address different levels or aspects of the situation). First, social cognitive theory describes factors that influence self-efficacy, which in turn has a direct relationship with goal setting and persistence on task. Self-efficacy, as a concept, is certainly related to metacognition- it is essentially a global metacognitive judgment or belief that is influenced by experiences and feedback, though research does support that it is a distinct construct from metacognition, relating to different kinds of skills (Moores, Chang & Smith, 2006). The hierarchical model predicts that in high goal conditions the discrepancy-reduction model is followed: low JOL (hard) items are prioritized, and items are learned until they reach the goal state. Under lower goal conditions, the hierarchical model aligns with the proximal learning theory if conditions are right (i.e. items presented simultaneously, not sequentially), and learners will prioritize the unlearned high JOL items and learn items until the perceived rate of learning drops to an unacceptably low level. If it is difficult for learners to enact this strategy (items presented sequentially, distracting environment), we should see behavior consistent with discrepancy reduction, but persistence should be lower – according to the social cognitive theory. It also follows from social cognitive theory that lower self efficacy should predict shorter giving up times for items, and higher self efficacy should predict longer giving up times; perhaps this is tied to a prediction of whether or not more study time would be fruitful (like a theory – based judgment of improvement informed by self efficacy). In a high self-efficacy, high goal condition, there should be less risk of laboring in vain as these learners tend to have more metacognitive strategies and skills at their disposal (Pintrich & DeGroot, 1990), and in these situations we may see strategy switching when perceived learning rate drops. A limiting overall factor (for any model) will be metamemory accuracy – both relative, for item choosing, and absolute, for goal reaching. If these judgments are inaccurate, performance may still fall short of learning goals.

Self-regulated learning may be more complicated than these models describe. Some learners may not strategically prioritize items, and study the material in the order that it is presented in a textbook or in their notes. Decisions about when to stop studying may often be made in other ways, such as feelings of boredom, or other priorities (for some students studying is a higher priority than others; some may stop studying because their favorite show is on TV, others might stop only for higher priorities like needing to go to work). Some students may even study for hours until they can’t anymore and fall asleep. Other students may stop studying when they think they are not making progress or have not mastered the material in a reasonable amount of time (though the accuracy of these judgments may be quite poor).

Ultimately, we predict that it requires a great deal of motivation and high overall self-efficacy to attempt and maintain studying in a strategic and thoughtful manner, or to even make metacognitive judgments at all. We are not suggesting that strategizing and making metacognitive assessments are an all or nothing matter; there is likely a continuum of how often or how many metacognitive judgments are made, how elaborate overall strategizing may be, how many kinds of study strategies are used, and how many self-regulatory behaviors students engage in. Motivation most likely changes on a daily basis, depending on when items are due, what is more interesting, how well one is doing in different courses, and before exams. For those students motivated enough to regulate their learning in a strategic manner, behavior within a study session may be predicted by the hierarchical model (Thiede & Dunlosky, 1999), but for students with low motivation, their behavior may be better described by the region of proximal learning model as long as it takes little effort to recognize the easy items, and they may be more prone to stopping when they are bored, tired, frustrated, or have more attractive priorities to attend to.

Experiments

The experiments examine aspects of these models of study time allocation that have not been fully examined in existent literature, and propose a new perspective on the management of study time: that strategic, self- regulated learning requires motivation, self efficacy, and metacognitive accuracy and skill. Important topics to be explored concern when to switch items, the metacognitive consequences of different study strategies, how to determine a stopping time, and to what degree motivation and reported use of metacognitive strategies and regulation are predictive of metamemory accuracy in the experiments. The models mentioned previously address the issues of switching and stopping, yet the evidence to support these claims is at times absent, unconvincing, or incomplete.

The first set of experiments concern switching decisions. Participants were given a list of items to study, and given the opportunity to change lists while studying. Item

JOLs at switch points will be analyzed to see if they occur at more difficult items, which would support the stopping/switching rule in Metcalfe's (2002) proximal learning theory that students are able to optimize study time and switch items when the learning rate drops. In the second experiment, when participants believe that time will not be limited, there should be much less switching behavior- there is time to study all items, so there is less need to strategize or prioritize. The third experiment should show more switching behavior, as it should be easier to prioritize items when the two lists differ in average difficulty.

The second set of experiments investigates the relationship of self-efficacy and motivation on learning with assigned goals. Experiment 1a involves a manipulation of self-efficacy through a practice trial and looks at persistence and performance. Self efficacy was manipulated between subjects, who were asked how they decided to stop studying, because we hypothesized that self efficacy may be related to JOIs. Experiment 1b modified the efficacy manipulation in an attempt to see stronger effects. Experiment 2 examined the effect of manipulated self-efficacy on judgments of learning, study time, and recall performance. We are interested in whether choosing a goal results in higher performance relative to the goal due to increased motivation, so in Experiment 3, self efficacy was experimentally manipulated, using anchoring, and variables of interest included goals chosen, persistence, performance relative to chosen goals, and reported use of metacognition. The final experiment consists of surveying students in a summer cognitive science course. Students were given a survey that asks about goals, workload, course importance/interest, and self-efficacy. In addition, they were asked about study habits, including strategy use, and self-regulation. Metacognitive accuracy was measured by exam predictions. Similar research has found that self-efficacy indirectly affects class performance through goal setting (Zimmerman, Bandura & Martinez-Pons, 1992; Zimmerman & Bandura, 1994; Pintrich & DeGroot, 1990). Isaacson & Fujita (2006) found evidence that higher performing students were more accurate in their test predictions. In these experiments, similar results should be found, and new relationships should be uncovered between motivational factors and the reported use of metacognitive judgments and strategic regulation of study time.

Additional work included examined the effects of framing on metacognitive judgments, specifically, how negative framing influences judgments of learning. Students were asked to estimate their amount of learning with either positive or negative frames, and estimated learning for themselves, or others. A typical, positive frame judgment would ask students to estimate how much of the material they know, or will get correct, on a recall test. Students are not typically asked to make these kinds of estimates in a variety of ways, such as how much of the material they do not know, or will get incorrect, on a recall test- despite the possibility that many students may conceptualize performance estimates in this way. Experiment 1 looked at positive and negative framing of judgments of learning; positive judgments of learning request participants to estimate the amount of material that will be correctly recalled on a test, while negative framed judgments of learning required the participant to estimate instead the amount of material they will get incorrect on a recall test. It was expected that negative framing would lead to more pessimistic expectations of performance. The second experiment was a survey designed to measure students' theories of learning, to examine if framing effects would persist in a hypothetical scenario. The third experiment was similar to Experiment 1, but also added item level judgments, for the purposes of measuring the influence of different cues on judgments in the different framing conditions. Experiment 4 solicited positive or negative framed judgments of learning, for oneself, or for others, to investigate the possibility that self-enhancement is a mechanism for framing differences.

The last project included investigated the metacognitive effects of different study situations; specifically, the effect of massing, spacing, or repeated testing on metacognitive judgments, to examine how spacing may contribute to lower metacognitive biases, and to what extent testing reduces confidence beyond the effects of mere spacing. It is known that repeated testing reduces confidence (Koriat, Sheffer & Ma'ayan, 2002), but part of this effect may be due to spacing of the study intervals, rather than an effect of recall experience. The first experiment compared metacognitive judgments in a repeated study-test condition to a study – trivia condition, where instead of recall tests, participants answered trivia questions in between study presentations. The second experiment compared massed study repetitions with a study – trivia condition; in order to show what effect mere spacing has on metacognition. The third experiment (a pilot experiment) investigated the effect of massed practice on judgments of improvement, comparing judged improvement to values observed in the first experiment.

Chapter 1

Improvement Judgments and Switching Decisions

Abstract

Judgments of learning rate have been suggested as a possible mechanism behind decisions about switching and stopping during learning. The shift to easier materials effect, which occurs under time constraint when students prioritize easy items, may be due to learners' sense of how quickly each item would be learned. In these experiments we examined the possibility that students could be sensitive to learning rates, and would show strategic behavior in a list-switching task. The effects of time constraint, and average list difficulty were examined, with the result that overall most learners did not strategically switch lists, but most chose to study longer on more difficult items, and that for those that did switch, that switch points were somewhat strategic, as they were more likely to occur at difficult items.

Improvement judgments and switching decisions

Self-regulated learning consists of multiple decisions on the part of the learner. Students must decide what material to study, how long to persist in study, and when to switch to other material. Judgments of learning are known to be used in decisions about what material to study, and influence length of study time, which in turn impacts performance (Thiede, Anderson, & Therriault, 2003; Kornell & Metcalfe, 2006; Metcalfe

& Kornell, 2005; Thiede & Dunlosky, 1999), but not much is known about how students make switching decisions about when to work on other material. It is possible that students consider the speed with which they are learning material to make these decisions, as detailed in the proximal learning theory (Metcalfe & Kornell, 2005), though these decisions may also be informed by other factors, such as length of time spent studying, current judgments of learning, or fatigue.

Judgments of how quickly one is learning would be very useful during study; in the region of proximal learning model, it is hypothesized that learners may use estimates of how quickly they are learning the material when making decisions about when they should stop studying. It is assumed to be more optimal to terminate study of an item when that item's learning rate is reduced, and more learning would occur on other material. A model of optimal learning behavior (Son & Sethi, 2006) illustrates how total learning would be maximized when learners focus on the item with the highest current learning rate. Essentially, learning will be greatest if students always focus on whatever material has the highest learning rate at any given time. The ability to judge learning rate would benefit self-regulated learning, as students could stop studying an item if the learning rate fell to a low level, to avoid working in vain when time is limited.

How accurate students' judgments of improvement are, and whether these judgments are used during the learning process to guide behavior is not yet clear. However, we do know that students use judgments of learning to guide behavior during self-directed learning (Kornell & Bjork, 2007; Kornell & Metcalfe, 2006; Nelson, et al., 1994; Son & Metcalfe, 2000). Specifically, students consider how well they have learned each item to determine how long to study it. Son & Metcalfe (2000) found that in most experiments (35 of 41 published experimental conditions), students chose to prioritize the more difficult, i.e. less well learned items. This prioritization is predicted by the discrepancy reduction model of study time allocation, (Thiede & Dunlosky, 1999), which states that people will choose to focus on items with the greatest discrepancy between current learned state and the goal state. Much evidence has been found to support this model; students generally choose to focus on the most unlearned items, and study them for a longer period of time (Koriat & Goldsmith, 1996; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988). This is not to say that no evidence has been found to support the proximal learning model, however, as conditions have been found where people will focus not on the most difficult items, but rather shift towards easier materials that presumably would be quicker to learn (Metcalfe & Kornell, 2003; Dunlosky & Thiede, 2004; Kornell & Metcalfe, 2006), though none of these experiments explicitly showed that students were thinking about how quickly they would be able to learn the different items.

Thiede and Dunlosky's (1999) hierarchical model also explains the shift to easier materials effect, but without an appeal to perceptions of learning rate. The hierarchical model states that students will generally seek to maximize the efficiency of study time, and while studying, they will regulate the study of individual items. Students' maximization of efficiency is also seen in the boundary conditions of the shift to easier materials effect, as students can only choose to prioritize the easier items when all items are shown in an array, and not when shown individually (Thiede & Dunlosky, 1999), perhaps because it is too difficult a task, or because the presentation style does not encourage such comparative strategizing.

Judgments of one's learning rate, or judgments of improvement, have not been investigated to the extent that other metamemory judgments have, though the existing research suggests that this type of judgment is not very accurate. A study by Kornell and Bjork (2008) found that when students were allowed to drop flashcards from study, they performed slightly worse than students who did not drop cards. In this case, students were weeding out items they judged to have very low potential improvement, and while they acted in a suboptimal manner it may still be the case that those items did have somewhat lower potential for improvement. Some evidence suggests JOIs are highly inaccurate, including predictive "how much more will I learn" and postdictive "how much more did I learn" type judgments (Townsend and Heit, 2009). However, on a study habits survey (unpublished) circulated in fall 2009, 22 percent of respondents answered that they choose to stop studying something when they feel like they are not learning any more, which indicates that some students *may* make this kind of judgment and rely on it to make decisions, regardless of its accuracy.

Other behavioral research concerning the idea of uptake rates suggests that in certain scenarios, such as information foraging (Pirolli & Card, 1999; Pirolli, 2005) and task switching (Payne, Duggan & Neth, 2007), people may be somewhat sensitive to changes in uptake rate. In information foraging, for example, the person attempts to maximize the amount of information retrieved per amount of effort (Pirolli & Card, 1999). When searching out information, one needs to find information from different sources, which may take differing amounts of effort; as such, a person may choose to first search on the Internet, and only attempt to find books from the library when absolutely necessary. In task switching experiments, Payne et al. (2007) also found that people were sensitive to uptake rates, and spent more time in easier tasks, and minimized costs by switching only after completing subgoals. Information foraging and task switching experiments suggest that behavior may show sensitivity to uptake rates; however, they do not predict whether or not people are capable of explicitly assessing these rates. For example, a person may move on to a different information source once the rate of finding relevant information decreases, yet they may not be capable of reporting exactly what that rate currently is, or what it previously was.

The included experiments investigate a kind of JOI that may be relevant to self-regulated learning. Instead of making a judgment of “I can’t learn this item any further”, “how much more will I learn”, or “how much more did I learn”, the participants were not asked to explicitly make judgments, but rather were instructed to try to optimize learning, which should involve strategic list switching. Students were given the opportunity to switch between two lists of word pairs, which they were learning under time pressure. They were informed to try to switch between lists and generally allocate time in such a way as to maximize the amount of items learned. The time pressure should motivate the students to think about relative learning rates, encouraging them to switch away from slow to learn items, and to focus on more quickly learnable material, which would support the proximal learning model. Frequent switching at very low JOL items that are difficult to learn, and spending more time on easier (medium to high JOL) items would show evidence in support of the proximal learning model. If switching instead occurs at easier items this may provide more support for the discrepancy reduction model, by showing further prioritization of difficult material.

Experiment 1

Judgments of improvement are relevant to between-items switching, not just judging the learning rate on a particular set of items. Despite our earlier findings that students were unable to make accurate judgments of improvement, they may be able to assess relative improvement, i.e. they notice when learning becomes more difficult. Students are presumed to study the items with the highest rate of learning (in the proximal learning model). For this experiment we examined a task switching situation, similar to Payne, Duggan & Neth (2007) who found that task-switching behavior was related to the relative rates of return. If switching follows relative rates of return, as suggested by the proximal learning model, we should find that participants choose to change lists when encountering difficult items, and that they persist when items are easier. If we do not find that switching occurs at difficult items, it may mean that either JOL accuracy is low, that participants chose to switch on easier items, or other criteria such as desiring to equate the time spent on each list. A pilot test on the materials ensured satisfactory JOL accuracy of the list, but beginning JOLs were also correlated with final performance to see if relative accuracy may be related to switching behavior.

Method

Participants. 50 participants from the subject pool at the University of California, Merced volunteered to participate for class credit.

Materials. Lists of paired associates, from the University of South Florida norms (Nelson, McEvoy & Schreiber, 1998), constructed from results of a pilot experiment. The initial study trial included 72 word pairs, and the restudy trial broke these pairs into two separate lists of 36 pairs. List construction on the restudy trial placed items in blocks of similar difficulty, so that four items in a row of easy, intermediate, or hard difficulties would be experienced.

Procedure. Participants were given a list of paired associates in an initial study trial, and were able to (and encouraged to) change lists during the restudy trial. The initial study trial presented a total of 72 items for 2 seconds each, and then solicited individual JOLs for each item. In the restudy (switching) trial, subjects were presented with one word pair at a time, and were allowed to switch to the other list, or proceed to the next item in the same list, at any point. Switching back to list A would take them to the point in list A where they had left off (so presentation will follow a set sequence for list A and list B, ensuring that no pair was viewed twice). Participants were instructed that there would not be enough time to study every word pair (they were limited to a total of 7 minutes, though they were not informed of the exact time limit). Subjects were asked to describe how and when they decided to switch lists after completing the experiment.

Results

JOLs for the items at switch points were analyzed to see if switching occurred at more difficult items, and reasons for switching were compared to data where applicable (for example, if a participant claims that they switched when the items were too hard or too easy, whether or not the data confirmed that behavior). We used paired samples t-tests for comparisons.

Study time. Average study time was 1.78 minutes, and no participants took the full amount of time available. Average recall was 44.26 out of 72 words, and the highest amount recalled was 66. No participants recalled all words, and average performance was far from ceiling.

Switching. Not all participants chose to switch lists; 23 participants did not switch at all, while 27 participants did choose to switch. Average number of switches per participant was 12.76. Switch data was not analyzed for participants with less than two switches, or if they switched on more than 65 trials (two outliers switched on nearly every trial).

Average switch JOLs were slightly lower than JOLs for other items, but not significantly so; means were 58.09 and 60.82, *t* (27) = -1.01, *p* = .32. Average recall score for switch items was also not significantly different, .56 versus .59, *t* (27) = -.58, *p* = .56. Participants most often reported switching at hard items, but this was not supported by the data.

Interestingly, though JOLs were not significantly lower, participants did spend more time on items at switch points, *MD* = 668 (milliseconds), *t* (26) = -2.19, *p* < .05. For all participants (those that switched lists and those that did not), more time was spent overall on the more difficult items, with the least amount of time spent on easy items, *F*

(2) = 3.94, $MSE = 160247$, $p < .05$, $\eta 2 = .074$. Paired samples t tests revealed significant differences among all three types of items (easy, medium, and hard). Participant reports agreed with the data, as most participants reported spending more time on the difficult items.

There was a moderate relationship between judgments of learning and actual recall performance, as average gamma was .51, significantly different from zero, $t\ (49) = 15.55$, $p < .01$

Histograms showed that the JOL distribution for items at switch points was very similar to that of regular items, as seen in Figures 1 and 2; both showed bimodal distributions. Switch items did not show consistently higher or lower JOLs than other items, demonstrating a lack of a consistent switching strategy- switches did not consistently occur at difficult items, and/or easy items.

Discussion

Switch points did not occur more frequently at more difficult items, which illustrates that participants did not follow a proximal learning strategy, though they frequently reported using that strategy. In contrast, discrepancy reduction predicts that students would spend more time on difficult items, and possibly switch away from easy items. We found that students spent more time on the difficult items, but not more frequent switching at easy items. Overall, we did not find evidence for the proximal learning model, and it is possible that participants imagined too much time pressure, and did not spend time strategizing beyond a simple discrepancy reduction strategy of "spend more time on hard items". Participants did not report confusion, so lack of strategic switching is not likely due to a misunderstanding of instructions. It is also not likely due to inaccurate judgments of learning, as they were significantly correlated with performance. Switching may be something that participants felt was unnecessary, and that their goals could be met by simply choosing to allocate more time to difficult items, and less time to easy items. It is also plausible that more optimal behaviors are actually sub-optimal, because strategy implementation may require too many cognitive resources that would simply be better spent rehearsing items. More optimal switching behaviors may only be exhibited when there are very obvious differences in difficulties between lists, or when time pressure is lifted.

Experiment 2

This experiment was identical to Experiment 1, but participants were *not* instructed that time was limited. It is of interest whether the same pattern of switching behavior will be observed (or if participants do not switch at all) when participants do not think that time is limited or that they have to strategize and spend time wisely. One potential reason for the null results of Experiment 1 is that participants imagined time pressure so severe that they felt that to fully implement a learning strategy would require too much time and cognitive effort that would be better spent on item rehearsal. If participants do not feel this time pressure, they may think there is more time that can be spent strategizing; however, it is also possible that a lack of any perceived time pressure will also result in a lack of strategic switching, because enough time is thought to be available to study each item for the necessary amount of time. If more strategic behavior is seen in this experiment, then the time pressure of Experiment 1 likely prevented strategizing; however, if strategizing is lessened, then time pressure is serving as a motivator for strategic behavior.

Method

Participants. 30 participants from the subject pool at the University of California, Merced volunteered to participate for class credit.

Materials. Identical to those used in Experiment 1.

Procedure. Identical to Experiment 1, except that participants were not instructed that time is limited.

Results

Study Time. Average study time in this experiment was slightly longer, at 2.42 minutes, and unlike before, one participant did take the full time available; *min* = .88, *max* = 7.0. Average recall was higher, at 49.6, *min* = 39, *max* = 61. Only 10 of the 25 participants chose to switch between lists, and the average number of switches was 3.17, *min* = 0, *max* = 25.

Switching. Mean JOLs were lower at switch points, means were 63 and 74.8, *t* (9) = -2.05, *p* < .05. Items at switch points were significantly more difficult, as they showed a lower proportion recalled than other items, .29 versus .69; *t* (9) = -4.11, *p* < .01. Less time was spent on items at switch points, *MD* = -1468 ms, *t* (9) = 4.96, *p* < .01.

Overall, however, more time was spent studying the (normatively) difficult items, *F* (2, 48) = 7.22, *MSE* = 778534, $p < .01$. Normative difficulty was determined based on data from associative norms (Nelson, McEvoy & Schreiber, 1998). Paired samples t -tests revealed a significant difference between the difficult items and others (time spent on easy and intermediate items did not differ).

There was a significant relationship between judgments of learning and performance, as mean gamma was .555, significantly different from zero, *t* (24) = 12.89, $p < .01$.

Histograms show that the distribution for switch items was slightly different from that of other items; switch items were bimodal, with peaks at extreme values, see Figure 3; while regular items were more clustered at higher JOLs, as shown in Figure 4. This suggests that participants chose to switch at very hard items, and occasionally at easy items, and would usually go on to the next item at very easy items.

As in Experiment 1, JOLs for the items at switch points were analyzed to see if they agreed with participants' explanations of switching preferences. Participants most often stated that they chose to switch more often at difficult pairs, and that they spent more time on difficult pairs, which does agree with switch JOLs and study time allocation. The data seem to support a modified discrepancy reduction strategy for those who showed switching behavior, with students spending more time on difficult items, but also being inclined to switch to the other list when items are too difficult. Most participants, however, did not strategize in terms of switching, but only on time allocation, spending more time on difficult items.

Discussion

Participants switched lists less often in this experiment, but those that did exhibit switching behavior appeared more strategic than participants in Experiment 1. The time pressure imposed on participants in Experiment 1 may have been overwhelming, resulting in overall more erratic study behaviors. Participants did spend more time learning the words in this experiment (2.42 versus 1.78), and recalled more words on average (49.60 as opposed to 44.32), but they were not taking the entire allotted time, or performing at ceiling. Participants followed a discrepancy reduction strategy, as they did in Experiment 1, spending more time on difficult items, yet in this experiment they also chose to switch lists at difficult items. The preference to switch lists at difficult items may suggest that participants were either trying to avoid wasting too much time on very hard items, or, they may have simply wanted to switch to the other list to take a break from the difficult items. Unfortunately, from the data it cannot be determined what exact strategy might underlie this behavior.

Experiment 3

This experiment used the same blocked lists as in Experiments 1 and 2, but in this case, the two lists differed in average difficulties of the items. One list was slightly more difficult than the other. Having different average difficulties of the lists was expected to make allocation easier for participants. The null result of Experiment 1, and the very small number of participants who exhibited list-switching behavior in Experiment 2, suggests that perceived time pressure was not a factor in preventing strategic list switching. Experiment 3 was designed to investigate the other potential explanation, which was that it may simply be too difficult to determine when switching would be beneficial. The lists were designed with different levels of difficulty so that participants would find it easier to know when list switching would be helpful (for example, switching away from the more difficult list is quite likely to result in seeing easier items next).

We expected to find more time spent on the easier list than the second list (except for those individuals who may have learned all of the easiest items during the practice phase).

Method

Participants. 31 participants from the psychology subject pool at the University of California, Merced volunteered to participate for class credit. Two were dropped due to computer problems.

Materials. Similar to those used in the previous experiments.

Procedure. Procedures were identical to Experiment 1, with the exception that the lists differed in average difficulty.

Results

Study Time. Average study time was 1.71 minutes, *min* = .40, *max* = 4.50. Mean recall was 42.75 words, *min* = 23, *max* = 62.

Switches. Average number of switches was 6.7, though only 11 participants of the 29 chose to switch between lists.

A paired samples t-test revealed no significant difference for JOLs at switch items, means were 50.8 and 57.63, *t* (10) = -.072, *p* = .24. Performance on switch items was not significantly different, .53 versus .59, *t* (9) = -.24, *p* = .41, and study time was not significantly longer for switch items, *MD* = 342 ms, *t* (10) = -1.25, *p* = .24.

Harder items were studied longer than easy and intermediate items, but not significantly so, *F* (2, 56) = 1.11, *MSE* = 10950000, *p* = .337.

Mean gamma was .485, significantly different from zero, *t* (28) = 9.77, *p* < .01, again showing a relationship between initial JOLs and final recall.

Distributions for “s” items and “n” items, as seen in Figures 5 and 6 (for the participants who chose to switch lists) revealed a bimodal distribution for the “s” items, as compared to the “n” items. This suggests that despite a lack of significant difference for JOLs of these items, participants may have been switching at items with low rates of learning, as they switched at both very easy and very hard items. The difference was more striking for Experiment 3 than Experiments 2 and 1, likely because in this experiment, participants were instructed that time was limited, but the lists differed in average difficulty, making strategic switching somewhat easier.

Discussion

Constructing the two lists with different average difficulties appeared to impact restudy behaviors, with students choosing to switch lists at both very high JOL (easy) items, and very low JOL (difficult) items, though mean JOLs were not significantly different for switch items. However, only a small proportion of students chose to switch lists; just as in Experiment 2, only a third of the participants exhibited any switching behavior. A lack of significant differences in this experiment prevents any firm conclusions from being drawn, but the distribution of values for regular items and switch items suggests that when two lists of different average difficulties are being studied, participants may adopt a proximal learning type switching strategy, where they switch at both low (very difficult) and high (already learned) JOL items.

General Discussion

These experiments were designed to investigate students' time allocation between two different study lists. Together, the three experiments do not conclusively answer how students decide to switch material; the first two experiments suggest that students generally prefer to switch at harder items, though the third experiment suggests that students might switch at both very easy and very difficult items when the two lists differ in difficulty. All three experiments revealed that in general, a majority of learners choose not to switch between lists at all, whether time is limited or not limited, and whether lists differ in average difficulty or not. Overall, more time was spent studying harder material in all experiments, a finding that is well supported in other metacognitive literature (Koriat & Goldsmith, 1996; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999).

The low number of students choosing to switch lists at all in the experiments raises a possibility that students are not attempting to be optimal in any way other than time- and effort- minimizing. Students may not have been motivated enough to perform well in the experiment, and performance may not be indicative of how time would actually be allocated between two sets of material to learn. Another explanation is that students may not have felt that to switch lists would be advantageous, and that, in fact, they could manage to learn all items within the time allowed. A third possibility is that because the two lists consisted of very similar material, differences in item difficulties may not have been as obvious as other kinds of material students might study; and because of this, the allocation decisions may require too much cognitive effort. If students were studying Spanish vocabulary and trigonometry, perhaps it would be easier for them to determine when to switch based on feelings of mental fatigue ("I'm tired of studying math now, so I will switch to Spanish"), and though such feelings may be loosely related to learning rates, these feelings likely arise only when the rate of learning drops very low. The ability to perceive small drops in learning rate and switching to other items may be beyond the capacities of most students, and to strategize in such a way may not, in fact, be optimal, as there is likely to be a cost to switching tasks.

The study choices of students did not provide support for the proximal learning model of study allocation; however, it is still possible that the proximal learning model may be descriptive of behavior in certain specific scenarios. It is known that students generally shift towards prioritizing easier items when under time pressure, known as the shift to easier materials effect (Metcalfe & Kornell, 2003; Dunlosky & Thiede, 2004; Kornell & Metcalfe, 2006), though this may be more because those items are closest to the goal state, not that the learner knows that the learning rates are higher for those items. The focus of the proximal learning model on prioritizing items based on their learning rates seems problematic, since no scenarios have yet been described under which students are able to articulate what learning rates are, or allocate time in such a fashion which serves to maximize the current learning rates.

The data appears to be best explained by the hierarchical model proposed by Thiede and Dunlosky (1999), as students generally spent more time on difficult items, and did not exhibit much switching behavior. This may occur for the same reason that the shift to easier materials effect is not observed unless items are presented together in an array (Thiede & Dunlosky, 1999), that it is quite difficult to keep goals and strategies in mind, try to execute the strategy, and learn the items being presented – it's easier to rely on simpler strategies that do not require as much cognitive resources, so that one's efforts can focus on learning the items. Data that would not be explained by the hierarchical model would be if participants chose to spend more time on medium difficulty items, or chose to switch at easy or medium difficulty items.

Overall, the preference of students to not switch lists, and the lack of clear switching preferences (in terms of JOLs), suggests that students may be unmotivated, thought switching was unnecessary, found it difficult to discern when switching would be advantageous, or were behaving as predicted by the hierarchical model. The model accounts for the cognitive difficulties encountered during self-regulated learning; learners want to dedicate the majority of their resources to the process of learning, rather than to strategizing, and are unlikely to strategize unless the situation supports a low resource mechanism for enacting a study strategy.

Figure 1. JOL distribution for regular items, Experiment 1.

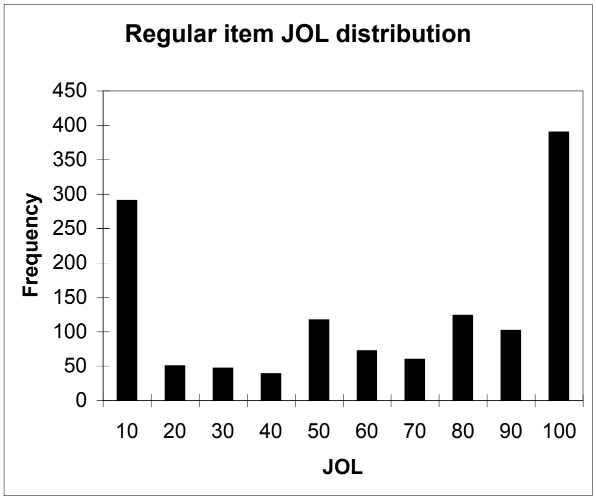


Figure 2. JOL distribution for switch points, Experiment 1.

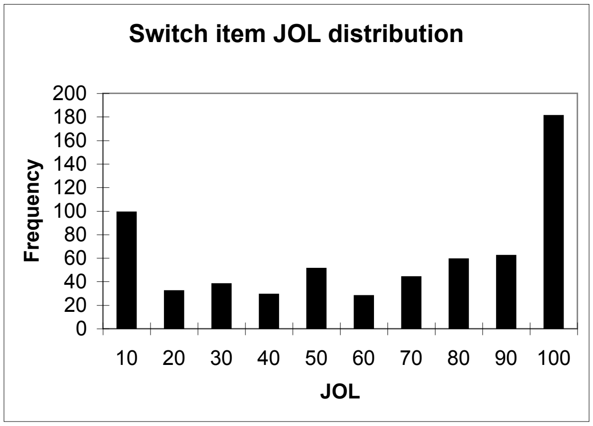


Figure 3. JOL distribution for regular items, Experiment 2.

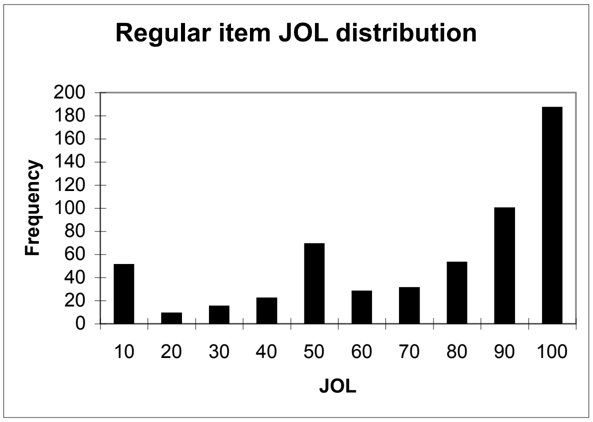


Figure 4. JOL distribution for switch points, Experiment 2.

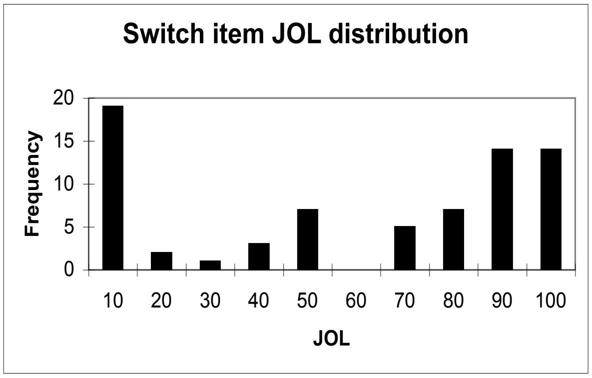


Figure 5. JOL distribution for regular items, Experiment 3.

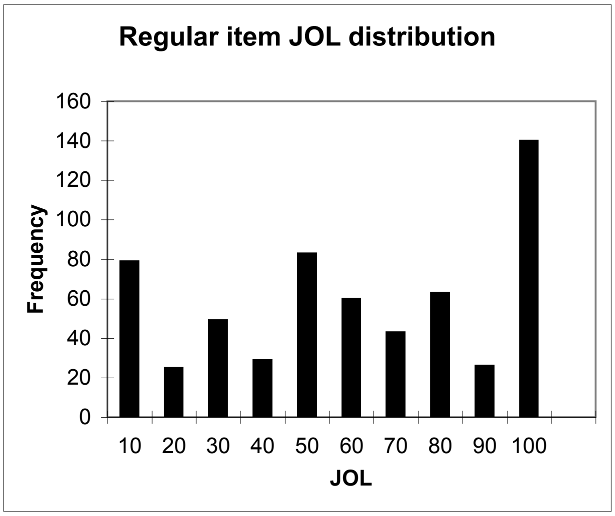
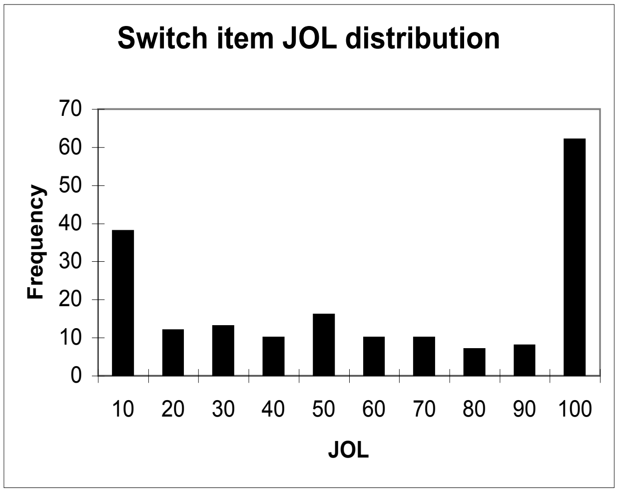


Figure 6. JOL distribution for switch points, Experiment 3.



References

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory and Cognition, 32*(5), 779-788.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219-224.

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: the benefits – and costs – of dropping flashcards. *Memory, 16*(2), 125-136.

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.

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*(4), 530-542.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*(4), 463-477.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*(4), 207-213.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the labor-in-vain effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676-686.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General, 136*(3), 370-388.

Pirolli, P. & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643-675.

Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive Science, 29*, 343-373.

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.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024-1037.

Townsend, C., & Heit, E. (2011) Judgments of learning and improvement. *Memory & Cognition* doi: 10.3758/s13421-010-0019-2

Chapter 2

Motivation and the use of Metacognition

Abstract

The impact of self-efficacy on judgments of learning, study time, and recall performance was examined in the included experiments. Self-efficacy was expected to relate to judgments of learning, as beliefs about one's abilities should be a source of information when making such judgments. It was also expected that self-efficacy would influence persistence, with higher efficacy relating to longer study times. These experiments manipulated self-efficacy, and examined effects on judgments of learning, performance, and restudy preferences. Overall, we found that self-efficacy was a source of information when making judgments of learning, but that effects on study time, performance, and restudy behavior were minimal.

Motivation and the use of metacognition

The effectiveness of a student who is studying for exams is likely to rely on several factors, including their level of motivation, the accuracy of their self-appraisals, and the effectiveness of their study strategies. More specifically, their goals, self-efficacy level, metacognitive accuracy, and self-regulatory skills all may contribute to the particular behaviors the student engages in, and their eventual success in learning the material. If a student has either a low performance goal, low motivation, does not feel capable of a task, or does not have the skills to strategize or concentrate on their learning, then less than ideal study behaviors will be chosen, and success is not likely to occur.

Self-efficacy is an individual's belief concerning how successfully one could perform some action (Bandura, 1977). Self-efficacy beliefs can be held for skills that a person currently has, or for some skill that one might consider learning. Self-efficacy can apply to any action that a person might peform, from ability to learn calculus, to ability to learn karate, drive a moving van, and so on. These judgments are informed by various sources of information, including past accomplishments, vicarious experiences, verbal persuasion (evaluations by others), emotional arousal, and interpretation of this information (Bandura, 1977). As self-efficacy describes a person's self evaluation of their capacity to perform some action, these predictions should influence the degree of effort one puts forth into those actions; time should not be wasted on a seemingly impossible task, and time is well spent attempting to do something that a person feels capable of mastering. Evidence has been found that feelings of high or low self-efficacy influence performance; Bouffard-Bouchard (1990) found that higher induced self-efficacy led to completion of more problems, and participants reported setting higher goals for themselves. Bandura and Cervone (1983) found that participants with high self-efficacy and high dissatisfaction with their current performance increased their effort and performance on an arm ergometer exercise machine much more than other groups, with the low self-efficacious participants who were not dissatisfied with their performance actually decreasing their performance. It appears to be important to find oneself capable of actually completing a task (high self-efficacy), and to be dissatisfied with current performance (via a discrepancy between current performance and goal performance).

Participants with low self-efficacy who do not feel capable of a task, yet are dissatisfied, are not likely to perform better – though the participants may feel anxiety, and those with high self-efficacy who are not interested in increasing performance will also not improve their performance.

Clearly, self-efficacy and discrepancies between performance and goals may be strong motivational factors. An additional consideration is feedback, as feedback may be necessary in certain cases to demonstrate the discrepancy between goals and current performance (Bandura & Cervone, 1983). The size of the discrepancy can also affect the motivation levels of an individual, as moderate discrepancies encourage increased efforts, while large discrepancies can be de-motivating (Bandura & Cervone, 1983), possibly because the large discrepancy indicates a slow learning rate, or reduces feelings of self-efficacy for the task.

Goals

When students set out to learn material, they generally do so with some kind of goal in mind, whether that goal is to pass an exam, fully understand and master the material, or to obtain an A grade. Goals are an important component of many models of study time allocation (Nelson & Narens, 1990; Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999; Bandura, 1993), as these models describe how students might regulate learning in attempt to reach a study goal. Study goals are often called a “norm of study” that the learner strives to meet (Nelson & Leonesio, 1988; LaPorte & Nath, 1976; Dunlosky & Thiede, 1998).

Previous research has explored the link between goals and study allocation and recall performance, finding that participants will generally perform better when given higher goals, though in these experiments, learners did not tend to actually meet these assigned goals (Koriat, Ma'ayan, & Nussinson, 2006; Dunlosky & Thiede, 2004; Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999; LaPorte & Nath, 1976). However, students report often setting goals for themselves, and to some degree the students must meet these goals. Potential explanations for the discrepancy between experimental goal achievement and real-world goal achievement are that either the higher – stakes environment of an actual class is more motivating, or the possibility of failure may inspire underconfidence, or simply that the longer time scale of a course may provide more opportunities to evaluate one's knowledge and realize the need for additional study.

Metamemory accuracy

If an individual is sufficiently motivated to achieve a learning goal, a remaining factor that may prevent them from accomplishing their goals is how accurate their self-appraisals are. If a student is striving to get an A on their upcoming exam, that student may study hard to meet the goal, but unfortunately fall short if the judgment of how well the material had been learned was incorrect. The accuracy of judgments of learning, which refers to judgments about how likely a piece of information is to be recalled, or how much information total will be recalled, shows fairly good relative accuracy – meaning that people can judge which material is more likely to be recalled. In terms of absolute accuracy, however, people tend to be biased, with initial overconfidence, and in repeated study – test experiments, underconfidence on subsequent trials (Koriat, et al., 2002; Meeter & Nelson, 2003; Koriat, 1997). Initial judgments of learning tend to be overconfident, with learners expecting to perform much higher than actual performance, and with repeated study and test trials, giving lower, underconfident judgments (Koriat et al, 2002).

Some models of study time allocation assume that learners make judgments of learning during study, deciding if items are learned well enough to stop or move on to other material (Nelson & Narens, 1990; Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 1998). Without feedback or multiple study opportunities, any ongoing judgments of learning may suffer from overconfidence; experiments that looked at goal attainment have found that when given low goals, participants learned more than required, but with high goals, learners fell short (Nelson & Leonesio, 1998; Thiede & Dunlosky, 1999; Dunlosky & Thiede, 2004). As some examples, when given a goal of learning 100% of items, participants learned 49% (Nelson & Leonesio, 1998); when given a lower goal, such as 24 of 30 items, recall still fell short at 10.5 words (Thiede & Dunlosky, 1999). A possible reason for such low goal attainment is that participants in these studies had only one study opportunity to reach the goal, and Thiede & Dunlosky (1999) had participants choose which items for restudy, so participants did not get an opportunity to restudy all the material.

Overview of Experiments

The following experiments focus on the effects of self-efficacy and motivation on learning with assigned goals. Self-efficacy was manipulated between subjects. It is possible that feelings of self efficacy may influence stopping times by influencing JOIs; learners with low self efficacy may feel their learning is slower, which may lead to feelings that the goal is less obtainable. Experiment 1a and 1b manipulate self-efficacy through a practice trial, and examine effects on judgments of learning and recall performance. Experiment 2 examines the effect of manipulated self-efficacy on judgments of learning, study time, restudy preferences, and recall performance.

Motivation to reach a particular goal may be a key component influencing persistence or giving up times. In a pilot experiment, many participants stopped studying before believing the goal would be achieved. This may occur because participants did not necessarily endorse that goal – participants may have held lower goals, or the goal of "being in the experiment for a seemingly reasonable amount of time so that I can get credit". To investigate this possibility, in Experiment 3, self efficacy was experimentally manipulated, to show whether choosing a goal results in better goal attainment, or higher performance relative to the goal (as it may be the case that the assigned goals are not necessarily accepted by participants), and what metacognitive judgments were used.

Experiment 4 consisted of surveying students in a summer cognitive science course. Students were given a survey that asked about overall goals for the course, workload, course importance, interest, self-efficacy for the material, strategy use, self-regulation, and other study habits. The students wrote down goals before each test, and predicted exam performance immediately before and after. Similar classroom surveys of self-efficacy and goal setting have found self-efficacy to indirectly influence class performance through goal setting (Zimmerman, Bandura & Martinez-Pons, 1992; Zimmerman & Bandura, 1994; Pintrich & DeGroot, 1990). Isaacson & Fujita (2006) found evidence that higher performing students had better metacognitive accuracy; the high performing students made more accurate test predictions, and were better able to choose test questions for which answers were known. Self-efficacy should be related to class performance, but it was also expected that the survey would find new relationships between performance, motivation, metacognitive accuracy, test anxiety, and reported use of metacognitive judgments and self-regulatory strategizing.

Experiment 1a

This experiment was designed to investigate the relationship between manipulated self-efficacy and metacognitive judgments of improvement and learning. Self-efficacy is a potential cue for judgments of improvement and judgments of learning; feelings of self-efficacy could also be informative for decisions concerning one’s learning, such as whether to take a particular course, feelings about how well one is learning, how well material is known, and whether to persist or terminate study. JOIs (or judgments of learning rate, or future learning, as they have been described elsewhere, are proposed as a mechanism for students’ decisions about persistence and study choice (Metcalfe & Kornell, 2003; Metcalfe & Kornell, 2005). Greater feelings of self-efficacy, just like JOIs, should correlate with longer study times, and preferential study choice (Bandura & Cervone, 1983; Bouffard-Bouchard, 1990). As these sound very similar, it may be possible that self-efficacy influences study decisions by way of informing JOIs, and potentially JOLs as well.

Method

Participants. 105 participants from the University of California, Merced subject pool participated for course credit.

Materials. A list of 50 Swahili – English translations from the Nelson and Dunlosky (1994) norms were used. A survey given to participants consisted of confidence ratings, a learning goal, and the BMIS mood scale (Mayer & Gaschke, 1998).

Procedure. The initial phase of the experiment consisted of a self-efficacy manipulation, disguised as a practice trial. Participants were in one of two conditions: high or low manipulated self-efficacy. Those in the high self-efficacy condition were given a practice list of ten normatively easier Swahili items, and were informed after the recall test that performance was better than the average participant's performance. Those in the low self-efficacy condition had a practice list of ten normatively more difficult Swahili items, and were told that performance was as good as the average participant.

Upon completion of the manipulation phase, participants completed a brief survey which asked for confidence in their ability to learn the items (self-efficacy), to concentrate, what goal the participant would set for themselves, and the BMIS mood scale (to confirm that mood was not significantly affected by the manipulation).

Participants had six study test trials, learning a set of 50 Swahili-English translations. After each study trial, participants made either judgments of learning (JOLs) or judgments of improvement (JOIs).

Scoring. Recall data were scored as correct or incorrect, with credit given for misspelled items.

Results

Participants did not exhibit significant differences in mood, *F* (1, 98) = 2.06, *MSE* = 88.16, *p* = .154, or self-efficacy, *F* (1, 102) = 0.35, *MSE* = 580.89, *p* = .56, or reported goals, *F* (1,102) = 2.42, *MSE* = 155.27, *p* = .12. Unfortunately, this shows that the manipulation was not successful, as reported self-efficacy did not differ.

Metacognitive Judgments. The main result of interest was whether or not the metacognitive judgments would be significantly influenced by level of self-efficacy. We compared mean JOIs and JOLs for the two self-efficacy conditions, and found no significant effects. Those in the higher self-efficacy conditions reported only slightly higher JOIs (*F* (1, 51) = .974, *MSE* = 286.50, *p* = .328, $\eta 2$ = .019). The difference in JOLs was not significant between the two conditions either, *F* (1, 38) = 1.904, *p* = .18, *MSE* = 1713.87, $\eta 2$ = .048.

Effects are seen, however, if instead of separating the data by experimental condition, the data is split by high or low reported self-efficacy measures, as shown in Figures 1 and 2. High reported efficacy participants felt that learning was faster, giving significantly higher JOIs, *F* (1, 51) = 4.45, *MSE* = 268.56, *p* < .05, $\eta 2$ = .08; and felt that more material was learned, reporting higher JOLs, *F* (1, 37) = 6.88, *MSE* = 1532.12, *p* = .013, $\eta 2$ = .157.

Accuracy. A 2-within subjects, repeated measures ANOVA revealed no significant difference between mean recall performance and mean JOLs, *F* (1, 38) = 0.417, *MSE* = 486.17, *p* = .52, *η2* = .011, and no interaction due to self efficacy condition, *F* (1, 38) = 3.00, *MSE* = 486.17, *p* = .091, *η2* = .073. If data is separated by self-efficacy score rather than condition, however, there is a significant interaction due to self-efficacy level, *F* (1, 37) = 10.013, *MSE* = 419.207, *p* = .003, *η2* = .213, with those in the low self-efficacy group underconfident, and those in the high self-efficacy group overconfident, as seen in Figure 1.

Gamma correlations were compared for JOIs and JOLs, to investigate whether relative accuracy of these judgments was also influenced by self-efficacy level. There was not significantly different relative accuracy for JOLs for the two self-efficacy conditions, *F* (1, 35) = .720, *MSE* = .423, *p* = .402, nor were there significant differences in relative JOI accuracy, *F* (1, 42) = 3.181, *p* = .082. However, since the JOIs showed a trend towards having higher relative accuracy for the high self-efficacy experimental condition, the data was examined for high or low reported self-efficacy and showed that participants with higher reported self-efficacy had slightly, though not quite significantly, more accurate JOIs, *F* (1, 42) = 3.56, *MSE* = .342, *p* = .066; JOLs were also still not significantly different in relative accuracy, *F* (1, 34) = 1.35, *MSE* = .335, *p* = .25.

No significant differences in recall performance were observed between the two self-efficacy conditions for JOLs, *F* (1, 39) = .01, *MSE* = 426.83, *p* = .92, or for self-efficacy level, *F* (1, 38) = .232, *MSE* = 431.18, *p* = .633.

Discussion

The self efficacy manipulation appeared to be unsuccessful between conditions, however, a median split by self efficacy score showed a relationship between reported self efficacy and metacognition. Self-efficacy was related to metacognition, not affecting the relative accuracy of metacognitive judgments, but rather being associated with underconfidence (in the case of low self-efficacy) or overconfidence (for those with high self-efficacy). There was a significant difference in both metacognitive judgments, depending on self-efficacy level, with higher reported self-efficacy associated with higher judgments of improvement and judgments of learning. This suggests that self-efficacy is related in some way to metacognitive judgments, and may be used as a source of information in making judgments of learning and improvement.

The differences in metacognitive judgments for the two self-efficacy conditions appear to support the hypothesis that self-efficacy may inform such judgments. If a person has high self-efficacy, feeling themselves quite capable of a task, one would expect that person to report higher judgments of learning, believing that they had learned more than a person with low self-efficacy, who felt much less capable of succeeding. These feelings of efficacy were also related to judgments of improvement, or how quickly participants felt they were learning the material, with participants reporting high self-efficacy estimating a faster learning rate than those with low self-efficacy. The observed relationship between self-efficacy and metacognition may simply result from both judgments being influenced by some other variable, however, and it is possible that changes in self-efficacy may not actually effect changes in metacognition.

Experiment 1b

The preceding experiment found a relationship between self-efficacy and metacognitive judgments, but the experimental manipulation was not sufficient to influence self-efficacy ratings. This experiment was an attempt to make a stronger self-efficacy manipulation, to examine causality. If changes in self-efficacy level were associated with changes in metacognitive judgments, this would support the notion that feelings of self-efficacy inform judgments about one's performance. This experiment also attempts to measure self-efficacy in a more thorough manner. Confidence ratings were also obtained on each trial. Self-efficacy items were adapted from the self-efficacy items on the Motivated Strategies for Learning Questionnaire (Pintrich & DeGroot, 1990), and the Need for Cognition scale (Cacioppo, Petty, & Cao, 1984) was added to the survey portion of the experiment.

Method

Participants. 74 students from the University of California, Merced subject pool participated for course credit.

Materials. The same stimuli as in Experiment 1A were used. Additional items were added to the survey portion of the experiment, including the Need for Cognition scale (Cacioppo, Petty, & Cao, 1984), and a self-efficacy scale adapted from the self-efficacy items on the Motivated Strategies for Learning Questionnaire (Pintrich & DeGroot, 1990).

Procedure. The experiment proceeded exactly as Experiment 1A, with two changes: the additional survey items, and slightly different wording on the positive self-efficacy manipulation. The manipulation for high self-efficacy was changed to "You did better than 91% of participants", rather than "you did better than average".

Scoring. Scoring was performed exactly as described in Experiment 1A.

Results

Participants did not differ significantly in self-efficacy, *F* (1, 71) = 1.76, *MSE* = 74.33, *p* = .19, need for cognition, *F* (1, 70) = 0.42, *MSE* = 412.46, *p* = .52, or reported goals, *F* (1, 70) = 0.28, *MSE* = 180.20, *p* = .60, but did differ significantly in mood, *F* (1, 67) = 5.74, *MSE* = 101.03, *p* < .05, with those in the easy condition having a slightly higher mood than those in the hard condition.

Judgments of Learning. JOLs were not significantly different for the two conditions, *F* (1,28) = 1.58, *MSE* = 1266.26, *p* = .22, *η2* = .053. Those reporting higher self efficacy reported significantly higher JOLs, however, just as in the first experiment, *F* (1,26) = 11.80, *MSE* = 924.89, *p* = .002, *η2* = .312. This is displayed in Figure 3.

Confidence. Overall, confidence was not significantly different between conditions, *F* (1, 61) = 0.047, *MSE* = 1607.15, *p* = .83, however, it was significantly different for those reporting higher levels of self-efficacy, *F* (1, 61) = 7.75, *MSE* = 1607.15, *p* < .01, *η2* = .113.

Judgments of Improvement. JOIs were not significantly different in this experiment between the two conditions, *F* (1, 33) = .277, *MSE* = 145.61, *p* = .602, *η2* =

.008; nor were JOIs different if separated by self-efficacy level, *F* (1,32) = 2.818, *MSE* = 136.15, *p* = .103, *η2* = .081.

It appeared that the differences in JOIs tapered off over the trials, as seen in Figure 4; essentially, the effect of the manipulation seems to have worn off over time.

Discussion

Similar results were found as in Experiment 1a, however, it appears that the change in the manipulation was made weaker, rather than stronger. Consistent with Experiment 1a, higher reported self-efficacy was associated with higher reported judgments of learning, but no differences were observed in judgments of improvement. This manipulation may have been too heavy-handed and obvious to participants, and as such resulted in weaker effects on self-efficacy, and weaker relationships to metacognitive judgments.

Experiment 2

This experiment examined the effects of self-efficacy on control processes, rather than simply examining JOIs and JOLs. Self-efficacy was manipulated to uncover its possible influence on behavior, specifically, task persistence and goal achievement. The total time students spent on task was of interest, as well performance on the recall task relative to the assigned goal. The two conditions, as in the previous experiments, were induced high self-efficacy, and induced low self-efficacy. Self-efficacy was measured to verify whether the manipulation was successful. Based on social-cognitive theory, it was hypothesized that participants with high self-efficacy would study for a longer period of time, and perform better relative to the assigned goal, overall recalling more words than those in the low self-efficacy condition.

Method

Participants. 89 participants from the subject pool at the University of California, Merced volunteered to participate for class credit.

Materials. Similar to those used in the previous experiments; Swahili-English word pairs from the Nelson and Dunlosky (1994) norms. The experiment questionnaire contained questions about study habits, beliefs about learning, mood, level of interest in the experiment, reasons for terminating the study trial, reported strategy use and metacognition. The items were adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & DeGroot, 1990), and Bandura's Self-Efficacy for Self Regulated Learning scale (Bandura, 2006), and the BMIS mood scale (Mayer & Gaschke, 1998).

Procedure. Self-efficacy was manipulated between subjects. Participants first experienced a short practice list (either easy or difficult words) and then given feedback about their performance, just as in Experiment 1.

All participants were instructed to study the words until the 80% goal could be reached. The restudy trial gave participants three passes through the list of items (word order was randomized), in which the participants could choose to study each item for as long or as short as desired. At the end of the three passes through the list, participants moved on to the test phase. Participants were reminded before each restudy pass if it was the first, second, or last study trial. After a 30 second distracter, the participants were tested on the words.

Results

Length of study time and performance on recall tests were analyzed in a between subjects ANOVA to test for significant differences. Results from the post experiment survey were analyzed, and length of study time, performance, and self-efficacy was compared. If high self-efficacy participants persist longer and perform better, this would support social cognitive theories; reported strategy use should also be higher in these participants. A manipulation check showed that self-efficacy was not quite significantly different between experimental conditions, *F* (1,70) = 2.09, *MSE* = 859.75, *p* = .15.

Study Preferences. Length of time spent on high, medium, and low JOL items was analyzed. The vast majority of students adhered to a discrepancy-reduction strategy, studying items previously given low JOLs for more time. In the second study trial, average time spent on low JOL items (from 0-33) was 5983ms, medium (from 34-65) 5675 ms, and high (66-100) 4612 ms. In study trial 3, times were 4541, 3696, and 2684 ms. Times were significantly different in study trial 2, F (2, 108) = 6.94, MSE = 4102163, $p < .01$, $\eta 2$ = .114; and in study trial 3, F (2, 100) = 13.23, MSE = 3394225, $p < .01$, $\eta 2$ = .21. There were no significant differences between conditions. Overall, participants preferred a discrepancy reduction strategy.

Study Time. Average study time (in milliseconds) was compared, and there was not a significant difference for the two conditions, *F* (1, 96) = 0.11, *MSE* = 4.034E11, *p* = .744.

Recall performance. There was no difference in mean recall for those in the low SE and high SE conditions, *F* (1, 93) = .004, *MSE* = 164.02, *p* = .952.

Judgments of Learning. Judgments of learning were compared, and did not differ between the two conditions, for aggregate, *F* (1, 76) = 1.94, *MSE* = 1515, *p* = .17, or for average individual item JOLs, *F* (1, 83) = .41, *MSE* = 1148, *p* = .41. However, if the data is compared on high or low reported self-efficacy level, there was a significant difference in aggregate JOLs, *F* (1, 74) = 6.50, *MSE* = 1455, *p* < .05, $\eta 2$ = .081, and in average individual item JOLs, *F* (1, 80) = 26.14, *MSE* = 883.93, *p* < .01, $\eta 2$ = .25. This is shown in Figure 5.

Discussion

It was expected that participants in the high induced self-efficacy condition would persist longer in studying, and would perform better relative to the assigned goal as compared to low self-efficacy participants. The analyses failed to show these results, and even the measured self-efficacy level was unrelated to recall and persistence. A partial explanation may be that students were not sufficiently motivated to reach the goal assigned in the experiment, or that motivation levels did not differ with self-efficacy.

Consistent with Experiments 1a and 1b, reported self efficacy was significantly related to judgments of learning, with those reporting higher self-efficacy giving higher judgments of learning, both for individual items, and list level judgments. Interestingly, though, average study time did not differ between conditions, or between reported self-efficacy level; while participants may have reported lower feelings of self-efficacy, the participants may have been motivated to study just as long as those with higher self-efficacy, to overcome the low performance expectations.

Reported self-efficacy level was significantly related to judgments of learning, just as in experiments 1a and 1b, but this influence on metacognition did not lead to differences in study behavior or recall performance. It is possible that students did not endorse the goal given in the experiment, even if they felt they were highly capable of reaching that goal. Alternatively, differences might be difficult to observe in an experimental situation because all participants are in the lab, provided with a study scenario, unlike when students are studying at home, where they do not feel pressure to conform to experimental demands.

Experiment 3

This experiment was identical to Experiment 2, with a modification in the efficacy manipulation stage, and participants were instructed to choose a study goal, rather than all participants assigned the goal of learning 40 of 50 words. In an effort to more strongly influence participants' expectations and feelings of efficacy, participants were asked to predict their performance in the experiment relative to an arbitrary anchor value. Those in the high self-efficacy condition were asked whether they would learn more than, less than, or equal to 45 words, while the participants in the low self-efficacy condition were asked whether they would learn more than, less than, or equal to 5 words.

This manipulation was designed to more strongly influence self-efficacy. As in Experiment 2, analyses focused on relationships between self-efficacy, judgments of learning, recall performance, and restudy behaviors. Participants with higher self-efficacy were expected to choose higher goals, give higher judgments of learning, persist in studying longer, and perform higher relative to those with low self-efficacy.

Method

Participants. 75 participants from the psychology subject pool at the University of California, Merced volunteered to participate for class credit.

Materials. Similar to those used in the previous experiments; Swahili-English word pairs from the Nelson and Dunlosky (1994) norms. The experiment questionnaire contained questions about study habits, beliefs about learning, mood, level of interest in the experiment, reasons for terminating the study trial, reported strategy use and metacognition. The items were adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & DeGroot, 1990), and Bandura's Self-Efficacy for Self Regulated Learning scale (Bandura, 2006), and the BMIS mood scale (Mayer & Gaschke, 1998).

Procedure. Self-efficacy was manipulated between subjects. In the beginning of the experiment, participants were asked to predict their performance in the experiment relative to an anchor. The experimental procedure was briefly described in the prompt: "In this experiment, you will be learning a set of 50 Swahili words. You will be presented with each word, one at a time, with its English translation. You will have three opportunities to go through the list and study each word pair for as long or as short as you like. Do you think you will be able to learn more than, less than, or equal to X words?" Those in the high self-efficacy condition had an anchor of 45 words, while those in the low self-efficacy condition had an anchor of 5 words. Participants entered their response, and then were queried as to how many of the 50 words they could learn in the experiment.

As in the previous experiments, participants experienced a short practice list (either easy or difficult words) and given feedback about their performance. Participants were then instructed to choose one of three goals, which appeared on the screen: 20, 30, or 40 words.

All participants were instructed to study the words until the chosen goal could be reached. The restudy trial gave participants three passes through the list of items (word order was randomized), in which the participants could choose to study each item for as long or as short as wanted. At the end of the three passes through the list, participants moved on to the test phase. Participants were reminded before each restudy pass if it was the first, second, or last study trial. After a 30 second distracter, participants were tested on the words.

Following the test phase, participants were asked to give their reason for ending the study session; in particular, if they really felt that they were prepared and would meet their goal, or if they stopped because of a perceived lack of improvement, boredom, or other reason. A questionnaire measured mood to check if mood was confounded with self-efficacy rating.

Results

As in Experiment 2, total study time and recall performance were analyzed in a between subjects ANOVA. Choice of learning goals was analyzed to see if those with high self-efficacy would choose higher learning goals compared to participants with lower self-efficacy. If the participants with high self-efficacy choose higher goals, study longer, and recall more words, this would support social cognitive theories. A manipulation check showed a successful manipulation of self-efficacy, *F* (1, 72) = 19.76, *MSE* = 94.19, *p* < .01.

Study Preferences. The length of time students choose to restudy items given high, medium, or low JOL items on the previous trial is indicative of what study strategy the participants were utilizing. The vast majority of students in this experiment, like Experiment 2, adhered to a discrepancy-reduction strategy, studying items previously given low JOLs for more time. In the second study trial, average time spent on low JOL items (from 0-33) was 5871 ms, medium (from 34-65) 4770 ms, and high (66-100) 4369 ms. In study trial 3, times were 3694, 3674, and 2663 ms. Times were significantly different from one another in study trial 2, *F* (2, 84) = 4.63, *MSE* = 3531497, *p* < .05, $\eta 2$ = .099; and in study trial 3, *F* (2, 82) = 13.71, *MSE* = 3621380, *p* < .01, $\eta 2$ = .251. There were no significant differences between conditions, *F* (1, 32) = 1.55, *MSE* = 3.45E7, *p* = .22.

Goals. Average goal did not differ between conditions, *F* (1, 62) = 0.92, *MSE* =

31.96, *p* = .34.

Study Time. Average study time (in milliseconds) did not differ significantly between conditions, *F* (1, 73) = 0.04, *MSE* = 2.489E11, *p* = .84.

Recall performance. Recall performance did not significantly differ between conditions, *F* (1, 73) = 0.011, *MSE* = 146.16, *p* = .92.

Judgments of Learning. Aggregate judgments of learning made after each study trial were higher, though not significantly so, for those in the high self-efficacy condition, *F* (1, 62) = 1.75, *MSE* = 2452.97, *p* = .19. These values can be seen in Figure 6.

Average individual item JOLs were also compared, and while those in the high self-efficacy condition gave higher judgments, as seen in Figure 7, this difference was not significant, *F* (1, 69) = 0.62, *MSE* = 1356.80, *p* = .44.

Discussion

In this experiment, there was evidence to suggest that manipulated self-efficacy may have some effect on metacognitive judgments, yet there were still no differences in actual self-regulation behaviors. Students still tended to study for the same amount of time, recalled the same amount of material, and preferred the same discrepancy-reduction study strategy, studying items previously given lower JOLs for longer periods of time.

The choice of experimental task may bear on the lack of differences- material was simple, participants were in an experiment, where they are expected to conform to the situation, and the task is not particularly difficult. A more informative task might be how self-efficacy influences metacognitive judgments, behaviors, and performance in a more complex learning situation. Self-efficacy may be more important when the material to be learned is conceptually complex, like material students learn in their courses. With more complex material, students must use different kinds of study strategies, and must regulate their own learning, deciding how and when to study. Experiment 4 was designed to investigate this possibility.

Experiment 4

For an ecologically valid demonstration of goal achievement and metamemory accuracy, a classroom survey was conducted in a summer cognitive science course to measure motivational factors such as grade goals, self-efficacy for learning the material, course load and other commitments, how important the course is to the student, and how interested or excited they are about learning the material. They also made predictions of exam performance. Survey items were adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich & DeGroot, 1990), and Bandura's Self Efficacy for Self Regulated Learning scale (Bandura, 2006).

Method

Participants. 31 students from a summer session course in 2009 volunteered to participate. Not all students filled in exam predictions, or survey questions, so 19 were included in the analyses.

Procedure. Students were given a survey during the first week of class, which asked about learning goals for the course (what grade the students hoped to achieve, what they hoped to learn), workload (how many classes students were taking, if also employed or generally busy with commitments), intrinsic value of the course (for their major, or for planned career, interest in course material), and self efficacy for the material. During the course, students were asked to predict test performance before and after completing each exam.

Results

The main issues of interest were whether higher performing students make more accurate test predictions, and also whether there were differences in motivational variables between high and low performing students.

Accuracy of performance judgments was calculated using the absolute accuracy index detailed by Schraw (2009), which is the squared deviation of the difference between performance and prediction. This value is useful, as it provides a measure of how much values deviate, rather than bias which can change sign from trial to trial and appear lower. Correlations were calculated between judgment accuracy and test performance, to measure if high performing students were more precise in their judgments. Correlations between each pair of variables are listed in Table 1, with significant correlations noted with an asterisk.

Correlations. Significant correlations were observed between self-efficacy for self regulated learning and grades, intrinsic value of the material and self-efficacy for self-regulated learning; between self-efficacy and intrinsic value, self-efficacy for self-regulated learning, and grades; a negative correlation between absolute accuracy and self efficacy (larger absolute accuracy values mean less accuracy); negative correlations between: anxiety level and grades, self-efficacy for self-regulated learning, and self-efficacy; and positive correlations between reported use of cognitive strategies and self-efficacy, and reported self-regulatory behaviors and anxiety. No relationships were found between other variables such as goals or workload.

Discussion

This experiment was designed to investigate the influence of self-efficacy in an ecologically valid, educational context where students learn complex material and likely have some experience and background knowledge to influence their expectations. What the correlations showed is that self-efficacy was related to self-efficacy for self regulated learning, the intrinsic value of the course, how accurate exam predictions were, and course grades. Grades were also strongly related to self-efficacy for self-regulated learning, and negatively correlated with anxiety level; those with high anxiety may be anxious because they know they are less capable of learning the material, which is then borne out by their lower grades. There is evidence to support this hypothesis, as anxiety levels were negatively related to self-efficacy, and self-efficacy for self-regulated learning, yet were positively related to reported self-regulatory behaviors. Anxious students may indeed try to compensate by engaging in study behaviors, but these behaviors may be less than optimal learning strategies. The fact that anxious students had low self-efficacy for self-regulated learning, yet reported to engage in more self-regulatory behaviors suggests strongly that the anxious students may be futilely engaging in poor study habits.

Ultimately, this experiment is not overwhelming evidence, as it is a small sample size, from one course, but it does suggest that there may be important relationships between self efficacy, self-regulatory behavior, metacognitive accuracy, and class performance. It is not clear if there is a causal relation between any of these variables, but the relationships do suggest there may be some degree of relationship. High self-efficacy may lead to increased intrinsic value for the material, as it becomes more enjoyable as one feels that it is understandable; self efficacy is understandably related to self-efficacy for self-regulated learning, as one should feel capable of engaging successfully in learning behaviors if they feel themselves capable of learning the material at all. It is unclear how self-efficacy might relate to accuracy of predictions, however.

General Discussion

In these experiments, it is evident that self-efficacy may influence metacognitive judgments, with higher self-efficacy leading to slightly higher judgments of learning, and higher judgments of improvement- those that feel more capable of the task feel that more is learned, and feel that learning is faster. There was no evidence to suggest that those with high self-efficacy study for a longer period of time, or that these participants use different learning strategies (at least not when learning simple materials such as Swahili vocabulary). This is problematic, as it is expected that changes in metacognition would be accompanied by changes in self-regulatory behaviors such as study time, and item prioritization. If such effects would be observed in an experiment, it may be necessary to have a much larger sample, or different materials that would require more involved study behaviors to achieve a high level of recall.

The difficulties in actually manipulating self-efficacy, while weakening any causal conclusions, were still accompanied by reported self-efficacy being related to metacognitive judgments. Interestingly, in Experiment 3, despite a successful manipulation of self-efficacy level, there was no observed difference in judgments of learning. The issue in question is whether or not self-efficacy influences metacognitive judgments, or if some other variable is influencing self-efficacy and metacognition; this pattern of results suggests that this may in fact be the case. If self-efficacy directly impacts metacognition, then any changes to self-efficacy should also influence metacognitive judgments. Results here showed that *measured* self-efficacy was related to metacognitive judgments, but when self-efficacy was experimentally manipulated, no such relationship was observed. It is difficult to conclude from one null result, as a much larger sample may have yielded significant effects, but it is possible that a third variable is responsible for the association between measured self-efficacy and metacognition. This factor may be past experience with similar tasks, or simply the participants' interpretation of all the sources of information that is informing self-efficacy. Though self-efficacy may have been experimentally manipulated, the interpretation held by participants may be such that the reported self-efficacy was more of a loose expectation that the participants were not particularly confident of.

Results from the class experiment, though obtained with a small sample, suggest that in different circumstances, differences may be seen between those with higher self-efficacy and lower self-efficacy. In the experimental results, those with lower self-efficacy achieved the same degree of learning as participants with high self-efficacy, yet in Experiment 4, students who showed higher self-efficacy also had higher grades in the course. They also showed higher self efficacy for self regulated learning behaviors, suggesting that these students may be more skilled in effectively strategizing and guiding their study behaviors—or that simply because they are more confident in their ability to do so, they are more likely to try to engage in successful study behaviors. In the experiments there was not opportunity for extensive self-regulation, so this may underlie the lack of significant differences. It may also be that self-efficacy is related to end performance only with more complex materials, or over a longer time course when a student has to decide whether to study or not—in the experiment, all participants are going to study to some degree.

There was also no evidence to suggest that self-efficacy is related to goal choice – even in the class survey (Experiment 4), students all chose the same goals (an “A” grade). In the case of Experiment, 3 this may be due to the experimental situation; participants are not likely to volunteer to learn more words, even if they feel perfectly capable of doing so. In terms of course goals, even though answers to the survey were somewhat anonymous, there could have been a social desirability bias coming into effect, with all students claiming to have an “A” goal, despite many of them not genuinely endorsing that goal.

The somewhat conflicting results of these experiments are difficult to draw firm conclusions from, but the results suggest a few points. Firstly, reported levels of self-efficacy may be related to metacognitive judgments, with high self-efficacy participants reporting that they had learned more (higher judgments of learning) and reporting that they are learning faster (higher judgments of improvement). Secondly, high self-efficacy may also be related to course performance, possibly due to these students engaging in more effective self-regulated learning strategies. Thirdly, manipulations to self-efficacy may not be related to performance or metacognitive judgments, suggesting that some external factor influences both self-efficacy and metacognition. Future experiments may investigate the role of other factors on self-efficacy, metacognition, and performance; self-regulatory skills may contribute to all of these variables, and students may have different levels of self-efficacy for self-regulated learning in different courses or learning tasks, feeling that they “don’t know how to study for this course” or that other material is clear in how it would be best learned.

Figure 1. Judgments of learning, and recall, Experiment 1A. Error bars represent standard errors of the means.

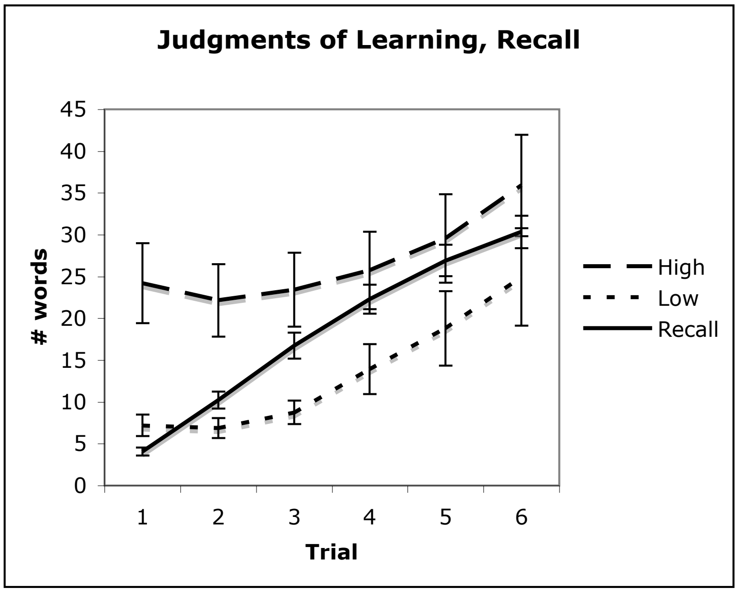


Figure 2. Judgments of improvement by self-efficacy level, Experiment 1a.

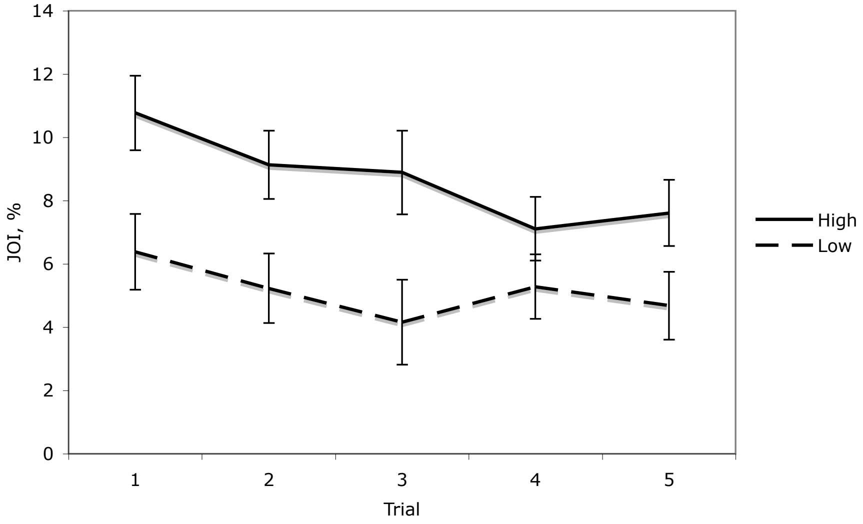


Figure 3. Judgments of learning, by self-efficacy level, and recall, Experiment 1b.

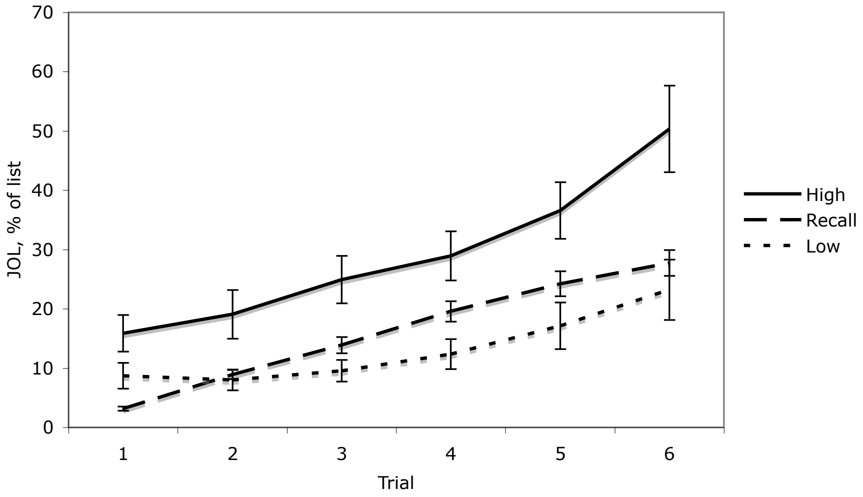


Figure 4. Judgments of improvement, by self-efficacy level, Experiment 1b.

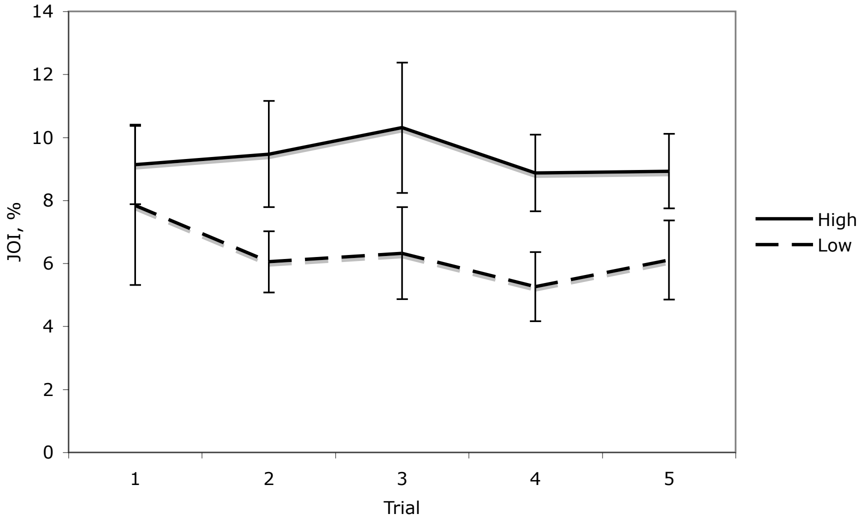


Figure 5. Judgments of learning, by self-efficacy level, Experiment 2.

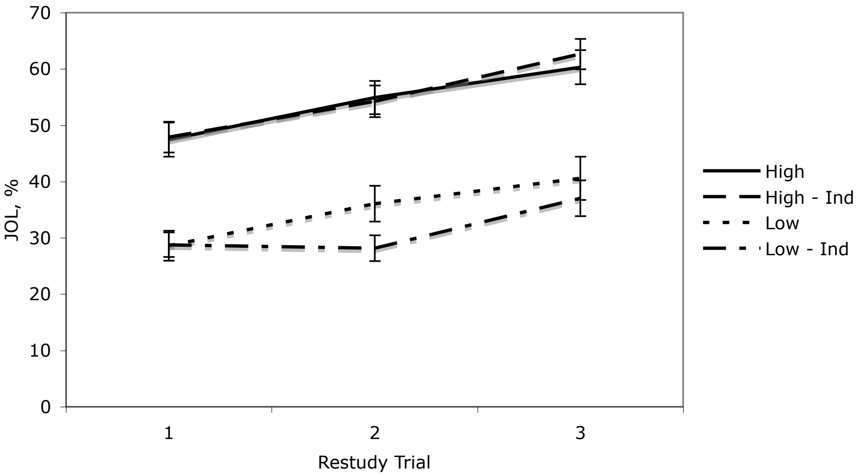


Figure 6. Average judgments of learning, high self-efficacy and low self-efficacy conditions, Experiment 3.

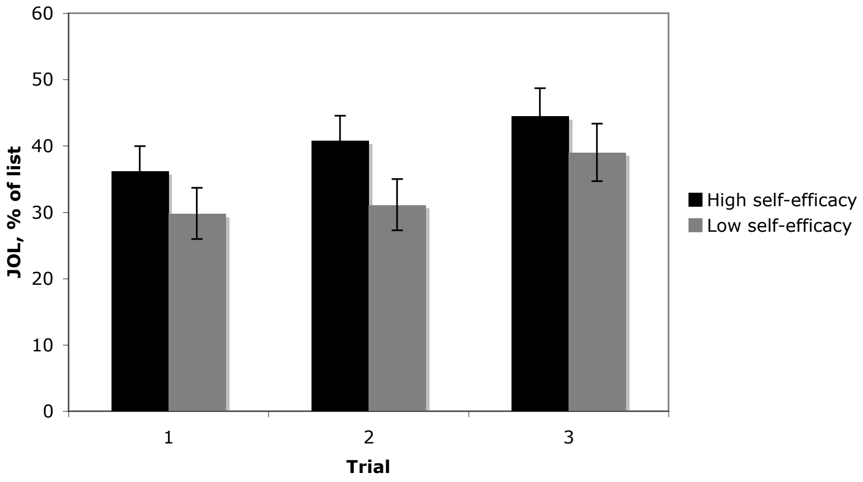


Figure 7. Average individual item judgments of learning, high self-efficacy and low self-efficacy conditions, Experiment 3.

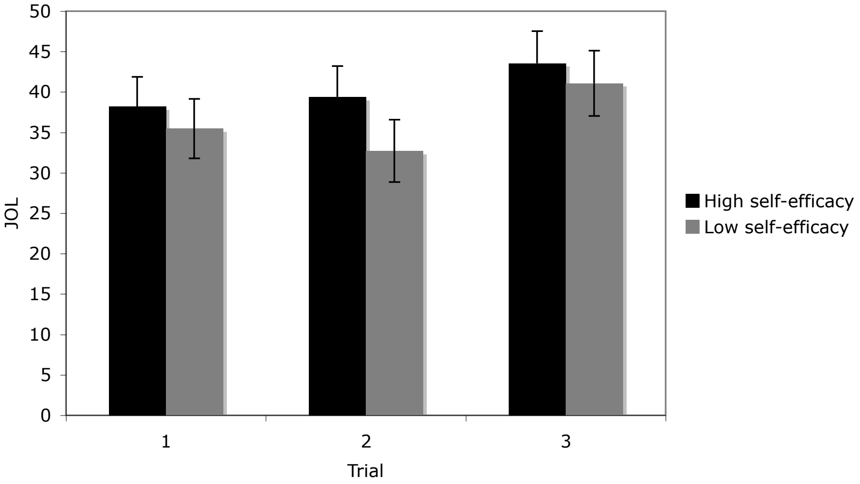


Table 1. Correlations, Experiment 4.

	Absolute Accuracy	Average Bias	Grades	Self-efficacy for self-regulated learning	Intrinsic Value	Self Efficacy	Anxiety	Cognitive Strategies	Self Regulation
Absolute Accuracy	1	-.311	-.255	-.295	-.420	-.474*	.194	-.366	-.274
Average Bias		1	-.282	.063	.283	.194	-.024	.265	.124
Grades			1	.494*	.175	.522*	-.533*	.122	.153
Self Efficacy for SRL				1	.621*	.791*	-.639*	.196	.064
Intrinsic Value					1	.647*	-.072	.391	.319
Self Efficacy						1	-.658*	.478*	.119
Anxiety							1	-.137	.471*
Cognitive Strategies								1	.380
Self-Regulation									1

References

Bandura, A. (1977). Self efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1993). Perceived self efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A. (2006). Guide for constructing self-efficacy scales *Self-efficacy beliefs of adolescents, 2006*(5), 307-337.

Bandura, A., & Cervone, D. (1983). Self evaluative and self efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology, 45*(5), 1017-1028.

Bouffard-Bouchard, T. (1990). Influence of self efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353-363.

Cacioppo, J. T., Petty, R. E., & Cao, C. F.(1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306-307.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky & A. C. Graessler (Eds.), *Metacognition in educational theory and practice* (pp 249-276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory and Cognition, 32*(5), 779-788.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self paced study. *Acta Psychologica,* 98, 37-56.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A., Ma’ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause and effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36-69.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147-162.

LaPorte, R. E. & Nath, R. (1976). Role of performance goals in prose learning. *Journal of Educational Psychology, 68*(3), 260-264.

Mayer, J. D., & Gaschke, Y. N. (1998). The experience and meta-experience of mood. *Journal of Personality and Social Psychology, 55,* 102-111.

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica, 113*(2), 123-132.

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*(4), 530-542.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*(4), 463-477.

Nelson, T. O., and Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory, 2*(3), 325-335.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the labor-in-vain effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676-686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-140). San Diego, CA: Academic Press.

Pintrich, P. & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024-1037.

Chapter 3

Framing Influences on Judgments of Learning

Abstract

Previous research has focused on what internal and external cues influence metacognitive judgments, but has failed to thoroughly explore the impact of question framing. Framing is known to influence other, varied, judgments such as product quality (Levin & Gaeth, 1988) and confidence in trivia answers (Koriat, Lichtenstein, & Fischhoff, 1980). In these experiments, students were asked to estimate their amount of learning with either positive or negative frames, and estimate the learning of themselves or other students. The results show that negative framing is associated with higher judgments of learning, potentially due to anchoring effects.

Framing Influences Judgments of Learning

Judgments of learning, or JOLs, are estimates of one's degree of learning. Aggregate JOLs are judgments of how much material is learned and will be recalled later, while individual item JOLs are estimates of how likely one is to recall a particular piece of information later. These judgments are made based on cues, such as familiarity of the material or the speed of recalling an answer (Benjamin & Bjork, 1996; Koriat, 1997; Koriat & Ma'ayan, 2005). Judgments of improvement, or JOIs, refer to estimates of one's learning rate, measuring how much more information one is likely to learn in an upcoming study trial. This judgment may be essential to decisions about further study; for example, when considering whether to continue, change strategies, or quit (Townsend & Heit, 2010, 2011). These judgments, like JOLs, are likely to be based on indirect cues.

The cue-utilization view described by Koriat (1997) describes metacognitive judgments as inferential in nature, and based on heuristics. The judgments rely on three kinds of cues: intrinsic cues, or characteristics of the item being learned that may be associated with the ease of learning that item; extrinsic cues, which are aspects of the learning situation, such as number of repetitions, and what study strategies were used; and mnemonic cues, such as ease of retrieval, familiarity, and ease of processing (Koriat, 1997). Intrinsic cues are known to influence judgments of learning; for example, related word pairs are given higher judgments of learning than unrelated word pairs (Metcalfe & Finn, 2008). Extrinsic cues include information about the learning situation, such as how many times material has been studied; more study repetitions are associated with higher judgments of learning (Koriat, 1997), though these judgments do not increase in magnitude to the extent that recall increases (Koriat, Sheffer, & Ma'ayan, 2002). Mnemonic cues include information about how quickly and easily a target is retrieved upon a recall attempt, how familiar the cue (or target) seems, and how easily material is processed during study, though this cue can occasionally be quite misleading, as Rhodes and Castel (2009) found that larger font sizes are associated with higher judgments of learning.

Another potential factor that might influence metacognitive judgments is how the metacognitive question is framed. Framing influences decisions, and decisions often change when the framing is changed (Tversky & Kahneman, 1981); when situations are framed in terms of losses, there are more risky choices, as opposed to framing in terms of gains, which shifts preferences towards risk-averse choices. Framing is also known to influence various other judgments, such confidence in one's answers to trivia questions (Koriat, Lichtenstein, & Fischhoff, 1980), perceptions of product quality (Levin & Gaeth, 1988), and evaluations of programs or issues (see Levin, Schneider, & Gaeth, 1998, for a review).

Framing something negatively generally results in lower evaluations, as seen in evaluations of products or programs, while framing in a positive light (for example, describing a 75% success rate, as opposed to a 25% failure rate) results in more positive evaluations (Levin, Schneider, & Gaeth, 1998). Similarly, focusing on why one's answer is likely to be wrong reduces the degree of overconfidence in one's answer (Koriat, Lichtenstein, & Fischhoff, 1980). On this basis, it was expected that framing metacognitive questions negatively, in terms of number of unlearned items, would show a similar effect, and show reduced performance estimates.

While framing questions negatively is likely to show lowered evaluations, it is possible that this negativity may, in fact, result in inflated estimates. Self-enhancement bias (Krueger, 1998) is an effect wherein people judge themselves more positively than an observer would rate them. This bias can occur when traits that are viewed positively by the participant are being evaluated, but not for traits that are part of negative self-views (Swann, Pelham, & Krull, 1989). By this we mean that certain traits that are important to an individual, that they feel they are particularly skilled at. Ability to learn and remember information is likely to fall under the positive self-views of a college student, so self-enhancement bias may impact judgments of learning when that ability is questioned in a negative light, just as having participants acknowledge risky behavior may trigger self enhancing personality ratings and self enhancing reports of health behaviors (Boney-McCoy, Gibbons, & Gerrard, 1999). Hence, on this basis, one might expect that negative framing would lead to self-serving biases when students make judgments about themselves, as opposed to hypothetical judgments about other students. Judgments about other students should not elicit a self-enhancing effect since the individual themselves are not being evaluated.

Inflated estimates of performance under negative framing conditions may also occur if the negative framing is interpreted as an anchor. Specifically, asking students how many of the 50 words they will get incorrect may serve to anchor judgments of learning at 50. Anchors are known to influence judgments, with high anchors associated with higher judgment values, and low anchors associated with lower judgment values (Zhao & Linderholm, 2010; Egidi & Hasson, 2009; Epley & Gilovich, 2006; Tversky & Kahneman, 1974). This effect occurs for many kinds of judgments, and whether the anchor is completely arbitrary or appears informative (see Chapman & Johnson, 2002, for a review). Anchors have specifically been shown to influence metacognitive judgments as well, such as metacomprehension (Zhao & Linderholm, 2010), multi-trial judgments of learning (Scheck & Nelson, 2005), and self-efficacy (Cervone & Palmer, 1990).

Negative framing may result in lowered estimates of performance, just as negative framing is shown to reduce varied estimates, such as: confidence in one's answers (Koriat, Lichtenstein, & Fischhoff, 1980), evaluations of programs (Levin & Gaeth, 1988), and judgments of product quality (Levin, Schneider, & Gaeth, 1998). Conversely, inflated estimates of performance may occur if students exhibit a self-serving bias (Krueger, 1998), or if the framing is interpreted more as a high anchor rather than as a negative frame. The following experiments were designed to investigate whether estimates would be increased or lowered, and if judgments are increased under negative framing conditions, whether this effect is due to self-enhancement or anchoring effects.

Overview of Experiments

The first experiment examines the effect of positive and negative framing on judgments of learning, with those in the positive condition asked how many of the words they would get correct on a recall test, while those in the negative condition were asked how many of the words they would get incorrect on a recall test. Experiment 2 investigates the possibility that distinctions between positive and negative framed judgments are due to theories of learning and remembering. This was accomplished by administering a survey to ask students to estimate learning of participants in an experiment, via positive or negative framing. The third experiment solicited item level judgments, as well as the aggregate judgments detailed in the other experiments, so that the influence of different cues could be examined. Participants were asked to make judgments for their own learning, or the learning of others, to investigate the possibility of self-enhancement. Experiment 4 was identical to Experiment 3, but with the removal of individual item judgments, as they appeared to change aggregate estimates.

Experiment 1

In this experiment, we were interested in the effects that framing might have on aggregate judgments of learning and improvement. Framing in this scenario refers to the specific way that students are asked to make judgments. For example, in terms of aggregate judgments of learning, students would typically be asked how much they *know*; they are not asked what percent of the material they do *not* know, or how many items they will get incorrect. The way in which students internally frame or conceptualize their judgments may have a profound effect on their judgment magnitude and/or calibration. On the basis of previous framing research, we expected that framing judgments in terms of how much material is *not* known, rather than how much material is known, would result in significantly lower judgments, but not impact accuracy statistics, as framing should not differentially effect trials. On the other hand, a self-serving bias would lead students to give higher judgments of learning when the question is framed negatively, because negative framing may lead students to downplay estimates of what they do not know, whereas positive judgments focus students on what they do know.

Method

Participants. 152 participants from the University of California, Merced psychology subject pool volunteered to participate for class credit. The number of participants in each condition was as follows: 13 in the positive JOL condition, 13 in the positive JOL and JOI condition, 52 in the JOI condition, 36 in the negative JOL condition, and 38 in the negative JOL and JOI condition. Sample sizes were unequal due to time constraints, and also because it was expected that negative framed judgments might show wider error variance (due to participant confusion, frustration, etc) and require a larger sample size.

Materials. The list of 50 Swahili – English word pairs was constructed from the Nelson and Dunlosky (1994) norms. These stimuli have been extensively used in previous metacognitive research. The list of word pairs was constructed in order to include a range of difficulty.

Design and Procedure. Participants were presented with the word pairs for a total of six trials. Each trial consisted of a study phase, during which each word pair was viewed in the center of the computer screen for two seconds each. After viewing all words, participants proceeded to a JOL judgment phase (if making JOLs), a test phase, and a JOI judgment phase (if making JOIs). The test phase showed participants each Swahili word, and they then typed in the English translations in response.

During the JOL judgment phase, participants made aggregate judgments of learning. Those in the negative-frame conditions were asked "How many words (out of the 50 word list) will you get incorrect (wrong) on the recall test? Your answer: I will miss ___ words on the test." Those in the regular-frame condition were asked "How many words (out of the 50 word list) will you get correct on the recall test? Your answer: I will get ____ words correct on the test." During the JOI judgment phase, participants were asked "Of the words you got incorrect (wrong) on the test, how many of those words will you learn in this study trial? Your answer: I will learn ____ words." Those in the JOL/JOI conditions made both JOLs and JOIs.

Scoring. Responses on the test trial were marked correct if they matched the target word. No points were deducted for misspellings.

Results

3 participants were removed from the analysis due to failure to learn any Swahili-English word pairs. 29 participants were removed due to either not entering any judgments, entering extremely outlying judgments, learning less than 5 words after all six trials, or technical errors. The final number of participants included for analysis in each condition was as follows: 28 in the negative JOL condition, 12 in the regular JOL condition, 11 in the regular JOL plus JOI condition, 29 in the negative JOL plus JOI condition, and 40 in the JOI condition, for a total of 120 participants.

Judgments of Learning. Figure 1 shows positive versus negative JOLs across trials, along with actual recall rates. For the sake of comparison, negative JOLs were converted to positive values by subtracting the value reported from 50, to get the number of words subjects felt that they would get correct. First, negative JOLs were compared to positive JOLs, to determine the impact of framing on judgment magnitude. Contrary to prediction, JOLs were actually higher when solicited with a negative frame than with traditional wording, $F\,(1, 38) = 5.23$, $MSE = 163.34$, $p < .05$, $\eta^2 = .121$. Collapsing across the conditions (whether or not JOLs were made in conjunction with JOIs) showed the same effect, with negative frame JOLs associated with higher values, $F\,(1,78) = 7.99$, $MSE = 249.78$, $p < .01$, $\eta^2 = .093$.

Regular JOLs were not significantly higher when made in conjunction with JOIs,

F (1, 21) = 0.68, MSE = 123.22, p = .42, η^2 = .031, nor were negative JOLs different when made in conjunction with JOIs, F (1, 55) = 1.43, MSE = 297.92, p = .24, $\eta 2$ = .025.

Accuracy of JOLs was first investigated by comparing JOLs to recall values in a 2 x 6 (repeated measures) x 2 (between subjects) analysis of variance. There was the typical underconfidence with practice effect (Koriat, Sheffer, & Ma'ayan, 2002), with mean JOL higher than mean recall on trial 1, but shifting to underconfidence in later trials. JOLs were significantly different from recall in both conditions (positive JOL and negative JOL), F (1, 38) = 11.55, MSE = 117.76, $p < .01$, η^2 = .233. There was not a significant effect of condition, F (1, 38) = 3.96, MSE = 441.41, p = .054, η^2 = .094, so this measure of accuracy did not significantly differ between the two conditions. In addition, collapsing across JOL conditions (whether or not they were made in conjunction with JOIs) shows the same effects, as illustrated in Figure 1, with JOLs being significantly lower than recall values, on average, F (1, 78) = 24.35, MSE = 142.72, $p < .01$, η^2 = .238, and no significant difference in absolute accuracy for the two framing conditions, F (1, 78) = 3.80, MSE = 472.28, p = .055, η^2 = .046.

Absolute Accuracy. Another measure of absolute accuracy, the average squared deviations between judgments of learning and recall, provides an estimate of how much judgments tend to differ from recall performance (Schraw, 2009?). The square root of this value is more interpretable, as it is essentially the average of the absolute values of difference scores between JOLs and recall. Here we found no significant differences in absolute accuracy between the two JOL framing conditions, F (1, 59) = 0.29, MSE = 38.51, p = .59. Means were 10.72 for positive framing, and 11.59 for negative framing.

Judgments of Improvement. As JOIs asked how many out of the wrong words would be learned in the trial, no score conversion was necessary. JOIs were not significantly higher when solicited in conjunction with negative JOLs, compared to when they are solicited with regular JOLs or without any JOLs; F (2, 67) = 1.35, MSE = 46.53, p = .268, η^2 = .039.

Accuracy of JOIs was investigated by comparing JOIs to actual improvements in recall between trials, in a 2 x 6 (within subjects, repeated measures) x 3 (between subjects) analysis of variance. Results show no significant difference between JOIs and actual improvement over the three conditions, F (1, 67) = .001, MSE = 24.88, p = .98, η^2 = .00, and no difference between conditions, F (2, 67) = 0.32, MSE = 42.14, p = .73, η^2 = .01.

Discussion

The key result in the first experiment was that judgments of learning were higher under the negative framing conditions as compared with regular, positive framing conditions. This suggests that participants actually seemed to feel that they learned more if they were asked to report how many words would be incorrect; this result is different than what would be expected from the framing literature. However, these findings can be explained in terms of a self-serving bias such as compensatory self-enhancement, the possibility that students use different cues for positive versus negative judgments, or that negative framing is akin to giving a high anchor for judgments (e.g. how many less than 50 will you remember versus how many more than 0).

Judgments of improvement were unaffected by the JOL framing manipulation, however, which lends some support to the anchoring and self-enhancement hypotheses, because the results appear as though learners were adding a constant to their estimates. If participants' estimates were influenced in a more complex fashion, due to relying on different mnemonic cues in the negative frame JOL condition, JOIs would likely be influenced in some way as well.

Experiment 2

For the second experiment, a survey was administered to a large sample of psychology students. The purpose of the surveys was to extend the findings of Experiment 1, and to evaluate students' general ideas about learning situations, without making judgments involving the self. If the same results appear, this would suggest an underlying explanation that is intrinsic to the nature of judgments of learning. On the other hand, if results differ when judgments are made about others, this would indirectly suggest that the results of Experiment 1 are due to a self-serving bias.

The surveys were constructed to see how the framing of questions might change responses. In this experiment, students were asked to make estimates of student learning (JOLs) and/or learning rates (JOIs) for a group of students participating in an experiment like that of Experiment 1, a multi-trial Swahili learning experiment. The four different forms were used to investigate two questions: what effect JOL framing might have on estimates, and what effect asking for both JOLs and JOIs might alter estimates, that is, are JOIs different when JOLs are made as well, and are JOLs different when JOIs are also made.

Method

Participants. 275 participants from the University of California, Merced subject pool volunteered to participate. 81 students completed survey A, 49 completed B, 91 completed C, and 54 completed D.

Materials. Four different surveys were constructed and administered in a between-subjects design. Survey A measured JOIs and positive JOLs, survey B positive JOLs, survey C negative JOLs, and survey D JOIs. Survey A asked participants to estimate for each of six study trials, of the words that are not learned, how many words students would learn during each study trial (a JOI), and how many words total they would know after each study trial (positive JOL). Survey B simply instructed participants to estimate how many words total would be known after each study trial (+JOL). Survey C solicited negative JOLs, in other words, how many words students would not know (get incorrect) after each trial (negative JOL). Survey D asked for JOIs only, of the words that are not known, how many words would be learned during each study trial. Students were also asked to indicate if they had participated in a Swahili memory experiment in the past, as this would be likely to influence their judgments of the task.

Design and Procedure. Each participant completed only one survey type. Surveys were included as part of a larger questionnaire packet for students to take home. Students were instructed to complete the surveys alone, and in a quiet place. Surveys were returned and entered a week later.

Scoring. As in Experiment 1, the negative frame JOLs were converted to positive values by subtracting the values reported from 50.

Results

22 participants were removed from analysis due to not entering judgments, misunderstanding instructions, or having far outlying judgments. Final numbers of participants for each survey was as follows: 74 for survey A, 44 for survey B, 81 for survey C, and 54 for survey D. Unequal samples were a result of many surveys not being returned.

Judgments of Learning. JOLs were significantly different among the surveys, $F\ (2, 187) = 5.30$, $MSE = 382.77$, $p < .01$, $\eta^2 = .054$, with post hoc tests revealing the difference being that survey A JOLs were greater than those of survey B; in other words, JOLs were higher when participants were also asked to provide JOIs. This also meant that survey C, which measured negative JOLs, was *not* significantly different than the JOLs in survey A or B.

JOLs also differed dependent on whether or not survey participants had participated in a Swahili learning experiment in the past, $F\ (1, 187) = 46.24$, $MSE = 382.77$, $p < .01$, $\eta^2 = .20$, with those who had done an experiment giving significantly lower JOLs.

A significant experience by survey interaction $F\ (2, 187) = 5.20$, $MSE = 382.77$, $p < .01$, $\eta^2 = .053$ revealed that the difference between the three surveys was much reduced for the participants who had experience with learning Swahili. This can be seen in Figures 3 and 4. No other comparisons were significant.

Judgments of Improvement. JOIs were not significantly different between surveys, F (1, 114) = 0.36, MSE = 52.92, p = .552, η^2 = .003. Experience with Swahili experiments showed lower mean JOIs, F (1, 114) = 5.17, MSE = 52.92, p < .05, η^2 = .043, though there was not a significant interaction between survey and experience, F (1, 114) = 0.10, MSE = 52.92, p > .05, η^2 = .002.

Discussion

This experiment did not show the difference between positive and negative framing that was observed in Experiment 1. A key difference between Experiments 1 and 2 is that rather than being asked to evaluate their own performance in a learning task, in Experiment 2, participants were asked about other students in a hypothetical learning situation. Although there are other differences between the two experiments, and given that caution is needed when inferring from a null result, the findings do not support the notion that positive versus negative framing has a general effect on JOLs that is independent of context, and due to theories of learning and forgetting. These results are consistent with the notion that framing effects may lead to a self-serving bias, so that they affect judgments about one's own learning but not the learning of others, possibly mediated by the use of different cues when making positive versus negative judgments.

Experiment 3

In order to satisfactorily examine whether framing effects are due to different cue usage, or self-enhancement, for the third experiment, we manipulated whether participants were making judgments about their own learning, or that of another student. Judgment framings were the same as in Experiment 1, with the addition of individual item JOLs: negative framed participants were asked how likely they were to get each word incorrect on the recall test, and positive frame participants asked how likely they were to get each word correct on the recall test. Individual item JOLs were included because if they are also significantly different, they allow for more detailed data analyses; influence of various cues can be calculated (of retrieval fluency, difficulty, etc) on individual JOLs, and the influence of individual JOLs on aggregate JOLs. These relationships should significantly differ between framing conditions, and in the negative framing condition, the correlations should differ depending on whether judgments are being made about oneself or another student.

Method

Participants. A total of 105 participants from the psychology subject pool at University of California, Merced, participated for course credit.

Materials. The same word pairs were used as in Experiment 1.

Procedure. The procedure was mostly the same as that of Experiment 1, with minor changes. Firstly, students were asked either to make judgments concerning their own performance, or the performance of a “student”. Secondly, item-level JOLs were solicited immediately after study of each item. As before, JOLs were framed either positively (in terms of correctness) or negatively (in terms of incorrectness).

Results

Aggregate judgments of learning. Aggregate level JOLs were not significantly different between framing conditions, *F* (1, 59) = .12, *MSE* = 511.69, *p* = .73, or between judgments made for self or another, *F* (1, 59) = .01, *MSE* = 511.69, *p* = .93, as seen in figure 6. There was no significant interaction, *F* (1, 59) = 05, *MSE* = 511.69, *p* = .83.

Individual judgments of learning. Average individual item JOLs were not significantly different between framing conditions, as illustrated in figure 7, *F* (1, 101) = 2.5, *MSE* = 2045.96, *p* = .12. They were not different for self or others, *F* (1, 101) = 0.10, *MSE* = 2045.96, *p* = .75, and there was no significant interaction, *F* (1, 101) = .32, *MSE* = 2045.96, *p* = .29.

Accuracy. A within subjects repeated measures ANOVA found no significant difference between recall and average individual item JOLs, *F* (1, 100) = 0.38, *MSE* = 995.72, *p* = .54, nor were there any significant interactions. This should not be taken as JOLs being extremely accurate, however, as these comparisons are between means; this only shows a lack of consistent bias among the subjects. There was a significant difference between aggregate JOLs and recall values, however, showing the underconfidence with practice effect, *F* (1, 59) = 4.99, *MSE* = 236.48, *p* < .05.

Absolute accuracy. Examining the absolute accuracy between framing conditions and focus conditions show a significant effect for both, as seen in figure 8;

positive frame judgments were more accurate than negative framed judgments, F (1, 102) = 11.39, MSE = 48.23, $p < .05$, $\eta^2 = .10$, and judgments were more accurate when made for oneself rather than another student, F (1, 102) = 5.23, MSE = 48.23, $p < .05$, $\eta^2 = .05$. There was not a significant interaction between these two factors, F (1, 102) = 1.59, MSE = 48.23, $p = .21$.

For average individual item JOLs, absolute accuracy was also calculated and compared between conditions; no differences in accuracy due to framing were observed, as seen in figure 9, F (1, 105) = 0.335, MSE = 32.83, $p = .56$. JOLs were significantly less accurate when made for another student, however, F (1, 105) = 7.67, MSE = 32.83, $p < .01$, $\eta^2 = .068$. There was not a significant interaction, F (1, 105) = 0.32, MSE = 32.83, $p = .57$.

Discussion

The results of Experiment 3 are inconclusive, as no differences in the magnitude of either aggregate or individual item JOLs were observed for either condition. There were slight effects on the accuracy of these judgments however, with participants being more accurate when making positive framed judgments, and when making judgments about their own performance.

Judgments of learning may not have been significantly different between framing conditions for numerous reasons, such as participant fatigue, as many had to be excluded from data analysis for noncompliance (not entering judgments, and/or not attempting recall). Another potential explanation is that adding the individual item JOLs influenced participants to make more informed judgments based on mnemonic cues, as their attention was drawn to the recall probabilities of the individual items. Since Experiment found significant differences between framing conditions when only making aggregate judgments of learning, yet Experiment 3 failed to show such differences in JOLs between framing conditions when both individual item and aggregate JOLs were solicited, it is likely that the addition of individual item JOLs is behind the lack of significant differences.

Experiment 4

In this experiment, we removed the individual item JOLs from the experiment, and shortened the experiment to three study-test trials due to concerns about participant fatigue (several participants were removed from analyses in Experiment 3 due to not entering judgments). If the lack of significant differences in Experiment 3 were due to the individual item JOLs causing participant fatigue, or increased reliance on mnemonic cues, removing the individual item JOLs should result in a significant difference being observed between positive and negative aggregate JOL framing conditions.

Method

Participants. A total of 70 participants from the psychology subject pool at the University of California, Merced participated for course credit.

Materials. The same materials were used as in Experiment 1.

Procedure. Identical to Experiment 3, with minor changes: individual item JOLs were removed from the procedure, and the experiment was shortened to three study test trials. A previous pilot experiment found that experiment length did not affect metacognitive judgments; so shortening the experiment should only influence compliance, and not actual judgment magnitude.

Results

Judgments of Learning. Judgments of learning for the negative frame condition were converted to positive values by subtracting from 50 (e.g. answering that you would get 45 incorrect equates to a value of getting 5 correct).

Judgments of learning were significantly higher for those in the negative frame condition, F (1, 67) = 9.43, MSE = 143.90, $p < .01$, $\eta^2 = .123$. The object of predictions (oneself or another student) did not significantly impact predictions, F (1, 67) = 0.49, MSE = 143.90, $p = .42$. There was no significant interaction, F (1,67) = 0.49, MSE = 143.90, $p = .49$.

A comparison of JOLs to recall values reveals no significant difference overall between JOLs and recall values, F (1, 65) = .021, MSE = 69.57, $p = .89$. There was a significant interaction with frame, however, with negative frame judgments being somewhat less biased, F (1, 65) = 5.32, MSE = 69.59, $p < .05$, $\eta^2 = .076$.

Absolute accuracy. The absolute accuracy statistics were not significantly different between framing conditions, as seen in figure 11, F (1, 79) = 2.13, MSE = 27.93, $p = .15$; nor were they different between judgment focus conditions, F (1, 79) = 0.20, MSE = 28.61, $p = .65$.

Discussion

In Experiment 4, we replicated the main result of Experiment 1, which is that negative-framed JOLs were significantly higher than positive framed JOLs. There was not a significant difference in the judgments made for oneself or for another student, suggesting that there is not a consistent self-enhancement effect that occurs when estimates are framed negatively. There were also not significant differences observed in absolute accuracy; between the extreme overconfidence of the initial negative frame JOL and the severe underconfidence of later, regular JOLs, both judgment framing conditions tended to differ from actual performance by similar amounts. Adding more trials to the experiment would likely show negative frame judgments to be more accurate overall, however, as these judgments would likely continue to suffer from less underconfidence in later trials.

General Discussion

The overall result of the experiments was that when presented with negative frames, participants gave higher judgments of learning. Different hypotheses were considered concerning the effects on JOLs, such as different cue utilization, self-enhancement, and anchoring effects.

In terms of cue utilization, the participants may have focused on slightly different mnemonic cues. For example, when making a positive, aggregate JOL, participants may consider how fluently the items were processed, how familiar the items seemed, and what study strategies they used. Negative aggregate JOLs may instead focus participants to think about different factors, such as how many of the items seemed difficult, unfamiliar, or were not studied well. When thinking in terms of these cues, students may not have a sense that a lot of items fall under these situations, and thus have inflated performance estimates (via low estimates of the number of incorrect items). We were not able to examine the relationship between retrieval fluency, or previous trial JOLs on judgments, as adding individual item JOLs to Experiment 3 eliminated the framing effect.

Differential cue utilization is not likely to underlie the differences between positive and negative JOLs, however, as these judgments did not differ in absolute accuracy, nor were judgments of improvement influenced in Experiment 1 by negative framing of judgments of improvement. If different cues are being accessed to make these metacognitive judgments, it is likely that there would be changes to metacognitive accuracy, and to perceptions of improvement. This suggests that it is unlikely that JOL framing changes the mnemonic cues that learners use when inferring their JOLs; these cues would likely influence JOIs as well. The lack of change in JOIs also reflects the (roughly) parallel slopes of the JOL curves seen in Figure 1, illustrating the bias shift. It is still possible that mnemonic cues underlie the shift in JOLs, but self-enhancement is another possible mechanism that would account for the lack of change in JOIs, but a self-promoting shift in JOLs.

Self-enhancement was proposed as a possible mechanism for the increased JOL estimates under negative framing conditions, however, the lack of difference between judgments made for oneself or for another student in Experiment 4 show that this explanation is unlikely. Self-enhancement would result in inflated estimates for one's own performance, but not for other students as well.

Though we found evidence that negative framing affected the JOLs of the material being learned, the effect was not predicted from the framing literature; it seemed that students believed they had actually learned more words. This data is also in contrast with the results of Finn (2008), who found less overconfidence (i.e. lower JOLs) when *individual* JOLs were made in terms of *forgetting*. Our findings also appear to be in opposition to typical findings found in attribute-framing experiments (Levin, Schneider, & Gaeth, 1998), which show overall less favorable evaluations with negative frames.

Anchoring effects appear to be the most plausible explanation behind the increased JOL estimates. Asking students how many of the items they will get incorrect out of 50 may inadvertently anchor their judgments of learning at 50. Asking students how many items they will get correct out of 50 may not elicit the same anchoring scenario, however, as it is a common judgment to make, and students may instead be relying on their own internal anchors, as proposed by Zhao and Linderholm (2008) in their anchoring and adjustment model of metacomprehension judgment.

Further experiments will attempt to reconcile these findings, however, and also look at restudy choice. It is possible that though it *appears* that participants are giving more favorable evaluations of their performance in the negative frame situation, they may attend more to the amount *not learned*, whereas in the positive frame they may attend more to the amount *learned* and thus they may have a more positive perception of their performance in a positive-JOL situation, and more pessimistic evaluations of their performance in the negative-JOL scenario. If this is the case, they may actually restudy more when making JOLs in the negative frame.

In terms of educational implications, what these results may show is that focusing on the number of errors, or the amount *not* learned, may result in more optimistic, inflated self-assessments when making performance predictions. This may be counterproductive, and encourage less time studying than is necessary, especially if students do not self-test (Finn & Metcalfe, 2007), as their JOLs will reflect the more overconfident trial 1 JOLs. Students are likely to judge their degree of learning when studying (and not explicitly asked to make JOLs in any way), but how accurate those judgments are is likely to depend somewhat on how students conceptualize the issue; people likely have their own question framing that they use to approach the topic of how well they have learned the material. Some students may ask themselves how many items they know well, while others may consider how many items they do not know well. Some students might gauge their learning by thinking about how long they have been studying, and still others could consider how likely they are to forget material on the exam. Future experiments will address whether anchoring or cue utilization underlies the effect of inflated negative-frame JOLs, and examine the impact of JOL framing on restudy preferences and recall performance.

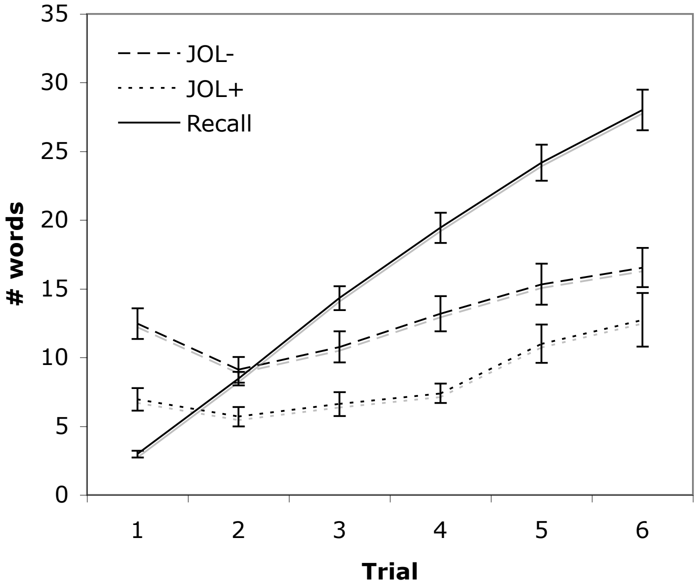


Figure 1. Mean judgments of learning and recall performance, Experiment 1.

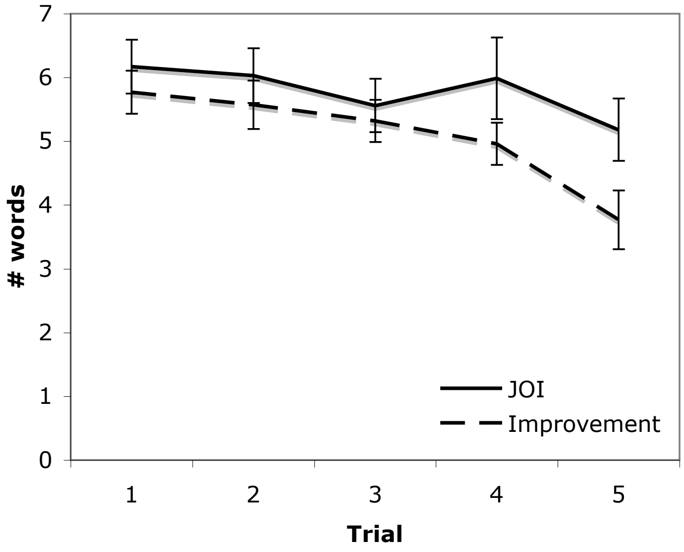


Figure 2. Mean judgments of improvement and recall improvements, Experiment 1.

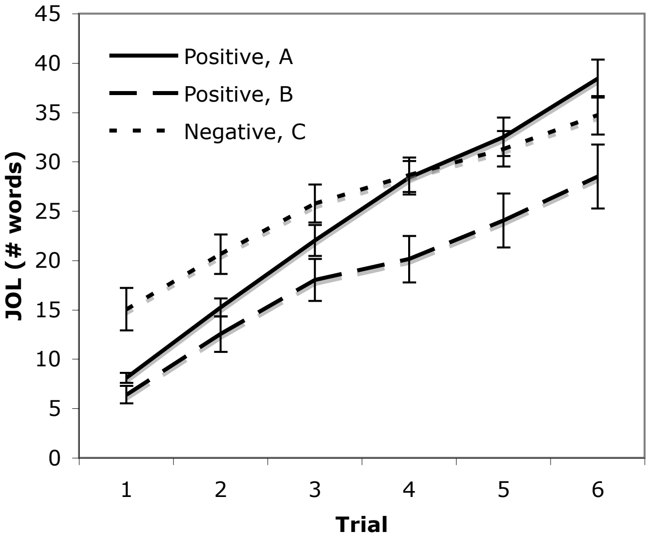


Figure 3. Mean judgments of learning by survey condition, no Swahili experience.

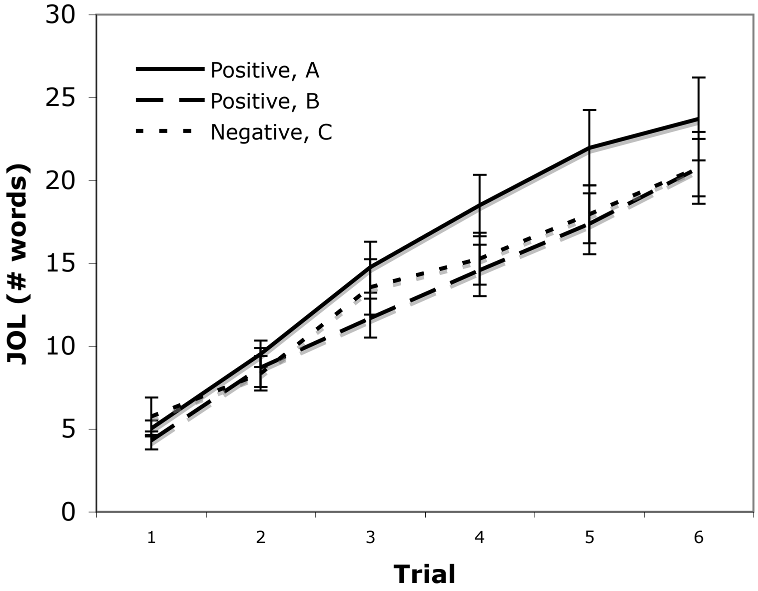


Figure 4. Mean judgments of learning by survey condition, with Swahili experience.

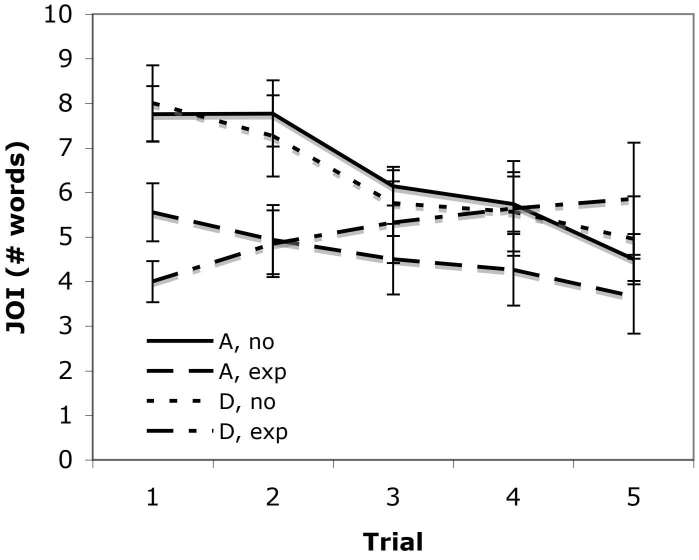


Figure 5. Mean judgments of improvement by survey condition and Swahili experience, Experiment 2.

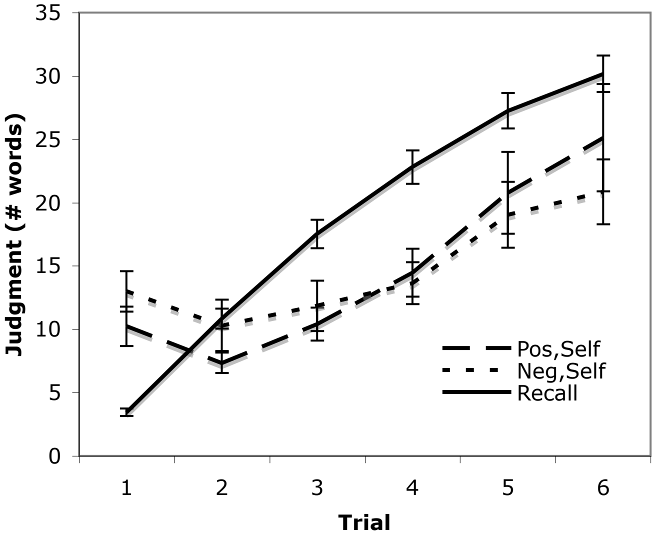


Figure 6. Mean aggregate judgments of learning by framing, compared to recall, for Self judgments, Experiment 3.

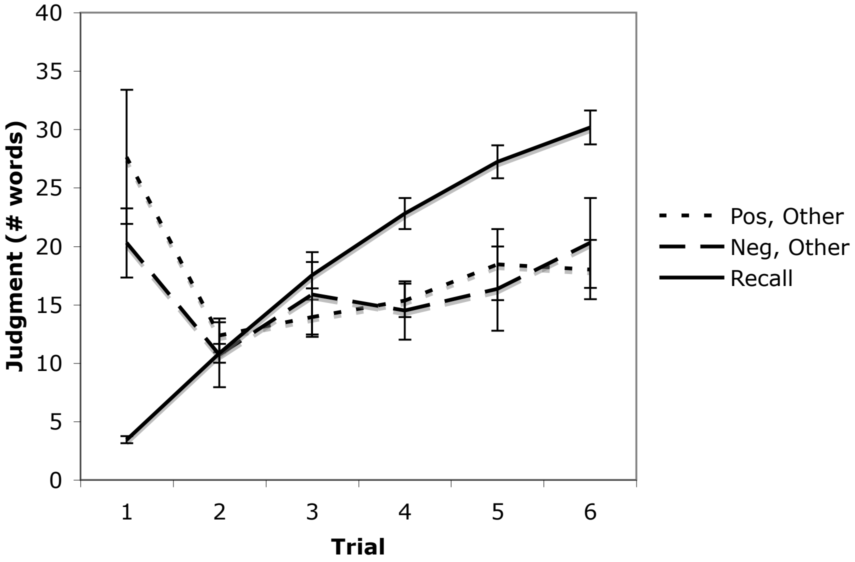


Figure 7. Mean aggregate judgments of learning by framing, compared to recall, for Other judgments, Experiment 3.

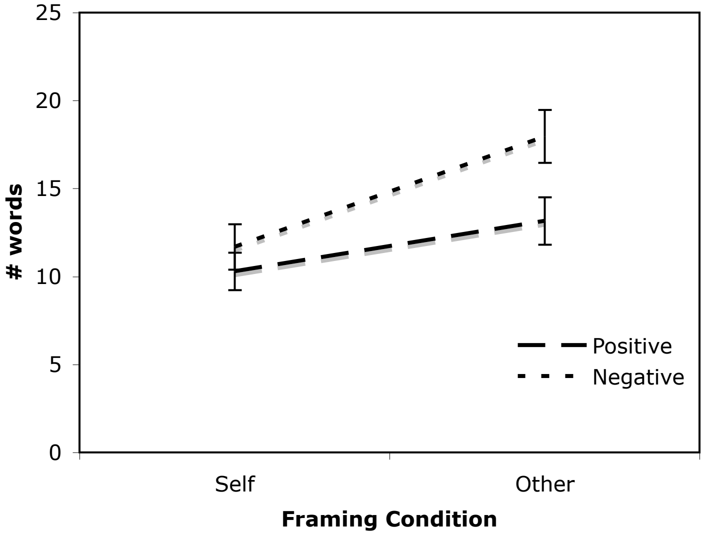


Figure 8. Absolute accuracy of aggregate JOLs by framing condition and judgment focus, Experiment 3.

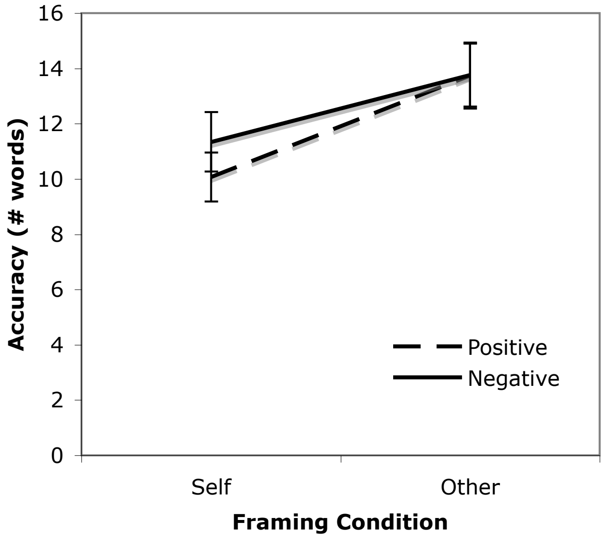


Figure 9. Absolute accuracy of individual item JOLs by framing condition and judgment focus, Experiment 3.

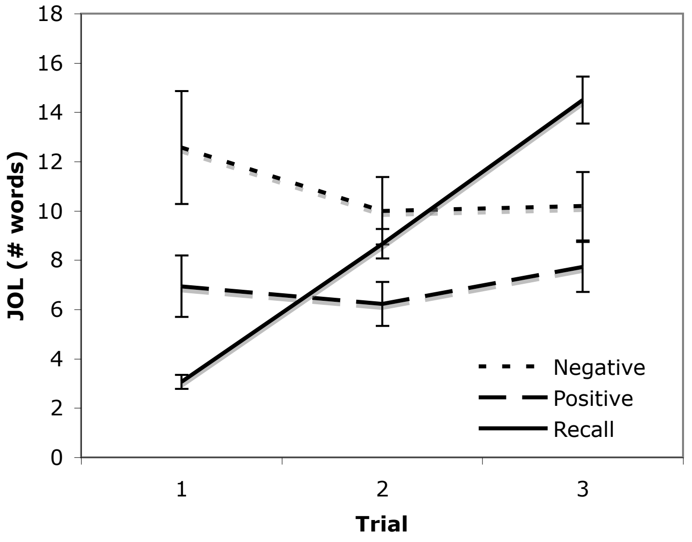


Figure 10. Judgments of learning by frame, Self judgments, Experiment 4.

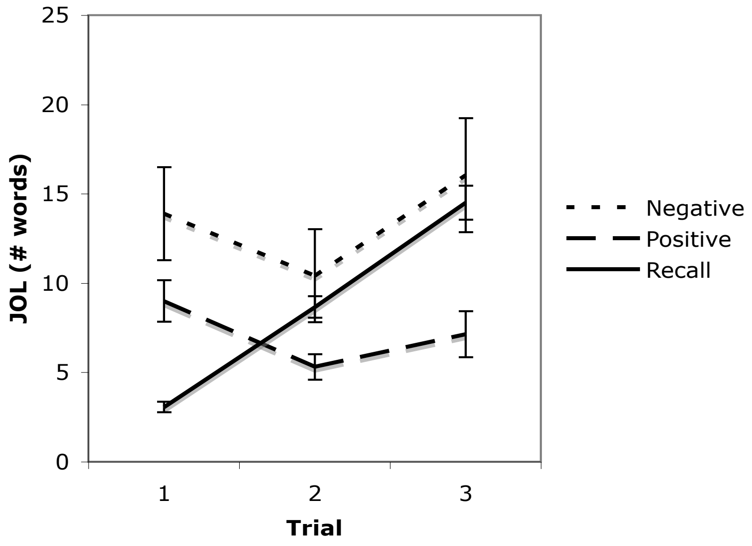


Figure 11. Judgments of learning by frame, Other judgments, Experiment 4.

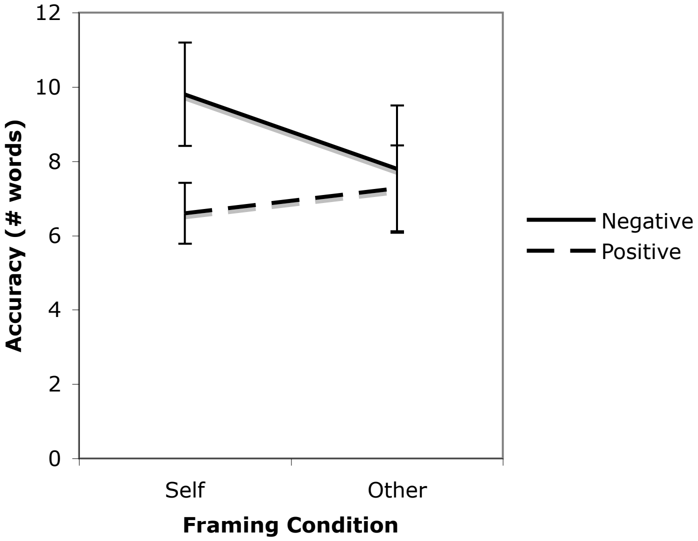


Figure 12. Absolute accuracy by framing and judgment condition, Experiment 4.

References

Benjamin, A., & Bjork, R. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit Memory and Metacognition*. Mahwah, NJ: Erlbaum.

Boney-McCoy, S., Gibbons, F., & Gerrard, M. (1999). Self-esteem, compensatory self-enhancement, and the consideration of health risk. *Personality and Social Psychology Bulletin, 25*(8), 954-965.

Epley, N., & Gilovich, T. (2006). The anchoring and adjustment heuristic: why the adjustments are insufficient. *Psychological Science, 17*(4), 311-318.

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory and Cognition, 36*(4), 813-821.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 238-244.

Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General. 126*(4), 349-370.

Koriat, A., Lichtenstein, S. & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory. 6*(2), 107-118.

Koriat, A., & Ma’ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*(4), 478-492.

Koriat, A., Sheffer, L., & Ma’ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147-162.

Krueger, J. Enhancement bias in descriptions of self and others. *Personality and Social Psychology Bulletin, 24*(5), 505-516.

Levin, I., & Gaeth, G. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*(3), 374-378.

Levin, I., Schneider, S., & Gaeth, G. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organiational Behavior and Human Decision Process, 76*(2), 149-188.

Metcalfe, J. & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology, Learning, Memory, and Cognition, 34*(5), 1084-1097.

Nelson, T., & Dunloskly, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, *2*(3), 325-335.

Rhodes, M., & Castel, A. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615-625.

Townsend, C. & Heit, E. (2010). Metacognitive judgments of improvement are uncorrelated with learning rate. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd annual conference of the cognitive science society. Austin, TX: Cognitive Science Society.

Townsend, C. & Heit, E. (2011). Judgments of learning and improvement. *Memory &*

Cognition doi: 10.3758/s13421-010-0019-2

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,*453-458.

Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychological Review, 20*(2), 191-206.

Zhao, Q., & Linderholm, T. (2010). Anchoring effects on prospective and retrospective metacomprehension judgments as a function of peer performance information. *Metacognition and Learning,6*(1)*,* 25-43.

Chapter 4

Study Scenarios and Metacognitive Judgments

Abstract

The experiments were designed to investigate the effects of repeated testing on metacognitive judgments. Repeated testing was compared with two alternative study methods, spaced practice and massed practice. Repeated testing was associated with significant underconfidence in both judgments of learning and judgments of improvement, while spacing with trivia rather than taking recall tests led to slight overconfidence. Massed study was associated with the highest judgments of improvement, but judgments of learning were no different from those in spaced practice. Results suggest that during repeated testing, metacognitive judgments become more accurate, and more underconfident, partly due to the experience of testing, and partly due to spacing of study trials.

Study Scenarios and Metacognitive Judgments

The question of how students choose to allocate their study time is a complex problem, and many studies have examined this question. Numerous published studies detail information on the accuracy of students' judgments of learning, how long students choose to study items, and if this systematically varies with judgments of learning, and the relative effectiveness of self-allocated versus externally allocated restudy. It is known that study allocation is indeed influenced by judgments of learning (Metcalfe, 2009), and in general students will choose to study difficult, low – JOL items for longer than easier, more well learned, high – JOL items (Son & Metcalfe, 2000) that more accurate metacognitive judgments lead to more effective restudy (Thiede, Anderson, & Therriault, 2003; Kornell & Metcalfe, 2006; Metcalfe & Kornell, 2005; Thiede & Dunlosky, 1999), and that students are likely to think about how well they know material as they are learning (Metcalfe, 2009; Metcalfe & Finn, 2008).

Study time allocation is influenced by judgments of learning (Metcalfe, 2009), but a problem is that such judgments may change dramatically over a study session. Ideally, judgments of learning should reflect the increases in how likely the information is to be remembered, but aspects of the study situation (or experimental situation) may influence judgments significantly, such as whether study sessions are massed together, spaced apart, whether recall is attempted in between study sessions, and so on. Metacognitive experiments that investigate metacognition over an extended study session often implement repeated testing, so that accuracy of judgments over time can be examined (Koriat, Sheffer & Ma'ayan, 2002; Serra & Dunlosky, 2005; Tiede, Derksen, & Leboe, 2009; Koriat, Ma'ayan, Sheffer, & Bjork, 2006; Koriat, 1997) This practice is informative in terms of how metacognition is influenced by this particular study practice, but it may not be informative of metacognitive judgments in general, as test trials themselves may significantly alter judgments (Finn & Metcalfe, 2008; Finn & Metcalfe, 2007). The test trial itself provides feedback to the learner that is not acquired by experiencing repeated study trials; students experience how easily the material is recalled, and if they recalled anything at all. Though some students may engage in some degree of self-testing, many students do not self-test (Karpicke, Butler, & Roediger, 2009). Among those that do engage in self-testing, most are unlikely to *repeatedly* self-test, so this particular type of experimental situation may not be informative of how accurate students' judgments really are over a typical study session. Rather than self-testing repeatedly, students may be more likely to engage in other kinds of study strategies. They may space their learning, either due to distractions, or moving between different materials. Students may also study via massed practice, studying material repeatedly without any self-testing or spacing of materials. Since students are likely to engage in different kinds of study, it would be useful to examine the effects that these different situations may have on metacognitive judgments.

Repeated Testing and Judgments of Learning

When soliciting judgments of learning in a repeated study – test procedure, the underconfidence with practice effect is generally observed (see Koriat, Sheffer, & Ma'ayan, 2002, for a review), where judgments shift from overconfidence on the initial trial, to subsequent underconfidence on all further trials. This effect is robust, and occurs for both aggregate and individual item JOLs, as well as under many different experimental manipulations (Koriat, Sheffer, & Ma'ayan, 2002). The underconfidence with practice effect may be caused by memory for past test (Finn & Metcalfe, 2008; Finn & Metcalfe, 2007), which will only be relevant in a repeated testing scenario. The memory for past test explanation essentially states that people remember their performance on the prior test, and use this information when making judgments of learning; for example, when making a JOL for the pair ORANGE – PEEL, one might remember that on the previous recall test they were able to remember “PEEL”, and as such, assign that item a high JOL. When studying another pair, such as PAINT – CEILING, perhaps the participant remembers that on the prior test they could not recall an answer, and assigns this pair a low JOL. Alternatively, Tiede, Derksen & Leboe (2009) suggest that increases in JOLs over trials only occur due to fluency cues, and that participants fail to appreciate that repeated study presentations benefit recall (though they do not address the possibility that participants fail to understand that the repeated testing is improving recall as well).

Testing may influence judgments of learning, but in addition, it also alters recall performance. Testing has been shown to improve memory, in terms of retention of information (Roediger & Karpicke, 2006a; Roediger & Karpicke, 2006b). Experiments with free recall of prose passages found that students who experienced extra study trials in place of tests had higher confidence levels in their performance relative to those who experienced repeated tests (Roediger & Karpicke, 2006a). In these particular experiments, students in the repeated testing condition experienced one study session and three test trials, rather than having study – test cycles, but see Roediger & Karpicke, 2006a for a review of testing effects across more varied experimental conditions. Though testing is associated with improvements in the relative accuracy of JOLs, it is associated with less confidence, as students who experienced extra study trials in place of tests had higher confidence levels in their performance relative to those who experienced repeated tests (Roediger & Karpicke, 2006b). Overall, the experimental evidence shows that repeated study – test cycles are associated with underconfidence; test trials may improve memory, but they seem to reduce estimates of performance.

In the generation effect, learners must generate the target of a cue-target pair, with only a few letters filled in (for example, CAT – M_ _ _ E). In contrast, the generation effect, which is related to the testing effect, also results in greater amounts of learning (Hirshman & Bjork, 1988), but may be associated with (appropriately) higher judgments of learning (Mazzoni & Nelson, 2005). Judgments of learning are likely higher in the generation task because it provides a very easy recall experience, as pairs usually rhyme, or are semantically related.

Spacing and judgments of learning

Students seem to underestimate the amount of learning that takes place over time in repeated study – test cycles, and they may also underestimate the benefit of spaced practice. Students gave judgments of learning that were higher after massed practice than for spaced practice when learning a motor task, suggesting that participants believed that more learning took place when massing practice (Simon & Bjork, 2001). Other research has found no significant difference between JOLs after massed or spaced study of word pairs (Dunlosky & Nelson, 1994). Participants in this experiment may not have believed that more learning took place when practice was massed, but they still failed to realize that performance was worse in that condition.

Although students may not understand that spacing is a more effective strategy than massing study repetitions, they frequently choose to space presentations of items. Pyc & Dunlosky (2010) examined how participants decide whether to space or mass practice of items, and found that participants generally chose to space items, preferring to mass presentations only for difficult items when the presentation times were very brief. This is not to say that they are strategic because they understand what is optimal; the authors point instead to a non-metamemory mechanism, that students choose to space presentations because they do not feel it necessary to restudy an item that was just studied (unless that item was very difficult, and it could not be processed within one presentation).

Students may often space material, but do not have sophisticated metamemory knowledge in regards to the benefits of spacing on memory performance (Pyc & Dunlosky, 2010). Massed presentations may be associated with similar (Dunlosky & Nelson, 1994), or higher judgments of learning when compared to spaced practice (Simon & Bjork, 2001), potentially due to increased feelings of fluency. Repeated testing is associated with underconfidence (Koriat, Sheffer, & Ma'ayan, 2002), as students may rely on memory for prior test performance when making judgments of learning (Finn & Metcalfe, 2007), and students may also discount the effects of testing and/or repeated presentations on memory (Tiede, Derksen & Leboe, 2009).

Overview of Experiments

The following experiments were designed to investigate the effects of repeated testing on metamemory judgments, and to what extent metamemory judgments are affected by mere spacing of study (as testing also spaces the study sessions apart), as compared to non-spaced (massed) list repetition. The first experiment compares a repeated study – test experimental condition to a condition where participants are asked trivia questions rather than taking a recall test. This particular comparison was designed to investigate the possibility that as testing may alter judgments partially because the test trial serves as a gap between study repetitions. The second experiment compares massed repetitions of the study list with trivia – spaced study, to determine if massing of study repetitions is associated with increased judgments of learning. The third experiment investigates the effect of massed practice on judgments of improvement. Repeated testing should result in the lowest metamemory judgments, while spaced practice results in higher judgments, and massed practice is expected to show even higher judgments that are significantly overconfident.

Experiment 1

In this experiment, we examined the effect of repeated testing on judgments of learning and improvement, by comparing a repeated study – test condition to a condition with general trivia questions in between study trials (lasting the same time as the average test session in previous experiments). We hypothesized that removing the intervening tests would result in overconfident JOLs, rather than showing the typical underconfidence with practice effects that are seen in repeated study – test cycles. Without the test sessions to provide information about recall, and recall latencies, judgments cannot be informed by memory for a prior recall test, and judgments are unlikely to become more accurate over trials.

Method

Participants. A total of 115 participants from the University of California, Merced subject pool volunteered to participate.

Materials and Procedure. A set of 50 Swahili-English word pairs from the Nelson and Dunlosky (1994) norms were used. In this experiment, there were 6 total conditions; 3 (judgment type) x 2 (test or trivia). Participants either made judgments of improvement (JOIs), aggregate judgments of learning (JOLs), or both individual-item JOLs and aggregate JOLs. Participants experienced either intervening recall tests or intervening trivia tests. Trivia sections were timed to take as much time as the average test for each trial.

Scoring. Word recall was scored according to whether responses matched the target words. No points were taken off for misspellings.

Results

Judgments of Improvement. First we compared judgments of improvement between the repeated testing condition and the repeated trivia condition. No difference was found between JOIs, F (1, 22) = 0.35, MSE = 8.59, p = .47, $\eta 2$ = .016, as seen in Figure 1.. This is interesting, as actual learning did differ; end recall was significantly different between conditions, with those in the repeated testing condition learning more words (23.6 versus 9.05); F (1, 35) = 18.98, MSE = 111.57, p < .01, $\eta 2$ = .352.

For the repeated testing condition, it was possible to compare JOIs to actual recall improvements. A within-subjects repeated measures ANOVA revealed that JOIs were underconfident, significantly lower than actual learning, *F* (1, 11) = 11.19, *MSE* = 7.01, *p* < .01, *η2* = .504.

Judgments of Learning. For participants who made only aggregate JOLs, JOLs were *slightly* higher, though not significant, when participants experienced trivia, *F* (1, 22) = 1.15, *MSE* = 522.52, *p* = .294, *η2* = .05. This can be seen in Figure 2. Final recall was significantly lower in the trivia condition, however, *F* (1, 26) = 12.14, *MSE* = 23.31, *p* < .01, *η2* = .318.

Paired samples t-tests showed that final recall and final JOL were significantly different for the test condition, *MD* = -11.2, *t* (14) = -3.57, *p* < .01; while final recall and final JOL did not differ significantly for the trivia condition, *MD* = 5.81, *t* (12) = 1.51, *p* = .16. This illustrates that despite the JOL values not differing between the two conditions, in the testing condition they were fairly underconfident, yet in the trivia condition they tended towards *overconfidence*.

A within-subjects repeated measures ANOVA between JOLs and recall for the repeated testing group revealed that JOLs were significantly lower than performance, *F* (1, 13) = 12.06, *MSE* = 136.00, *p* < .01, *η2* = .481. There was an increasing gap between judgments and recall over trials, showing the underconfidence with practice effect.

Individual and Aggregate Judgments of Learning. Another set of participants made both individual and aggregate JOLs, and experienced either repeated trivia or repeated tests. These values are displayed in Figure 3. Aggregate JOLs were significantly different between these two groups, with the trivia condition reporting higher values, *F* (1, 24) = 13.83, *MSE* = 582.61, *p* < .01, despite lower final recall performance, *F* (1, 29)

= 3.35, $MSE = 137.68$, $p = .08$, $\eta 2 = .10$. Paired samples t tests showed that final JOLs were significantly lower than final recall for the testing condition, $MD = -17.83$, $t\ (14) = -6.09$, $p < .01$, and slightly (not significantly) higher for the trivia condition, $MD = 5.47$, $t\ (14) = 1.44$, $p = .17$.

A comparison between JOLs and recall for the repeated testing condition showed that JOLs were significantly lower than recall values, $F\ (1, 14) = 16.19$, $MSE = 277.99$, $p < .01$, $\eta 2 = .536$. The graph shows the usual underconfidence with practice effect, with judgments underconfident in later trials.

Speculation that aggregate JOL values would differ between the conditions with individual JOLs and the conditions with only aggregate JOLs was not supported, as a repeated measures ANOVA revealed that only the condition with both individual level JOLs and trivia rather than tests differed significantly from the other three conditions; post hoc analyses showed significant differences between that condition and the other three JOL conditions at .05 significance.

Average individual item JOLs were analyzed for differences between the trivia and test conditions, with the finding that individual item JOLs were significantly higher for those in the trivia condition, $F\ (1, 27) = 26.29$, $MSE = 1412.57$, $p < .01$, $\eta 2 = .493$. A within subjects ANOVA comparing mean JOLs (converted to # of words) and recall across trials for the testing condition show that mean individual JOLs were significantly lower than performance, $F\ (1, 13) = 11.84$, $MSE = 163.06$, $p < .01$, $\eta 2 = .477$.

Discussion

Overall, metacognitive judgments were more confident for those in the trivia conditions, despite lower recall performance. Participants in the repeated testing condition showed significant underconfidence when judgments were compared to actual learning. This finding supports the idea that judgments of learning may be informed by memory for prior test performance, as Finn and Metcalfe (2007) suggest. Interestingly, judgments of improvement did not differ, yet showed significant underconfidence in the repeated testing condition, suggesting that participants failed to appreciate the benefits of testing on memory performance, unaware that testing improved recall performance.

Participants who reported only aggregate JOLs did not show significantly higher JOLs for the repeated testing group, despite their significantly higher performance. A closer examination of the data suggests that although participants in the trivia condition reported similar judgments of learning as the participants in the repeated testing condition, the trivia participants may not have been particularly overconfident, but in the testing condition, people failed to appreciate the benefit of repeated tests. There is much evidence showing underconfidence in repeated study – test procedures, referred to as the underconfidence with practice effect (Koriat, Sheffer, & Ma'ayan, 2002). Explanations for this effect refer to the feedback that one gets through a testing experience as the cause, known as memory for past test (Finn & Metcalfe, 2007) and indeed, the average JOL tends to be more closely related to the performance on the previous test rather than the current one (Finn & Metcalfe, 2007, 2008). The lack of clear bias (on the final JOL) for the trivia group is interesting, since trial 1 JOLs in repeated testing experiments are generally quite overconfident (Koriat, Sheffer, & Ma'ayan, 2002); this suggests that either repeated experience or the spacing of study may discourage systematic overconfidence. This may be due to either more accurate mnemonic cues from study experience, or because the spacing of study results in lower perception of processing fluency. However, it is important to note that spaced study was not as fruitful for actual learning as compared to repeated testing.

The addition of individual item JOLs was associated with significantly higher aggregate JOLs in the trivia group, though again, the final JOL was not significantly different from recall, showing the lack of a consistent response bias. Aggregate JOLs were underconfident in the testing group, just as in the aggregate – only condition. The individual item JOLs showed the same pattern of results as aggregate JOLs; significantly underconfident in the testing condition, and those in the trivia condition were significantly higher than those in the test condition.

The results of Experiment 1 confirm our hypotheses that removing repeated tests results in greater confidence relative to performance. The trivia – spaced condition showed more confident JOLs, both on the individual item and aggregate level, and judgments of improvement were significantly more confident in the trivia – spaced condition.

Experiment 2

For the second experiment, we compared the trivia-spaced condition with a massed study condition. It is informative that repeated tests are associated with underconfidence, and that spacing study with trivia instead of repeated testing is associated with less bias; however, massed practice may also be a common practice during self-regulated learning. Massed learning is expected to result in overconfident judgments of learning as compared with the trivia-spaced condition, as many mnemonic cues will be misleading. Judgment cues such as familiarity and processing fluency should be deceptive during massed practice, with items feeling much more fluent and familiar despite little increase in recall-ability. These effects should be mitigated, and cues should be somewhat more informative in the trivia-spaced condition, where there are larger gaps between presentations of material.

Method

Participants. A total of 54 participants from the University of California, Merced subject pool volunteered to participate.

Procedure. Materials were identical to those in Experiment 1, and the trivia-spaced condition proceeded identically to that of Experiment 1. Those in the massed practice condition experienced the same learning procedures, but without trivia spacing in between study presentations of the list.

Scoring. Word recall was scored according to whether responses matched the target words. No points were taken off for misspellings.

Results

Judgments of Learning. A repeated measures between subjects ANOVA revealed that judgments of learning were not significantly different between spaced and massed conditions, *F* (1, 52) = 1.40, *MSE* = 1689.56, *p* = .24; means are displayed in figure 4.

Recall. There was no difference in recall performance, *t* (52) = -.386, *p* = .70.

Accuracy. Judgments of learning on the final trial were compared to recall values. Paired samples t-tests revealed no difference between JOL and recall for the massed condition, *t* (28) = .71, *p* = .49, or for the spaced condition, *t* (24) = -.14, *p* = .89.

In case the null results are due to insufficient sample size, additional participants will be run, and statistics re-calculated.

Discussion

Current results indicate that participants did not perform differently in the massed and spaced conditions, nor were judgments of learning different between the two conditions. It is fairly likely that the sample size is too small, and to address this issue, additional participants will be run. We should find that performance is lower for participants in the massed practice condition, while their judgments of learning should be higher than those in the spaced practice condition. Massed practice should result in significantly more overconfident judgments of learning than spaced practice; even if JOLs are not significantly different from JOLs during spaced practice, they should be significantly higher than actual recall. Previous research suggests that participants are unaware that massed practice is associated with less learning when compared to spaced practice (Simon & Bjork, 2001; Dunlosky & Nelson, 1994), and should expect the same, or better, levels of learning while actually performing worse.

Experiment 3

While JOLs may be significantly more overconfident during massed practice as compared to trivia – spaced practice, the actual judged learning rate was not examined in Experiment 2. Judged improvement values should also be influenced by massed practice, with participants rating their learning much higher in this condition. This pilot experiment was designed to investigate the effect of massed study trials on JOIs (a new experiment comparing JOIs for massed practice, and trivia-spaced practice will be run this summer; this experiment is treated as a pilot experiment, as it was initially designed to look at the effect of the number of study trials on judgments, so not all participants experienced all six study trials). Judgments of improvement were compared to those of a previous experiment, to investigate if massed study repetitions might change JOIs, through an increased sense of fluency. If massing the study trials results in increased feelings of processing fluency and/or familiarity, JOIs should be much higher when study trials are massed. In addition, participants completed a survey which asked them to predict how a group of students would perform in a similar experimental scenario, to see if students' expectations of others would also be affected by their experience (would they rate others' learning similarly to how they rated their own learning in the experiment?). The survey had two forms, with participants completing one of them; the first form asked them to predict how many words students would know after each of six study trials, while the other asked how many words students would learn on each of six study trials.

Method

Participants. 155 participants from the University of California, Merced psychology subject pool volunteered to participate for course credit.

Materials. A set of 50 Swahili-English word pairs from the Nelson and Dunlosky (1994) norms were used.

Procedure. Number of study trials was manipulated between participants: each participant experienced one of six experimental conditions, with from 1-6 study trials. Participants made predictive judgments of improvement before each study trial, which were prompted with the question "if you study the list now, for the same amount of time as the previous study trial, what percent more of the list do you think you will learn?" After experiencing all study trials, each participant completed one recall test, in which they were shown each Swahili word, and typed in the appropriate English word in response. At the conclusion of the experiment, all participants completed a survey in which they predicted the performance of other learners. The survey was either predicting knowing, or how well other students would do on a recall test after each of six study trials, or predicting learning, or how many more words other students would learn on each of the six study trials.

Scoring. Words on the recall test were scored as correct if they matched the target word, no points were taken off for misspellings.

Results

Judgments compared with previous experiment. A comparison with the JOIs of Experiment 1 (condition 6 of this experiment was used, as the others had less than 5 JOIs) showed a significant difference among the three JOI conditions, *F* (2, 40) = 12.36, *MSE* = 73.04, *p* = .13, $\eta 2$ = .382. Post hoc tests revealed a significant difference between the JOIs of Experiment 2 and the two conditions of Experiment 1, at a .01 level of significance. Results are displayed in Figure A.

Survey. JOLs on the “know” post experiment survey, which asked for estimates of how much participants would learn on each study trial, were not significantly different from actual performance of experiment participants, with the exception of trial 5. Results are presented in Figure 6; t-tests are in Table 1.

For the “learn” post experiment survey, which asked for estimates of the rate of improvement of participants on each study trial, the initial JOL was slightly, but not significantly higher than trial 1 performance, *F* (1, 76) = 3.46, *p* = .067 (*MD* = 1.52). Similarly to judgments of improvement given during the experiment, JOIs were still significantly higher than mean improvement between trials, as seen in Figure 7. Mean improvement between trials is the difference between mean recall scores, and as each participant only completed one recall test, these improvement values appear somewhat erratic, not showing a smooth line with diminishing value that would be expected.

Discussion

In this experiment, we found that judgments of improvement in a massed practice scenario were significantly higher than judged improvement during repeated testing and spaced practice scenarios. Massed practice is clearly associated with a higher perceived learning rate; based on the results of Experiment 1, which found no difference in JOIs between trivia spaced and testing conditions, it is likely that the lack of spacing lead to increases in JOIs.

One possible explanation for this effect of higher JOIs without repeated tests is that testing provides feedback to the learner. If test feedback were a cue used in JOIs, it would differ somewhat from the memory for past test heuristic, where JOLs tend to be correlated with trial N-1 (previous trial) test performance. In prior experiments, we had found that JOIs were not correlated with changes in JOLs from trial N-1 to trial N; however, a between subjects analysis may not be sufficient to suggest that participants do not make covert JOLs, which would be informed by MPT, and those JOLs could in turn influence JOIs. This explanation is not sufficient to explain why JOIs should be higher for massed versus spaced practice, however, as there is no recall feedback gained from spacing of study.

Corinne Townsend 6/19/11 11:56 PM
Deleted: which is

The pattern seen on the surveys, which were *also* lower than JOIs given during the experiment, also suggests that the massed study situation results in experiential cues that result in very high JOIs, such as an increased sense of fluency of the material, which is mistakenly perceived to be informative of how quickly they are learning the material. This may help explain why massed practice is something students often mistakenly feel is very effective.

General Discussion

In these experiments, we found evidence that repeated testing is associated with significantly underconfident metamemory judgments, while spacing presentations may lead to reduced bias, and massed practice is likely to lead to significant overconfidence. The underconfidence associated with repeated testing, known as the underconfidence with practice effect (Koriat, Sheffer, & Ma'ayan, 2002), is thought to be primarily due to memory for past test (Finn & Metcalfe, 2007, 2008), as participants can refer to prior recall experience when making judgments of learning. Our results suggest an additional mechanism by which repeated testing may show a reduction in confidence, specifically, the spacing of study. Repeated testing not only gives feedback, but also spaces the study presentations of material, and the trivia – spaced practice condition was associated with reduced bias in both Experiments 1 and 3. Massed practice, when compared with spaced restudy, was associated with particularly overconfident predictions concerning the rate of improvement, though it is unclear how (or if) judgments of learning are affected.

The differences in metacognitive judgments between massed and spaced practice are likely due to changes in processing fluency – the massing of list presentations would increase feelings of fluency, as items were recently processed, and this increase in fluency due to recent presentation may be incorrectly attributed to how well learned the items are. In massed practice, the fluency cues will be strong, but uninformative, while in spaced practice, these cues will be somewhat more informative, as feelings of fluency will increase much slower due to the longer gap between presentations. Enhancing processing fluency has been found to be associated with increases in judgments of learning, despite a lack of improved recall performance (Tiede, Derksen & Leboe, 2009), suggesting that participants may rely on fluency cues, misattributing the fluency as a mnemonic indicator rather than a sign of recent presentation, or priming (Tiede, Derksen & Leboe, 2009).

The implications of this research is that unfortunately, students are overall the least confident when they should be the most confident about their performance, as repeated testing resulted in higher recall performance, and is known to improve recall performance under many conditions (Roediger & Karpicke, 2006a, 2006b). Massed practice, while associated with the greatest confidence, was the least effective study scenario. This misguided confidence may lead students to choose ineffective methods of study (massed practice) because these methods feel more rewarding. Students do not generally choose to self-test (Karpicke, Butler, & Roediger, 2009) and one reason why they do not opt to self-test may be that after self testing, they feel as if they had learned less material, though self testing does provide an opportunity for a student to give themselves an accurate self-appraisal if answers are on hand. However, for students that are highly motivated and do choose to self test, they are much less likely to have inflated estimates of their ability, and may study longer because of their reduced confidence

A different interpretation of the results would be that although spacing restudy may be associated with reduced (or no) underconfidence, the underconfidence with practice effect observed in repeated-testing conditions is only due to memory for past test and is not influenced at all by spacing of study. To address this criticism, a follow up study may compare a space + test condition with a long space condition. It has been found that delayed judgments of learning do not display the underconfidence with practice effect (Finn & Metcalfe, 2007), but in this case the judgments themselves would not be delayed. A study/jol – test – space condition will be compared with a study/jol – space – test condition, as well as a study/jol – space condition (this spacing will be equal to the test and space in the other conditions, for equal time delay between restudy). Judgments of learning in these two conditions should be lower than those in the spaced only condition; there should be reduced confidence that results from spaced restudy, and additional underconfidence added from memory for past test. The additional underconfidence from memory for past test may be somewhat reduced in the condition where the space occurs after testing, as more of the test experience will be forgotten. If, however, there is not evidence that these factors are additive, such as the test and spaced conditions are associated with less underconfidence than testing alone, this would provide evidence that the underconfidence with practice effect observed in repeated testing is not at all due to spacing effects.

Tables and Figures

Individual tests are listed because repeated measures comparisons could not be calculated- not all participants experienced all six trials. Because of the structure of the experiment, individual comparisons had to be made.

Table 1.

Comparisons between survey JOLs and recall performance in Experiment 3.

Trial	*df*	*F*	*p*	*MD*
1	1, 103	.007	.94	.063
2	1, 100	.007	.93	.086
3	1, 102	.629	.43	-1.0
4	1, 106	.356	.55	.83
5	1, 105	7.48	.007*	-5.17
6	1,100	.028	.87	-.37

Figure 1. Judgments of Improvement, trivia spaced and repeated testing conditions, Experiment 1.

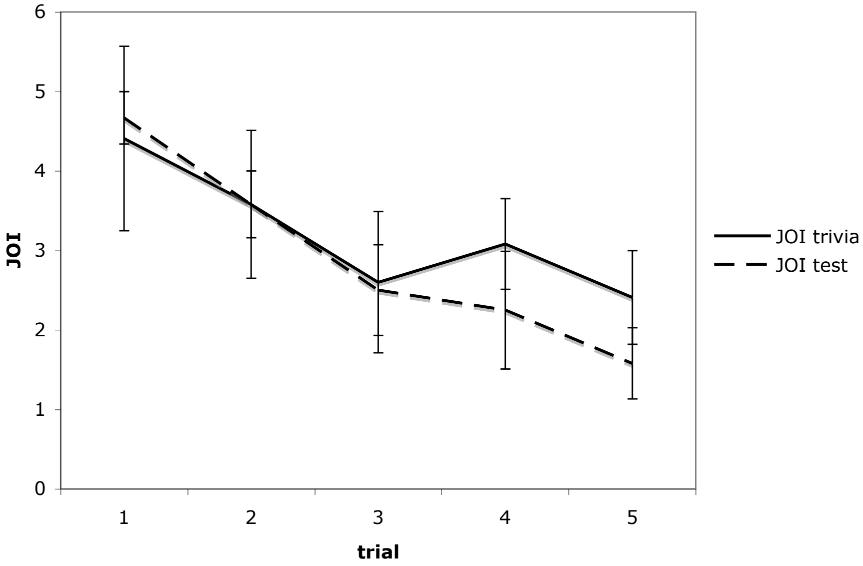


Figure 2. Judgments of Learning, trivia spaced and repeated testing conditions, aggregate-only JOL condition, Experiment 1.

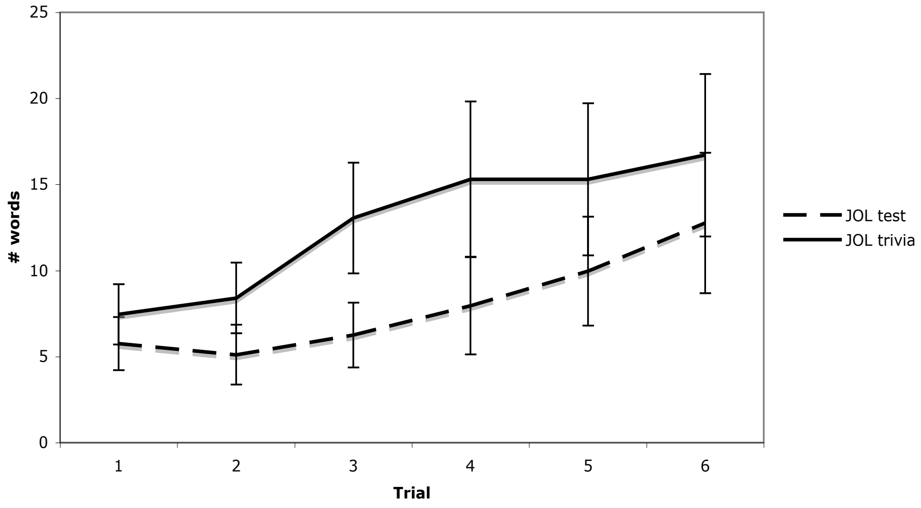


Figure 3. Judgments of Learning, individual and aggregate, trivia spaced and repeated testing conditions, aggregate and individual JOL condition, Experiment 1.

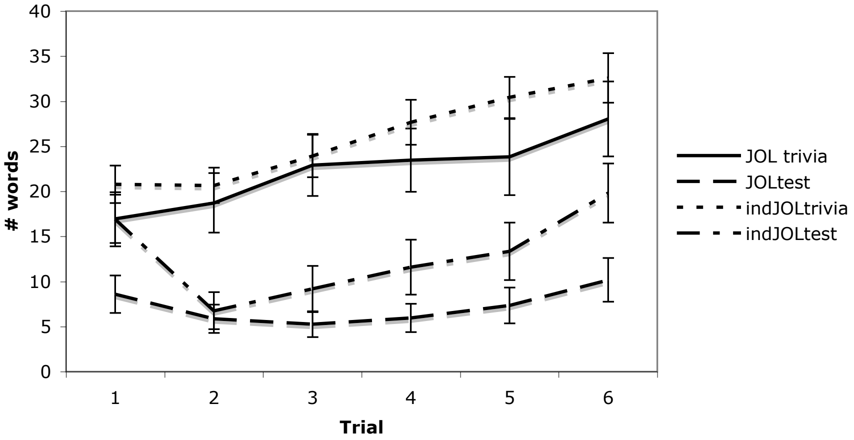


Figure 4. Judgments of Learning, trivia spaced versus massed study conditions, Experiment 2.

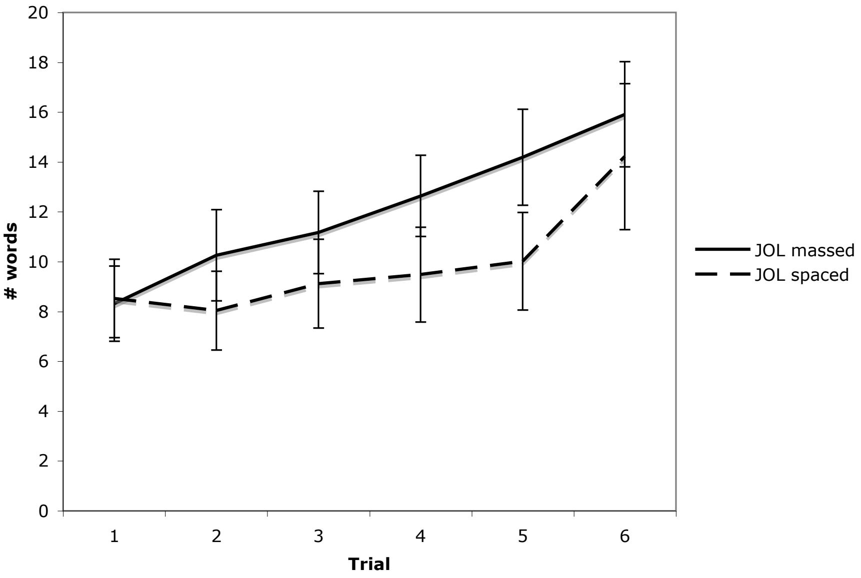


Figure 5. Experiment 3, Judgments of Improvement, massed, compared to JOIs in Experiment 1.

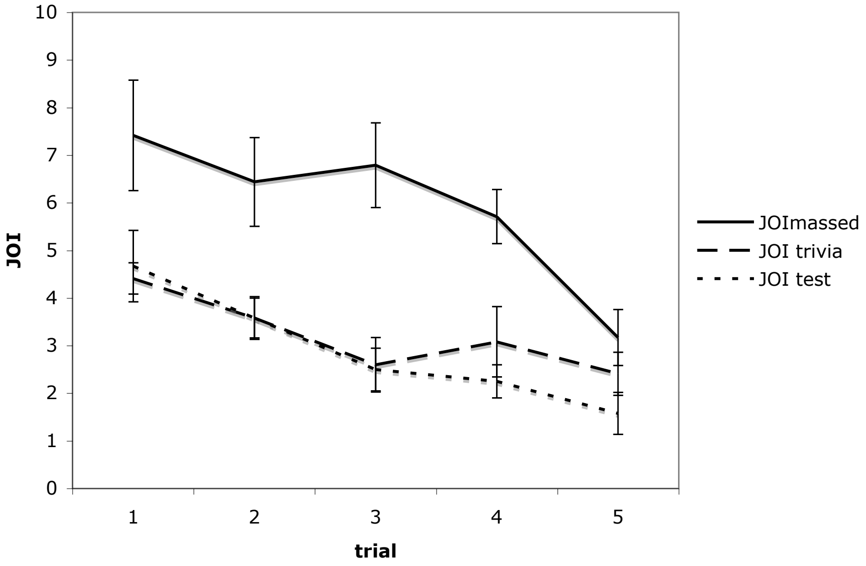


Figure 6. Survey judgments of learning compared to performance, Experiment 3.

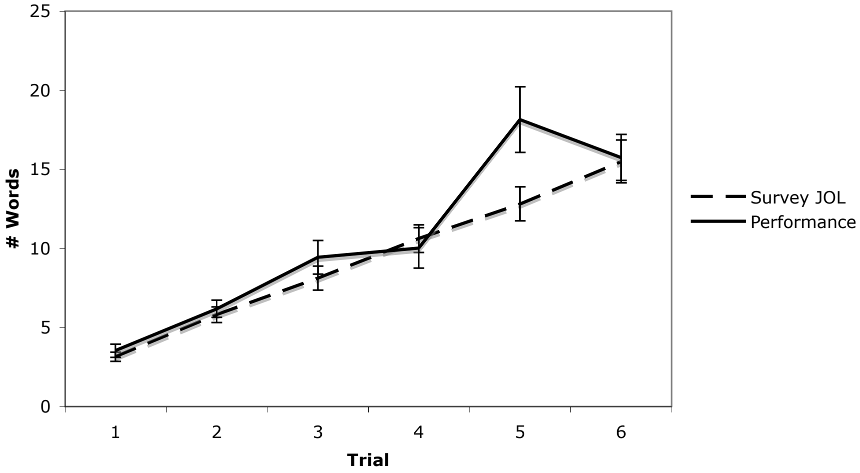
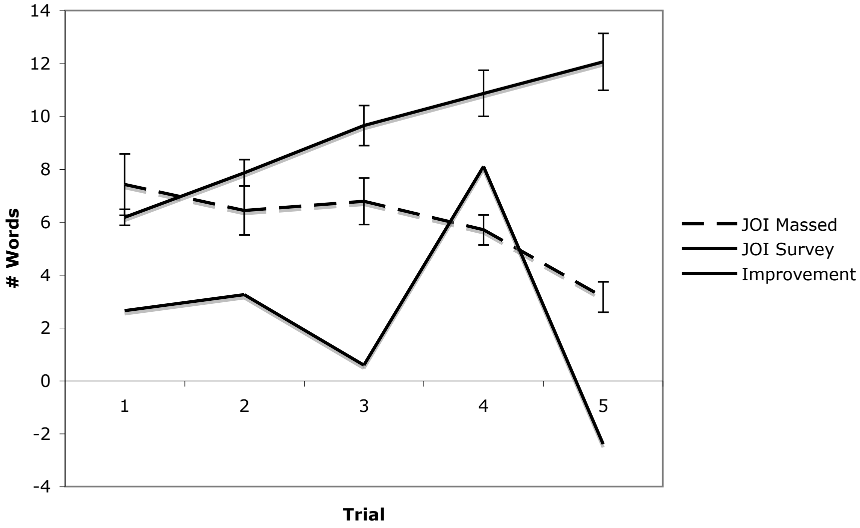


Figure 7. Survey judgments of improvement compared to recall improvements and JOIs, Experiment 3. Improvement values are the differences between mean recall values between trials.



References

Dunlosky, J. & Nelson, T. (1994). Does the sensitivity of judgments of learning (jols) to the effects of various study activities depend on when the jols occur? *Journal of Memory and Language, 33,* 545-565.

Finn, B. & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 238-244.

Finn, B. & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*(2008), 19-34.

Hirshman & Bjork, (1988). The generation effect: support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 484-494.

Karpicke, J., Butler, A., & Roediger, H.(2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17*(4), 471-479.

Koriat, A. (1997) Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147-162.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. (2006). Exploring a mnemonic debiasing account of the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 595-608.

Mazzoni G., & Nelson, T., (2005). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1263-1274.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159-163.

Pyc, M., & Dunlosky (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory & Cognition, 38*(4), 431-440.

Roediger, H., & Karpicke, J. (2006a). Taking memory tests improves long-tem retention. *Psychological Science, 17*(3), 249-255.

Roediger, H., & Karpicke, J. (2006b).Basic research and implications for educational practice. *Perspectives on psychological science, 1*(3), 181-210.

Serra, M. & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence with practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition. 31*(6), 1258-1266.

Simon, D., & Bjork, R., (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(4), 907-912.

Tiede, H., Derksen, C., & Leboe, J. (2009). An investigation of increases in metamemory confidence across multiple study trials. *Memory, 17*(3), 288-300.

General Discussion

The choices that students make during self-regulated learning are numerous and complex. Decisions about switching between materials, how long to study materials, how to prioritize materials, what study goals to set, and the use of metacognitive judgments in order to make these decisions have been examined in these experiments.

In the first chapter, three experiments show that learners tend to switch more often at difficult items, though in some cases effects were not significant, and overall the vast majority of participants did not choose to switch. These results suggest that complex switching decisions may not be made at all unless the scenario readily lends itself to quickly making such decisions, or if switching behavior would result in a large advantage. It is unlikely that the non-switching participants believed that switching would be advantageous, and they likely felt that it was a waste of cognitive resources that would be better spent on rehearsal of items. The default strategy appears to simply be discrepancy reduction: study more difficult, less well learned items, for longer. This explanation aligns with the hierarchical model of study time allocation (Thiede & Dunlosky, 1999) and with the more recent agenda-based regulation model of study time allocation (Ariel, Dunlosky, & Bailey, 2009), which is an evolution of the hierarchical model. If participants chose to allocate more time to easy and moderate difficulty items, and/or chose to switch more often at these items, that would cast doubt on these models (unless those items carried a higher point value on an exam). The overall tendency to not switch lists poses a problem for the proximal learning model, which assumes that learners are driven to maximize current learning rates (Metcalfe & Kornell, 2005). Lack of switching, and greater time allocation to normatively difficult material aligns well with the discrepancy reduction model (Dunlosky & Hertzog, 1998), which emphasizes the learner's drive to reduce items with the greatest discrepancies between current learning and the goal state.

The second chapter investigates the contribution of self-efficacy to metacognitive judgments and regulation of behavior. In several experiments, self-efficacy was experimentally manipulated, and when manipulations were successful, effects were seen on metacognitive judgments such that participants in the high self-efficacy condition gave higher judgments of learning and improvement. When experimental manipulation of self-efficacy was ineffective, differences were still observed between metacognitive judgments of those with high or low reported self-efficacy. Though self-efficacy appeared to be related to judgments of learning and improvement, no effects were observed on study time, item prioritization (all participants spent more time restudying low-jol items), or goal setting. Like chapter 1, these results appear to support discrepancy reduction (Dunlosky & Hertzog, 1998) and the hierarchical model of study time allocation (Thiede & Dunlosky, 1999), which emphasizes that learners seek to minimize effort while maximizing the chances of meeting the study goal; because all students showed similar strategizing, study time, and performance, they likely followed the same general study plan: study difficult items longer, and study all items long enough to reach the specified goal. Deviating from this default discrepancy-reduction plan likely does not occur without differentially weighting items (Ariel, Dunlsoky & Bailey, 2009) or without enacting time constraints (Son & Metcalfe, 2000; Dunlosky & Thiede, 2004). These results do not support the social-cognitive model of study time allocation (Bandura, 1991; Zimmerman, 1990), though perhaps the effects of self-efficacy on self-regulatory behaviors would be more apparent in more ecologically valid learning materials that participants have a longer history with. Self-efficacy for one's capacity to learn translations might not have a large impact, as there is little experience with the task, and only a small amount of information that self-efficacy is based on. Additionally, differences in self-regulatory behaviors might be more readily apparent within subjects, if given two kinds of material to learn, with differing efficacy levels for the two materials.

Chapter 3 focuses on how metacognitive judgments are made, examining the role of framing in influencing judgments of learning. Results showed that negative framing resulted in increased judgments of learning, but that this was limited to experimental situations and not to theories of learning, as no effects were seen on survey data. Framing effects were not due to self-enhancement, as they were also observed for judgments made about other students, and not just for judgments made about one's own performance. Framing effects may have been caused by an anchoring effect; asking students how many of the 50 items they missed may have effectively served as a high anchor for performance estimates as compared with regular estimates, which may rely on students' own internal anchors that are based on their expectations of performance due to past experiences. Whatever the cause of the framing effects observed, it is clearly important to consider the importance of how metacognitive inquiries are phrased, as the phrasing may significantly influence the answers that are given. Additionally, this calls into question the current knowledge on metacognitive judgments; students are likely to pose their own metacognitive questions in different ways, based on their experience. Some students may have experience with parents and teachers asking them how well they know material, and how well they think they will do on exams; others may have the experience of being prompted to think about how much material they do not know well, or how likely they are to do poorly on an exam. Still other students may conceptualize the question as whether they have studied for a requisite amount of time that is associated with good performance. Metacognitive regulation will depend on the accuracy of metacognitive judgments, which may be significantly influenced by how these judgments are conceptualized.

Chapter 4 examined the impact of different types of study scenarios on metacognitive judgments. Repeated testing was associated with significant underconfidence, which is widely supported in the literature (Koriat, Sheffer & Ma'ayan, 2002), but part of this underconfidence may be due to spacing of study rather than solely due to memory for past test performance (Finn & Metcalfe, 2007; 2008). Spaced restudy was associated with reduced biases in judgments of learning, while massed restudy led to overconfident judgments of improvement (though unclear results on judgments of learning). These results highlight the importance of the study situation on metacognitive judgments, and that different situations may have drastic effects on metacognitive accuracy. In addition, certain kinds of study behavior may feel deceptively productive (such as massed restudy) while very productive methods, such as repeated study and testing, may feel deceptively unproductive and leading to severe underconfidence despite being superior to other methods of study. This has implications for models of study allocation, such as the hierarchical model (Thiede & Dunlosky, 1999) which state that students will likely behave in a way to maximize learning while minimizing effort – this would predict that students might choose methods like cramming (massing practice) as it feels more productive than repeated testing, and repeated testing certainly requires more effort on the part of the learner. Students do report that they do not regularly choose to self-test (Karpicke, Butler, & Roediger, 2009), so this prediction may be accurate.

A potential problem with these experiments is that there were no observed differences in behavior in some instances where they were expected; differences in self efficacy did not lead to differences in study time, or in study time allocation. One explanation for this result is that no differences were observed simply because the manipulation was between subjects, and it would be necessary to look at a within subjects manipulation. As an example, we might not see differences in study time, or allocation behaviors if we observe two different people with differing self efficacy levels, or differing judgments of learning, but if we look at these same people with different materials, and those materials are given different self efficacy ratings and judgments of learning, we are more likely to see differing amounts of study time and strategies. In regards to judgments of improvement, these have not shown themselves to be useful in allocation decisions so it is likely that examination of these judgments is less informative.

Another point that has not been addressed concerns the development of these metacognitive skills, and whether students learn how to make metacognitive judgments and how to strategize, or whether these abilities are innate. From these included experiments, it is impossible to draw any conclusions regarding such matters; however, it seems more likely that such skills are ones that must be learned and acquired over time. Many students do not possess good study skills, or strategize while studying, but this may be due to different educational experiences; the ways in which teachers and parents encourage students to engage with materials may vary widely in different cultural contexts. The basic ability to reason about mental states of oneself and others (Theory of Mind) is one that is universal (Dunlosky & Metcalfe, 2009). The more specific ability to reason about the contents of one's memory appears during early childhood, with research showing that different ages of children (kindergarten, second, and fourth grade) showed similar relative accuracy in their of judgments of learning made for picture pairs (Schneider, Vise, Lockl, & Nelson, 2000). Metacognitive accuracy does not always mean optimal control of learning, and we do see that actual regulation of study does change quite a bit with age; younger children spend similar amounts of time on easy and difficult items, while older children begin to spend longer on the more difficult items, have more strategy knowledge, and utilize more study strategies (Dunlosky & Metcalfe, 2009). This supports the idea that metacognitive strategies and control of study time is likely learned behaviors, acquired from years of schooling. Cross cultural research could help illuminate this topic; students with similar performance may show similar skill in their judgments of learning, but vary widely in their strategy knowledge and study allocation and strategizing.

In terms of whether people are generally good, or bad, at metacognition is a difficult question to answer. The ability to think about one's thoughts and the contents of one's memory is a useful skill, and most people will know that they possess no knowledge or skill in certain areas that they lack training in (e.g. a librarian who has never worked on cars will know that they do not understand how their engine works, and will not attempt to fix it). Where metacognition starts to appear less accurate is when making more precise estimations of how well you understand or know different materials. Many researchers argue that relative accuracy is fairly good, and certainly above chance (see Dunlosky & Metcalfe, 2009, for an extensive review). This ability could certainly aid learners if they wanted to optimally distribute study time among items; however, actual study behaviors seem far from optimal. Even when learners spend more time on difficult items (as found in these enclosed experiments), allocation is not necessarily *ideal*, nor are the rehearsal strategies that are being used. One could argue that the goal is to minimize effort while optimizing learning, but this argument could be used to explain just about any kind of study behavior, from cursory, lazy restudy of difficult items only, to a learner actively studying difficult material, briefly revisiting easy materials, and quizzing themselves. In any case, learners have room for improvement in their metacognitive accuracy, and in their knowledge and implementation of different rehearsal strategies and study allocation strategies. One must also keep in mind that our participants are by no means representative of the general population—these participants from whom we are drawing conclusions are college students, who are ostensibly the most metacognitively skilled participants that are available, and the fact that they have much room for improvement suggests that the population as a whole is likely to be lacking even more in terms of metacognition.

Generalization to practical, educational settings is always somewhat difficult when experiments involve simple word pairs, whereas actual learning generally encompasses comprehension and retention of complex concepts, learned both in lecture and in reading textbooks and class notes. However, it is likely that many relationships will still hold, provided that the learner does not have problems with reading comprehension. Students are likely to be influenced by their judgments of learning; one would assume that given two midterms the next day, the student is likely to study for the course with the lower current level of learning (if all else is held equal). Self efficacy judgments and test expectations may even be more accurate, as there is often much past experience that students can draw from when making these evaluations. In addition, the restudy behaviors one must engage in with course materials may by default be more complex than simple repetition; rereading of notes or textbook chapters will involve more active, complex reading processes than does repeating a word's definition. Evidence suggests that in educational contexts, self efficacy is an important measure to look at; Pintrich and DeGroot (1990) have found significant relationships between self efficacy and classroom engagement and performance in a sample of 173 seventh graders in Science and English courses, and Pintrich, Marx and Boyle (1993) review further evidence that self efficacy is correlated with classroom achievement. In regards to judgments of learning, Hacker, Bol, Horgan, and Rakow (2000) examined students' abilities to predict exam performance and found that the most accurate students were the highest performing, confirming the relationship between metacognitive accuracy and performance. In addition, research on metacomprehension (judgments for how well text materials are understood) show that more accurate metacomprehension judgments result in more effective restudy behaviors and greater levels of performance (Thiede, Anderson, & Thierrault, 2003).

Overall, the included experiments show several conclusions regarding student behavior during study time allocation. Students were unlikely to engage in repeated switching behavior, instead enacting a low effort strategy to reach study goals, as suggested by the hierarchical model (Thiede & Dunlosky, 1999). While learners appeared to use their feelings of self-efficacy to inform metacognitive judgments, effects on overall study time and strategizing were minimal, challenging the social-cognitive theory of self-regulated learning (Bandura, 1991; Zimmerman, 1990), while further supporting the hierarchical model (Thiede & Dunlosky, 1999). Framing of metacognitive judgments appeared to significantly influence judgments of learning, a key finding that suggests that different conceptualizations of metacognitive questions may have profound impacts on estimates regarding one’s performance, and in turn, self-regulation behaviors. Study scenarios also appear to influence metacognitive judgments, with repeated testing leading to underconfidence, spacing repetitions associated with reduced bias, and massed presentations associated with inflated estimates of improvement. The complex relationship between metacognitive judgments, influences on metacognitive judgments, allocation behaviors, performance, and self-efficacy beliefs cannot be fully explained by these experiments, but these results suggests that students prefer a low – effort study strategy that they feel is productive, and that their judgments are significantly influenced by question framing and study conditions.

Corinne Townsend 7/16/11 3:14 PM
Deleted: ... [1]

References

Bandura, A. (1977). Self efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1991). Social cognitive theory of self regulation. *Organizational Behavior and Human Decision Processes, 50,* 24-287.

Bandura, A. (1993). Perceived self efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A., & Cervone, D. (1983). Self evaluative and self efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology, 45*(5), 1017-1028.

Berry, J. M. & R. L. West (1993). Cognitive self efficacy in relation to personal mastery and goal setting across the life span. *International Journal of Behavioral Development, 16*(2), 351-379.

Bouffard-Bouchard, T. (1990). Influence of self efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353-363.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self efficacy for self regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100*(3), 525-534.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control process view. *Psychological Review, 97*(1), 19-35.

Charnov, E. L. (1976). Optimal foraging: The marginal value theorem. *Theoretical*

*Population Biology, 9,*129-136.

Collins, J. L. (1982). Self efficacy and ability in achievement behavior. Paper presented at the meeting of the American Educational Research Association, New York.

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky & A. C. Graessler (Eds.), *Metacognition in educational theory and practice* (pp 249-276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228-232.

Dunlosky, J. & Metcalfe, J. (2009). Metacognition. Thousand Oaks, CA: SAGE publications.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self paced study. *Acta Psychologica,* 98, 37-56.

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory and Cognition, 32*(5), 779-788.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*(1), 19-34.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 238-244.

Fletcher, W. “Phrases in English” database. Retrieved October 2006, from http://pie.usna.edu/index.html

Green, R. F. (1984). Stopping rules for optimal foragers. *The American Naturalist, 123*(1), 30-40.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior, 13*(2), 243-266.

Isaacson, R. M. & Fujitsa, F. (2006). Metacognitive knowledge monitoring and self regulated learning: Academic success and reflections on learning. *Journal of the Scholarship of Teaching and Learning, 6*(1), 39-55.

Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of near threshold priming on metamemory and recall. *Acta Psychologica,* 55-68.

Kearney, E. M., & Zechmeister, E. B. (1989). Judgments of item difficulty by good and poor associative learners. *American Journal of Psychology,* 102, 365-383.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory and Cognition, 36*(2), 416-428.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A., & Bjork, R. (2006). Mending metacognitive illusions: A comparison of mnemonic based and theory based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 1133-1145.

Koriat, A., Bjork, R,, Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience based and theory based processes. *Journal of Experimental Psychology: General, 133*(4), 643-656.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause and effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36-69.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3) 595-608.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147-162.

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: the benefits – and costs – of dropping flashcards. *Memory, 16*(2), 125-136.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219-224.

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.

LaPorte, R. E. & Nath, R. (1976). Role of performance goals in prose learning. *Journal of Educational Psychology, 68*(3), 260-264.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 464-470.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 663–679.

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5) 1263-1274.

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica, 113*(2), 123-132.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*(3), 349-363.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174-179.

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132*(4), 530-542.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*(4), 463-477.

Moores, T. T., Chang, J. C., Smith, D. K. (2006). Clarifying the role of self efficacy and metacognition as predictors of performance: Construct development and test. *The DATABASE for Advances in Information Systems, 37*(2 & 3), 125-132.

Moss, H., & Older, L. (1996). *Birkbeck Word Association Norms*. Hove, UK: Psychology Press.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the labor-in-vain effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676-686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125-140). San Diego, CA: Academic Press.

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

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*(4), 207-213.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General, 136*(3), 370-388.

Pintrich, P. R., & DeGroot , E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Pirolli, P. & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643-

675.

Pirolli, P. (2005). Rational analyses of information foraging on the web. *Cognitive Science, 29*, 343-373.

Schneider, W., Vise, M., Lockl, K., & Nelson, T.O. (2000). Developmental trends in children's memory monitoring. Evidence from a judgment-of-learning task. *Cognitive Development, 15,* 115-134.

Sexton, T, & Tuckman, B. W. (1991). Self-beliefs and behavior: The role of self-efficacy and outcome expectation over time. *Personality and Individual Differences, 12*(7), 725-736.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27,* 379-423.

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.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73.

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study:

An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1024-1037.

Thiede, K. W., Dunlosky, J., Griffith, T. D., & Wiley, J. (2005). Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 31*(6), 1267-1280.

Treadwell, J. R., & Nelson, T. O. (1996). Availability of information and aggregation of confidence in prior decisions. *Organizational Behavior and Human Decision Processes, 68*(1), 13-27.

Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology, 71*(5), 673-679.

Watson, D., & Clark, L. A. (1994). The PANAS-X: Manual for the positive and negative affect schedule-Expanded Form. Iowa City: University of Iowa.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25,* 82-91.

Zimmerman, B. J. (1990). Self regulated learning and academic achievement: an overview. *Educational Psychologist. 25*(1), 3-17.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self regulatory influences on writing course attainment. *American Educational Research Journal, 31*(4), 845-862.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663-676.

Appendix 1. Swahili Word Pairs

Easy Practice list:

Mbwa – Dog

Lulu – Pearl

Wingu – Cloud

Iktisadi – Economy

Goti – Knee

Yai – Egg

Pombe – Beer

Godoro – Mattress

Fagio – Broom

Tabibu – Doctor

Hard Practice List:

Ubini – Forgery

Ladha – Flavor

Ruba – Leech

Mshoni – Tailor

Fahali – Bull

Nafaka – Corn

Lawama – Blame

Kamba – Rope

Hadithi – Story

Gutu – Stump

Study List:

Sumu – Poison

Roho – Soul

Chakula – Food

Maiti – Corpse

Leso – Scarf

Chimbo – Quarry

Yamini – Oath

Tajiri – Merchant

Kasuku – Parrot

Adha – Trouble

Sanda – Shroud

Mfupa – Bone

Handaki – Trench

Desturi – Custom

Paji – Forehead

Lango – Gate

Inda – Spite

Bandari – Harbor

Vumbi – Dust

Fumbo – Mystery

Bahasha – Envelope

Wakili – Agent

Lozi – Almond

Jani – Leaf

Baharia – Sailor

Pamba – Cotton

Nabii – Prophet

Dafina – Treasure

Adui – Enemy

Zeituni – Olives

Ziwa – Lake

Punda – Donkey

Kaputula – Shorts

Rafiki – Friend

Wasaa – Leisure

Tumbili – Monkey

Pazia – Curtain

Pipa – Barrel

Rembo – Ornament

Hariri – Silk

Farasi – Horse

Kaa – Crab

Duara – Wheel

Nyana – Tomato

Buu – Maggot

Jibini – Cheese

Rushwa – Bribe

Talaka – Divorce

Gharika – Flood

Ankra – Divorce

Appendix 2. English Word Pairs

Pepperoni – Pizza

Saltine – Cracker

Glove – Hand

Pistol – Gun

Instructor – Teacher

Knob – Door

Railroad – Train

Flight – Airplane

Lobe – Ear

Icing – Cake

Hill – Mountain

Web – Spider

Jealousy – Fear

Limit – Capacity

Manner – Method

Knowledge – Strength

Pride – Loyal

Impulse – Whim

Destiny – Eternity

Advice – Idea

Spirit – Being

There – Where

Panic – Anxiety

Attention – Discipline

Friends – Enemies

Shock – Scare

Galaxy – Space

Tunnel – Hole

Referee – Game

Market – Place

Goblin – Halloween

Parade – Float

Inch – Mile

Weather – Rain

Angle – Triangle

Musk – Scent

Ash – Fire

Dentist – Teeth

Pony – Horse

Bark – Tree

Mom – Dad

Sparrow – Bird

Omelet – Eggs

Inn – Hotel

Text – Book

Mud – Dirt

Freckle – Face

Cigarette – Smoke

Virtue – Quality

Morals – Dilemma

Compulsion – Tendency

Attitude – Arrogance

Honor – Justice

Impression – Mood

Lack – Confidence

Passion – Fulfillment

Recognition – Memory

Belief – Betrayal

Thought – Wonder

Foresight – Wisdom

Haze – Mist

Disaster – Catastrophe

Red – White

Junk – Yard

Mob – Riot

Exit – Entrance

Gallon – Gas

Shape – Form

Molecule – Science

Age – Year

Curved – Round

Oats – Wheat

Appendix 3. Survey

A: Judgments of how much students will know after X many trials

Imagine that more students were going to participate in a very similar experiment. They will also be learning Swahili vocabulary, but they will have 6 study trials, of a 50 word list.

Please answer all of the following questions:

a) How many words (total) do you think they would know after 1 study trial?

b) How many words (total) do you think they would know after 2 study trials?

c) How many words (total) do you think they would know after 3 study trials?

d) How many words (total) do you think they would know after 4 study trials?

e) How many words (total) do you think they would know after 5 study trials?

f) How many words (total) do you think they would know after 6 study trials?

g) On which study trial (1,2,3,4,5,6) do you think students are learning the most?

B: Judgments of how much students will learn on any given trial.

Imagine that more students were going to participate in a very similar experiment. They will also be learning Swahili vocabulary, but they will have 6 study trials, of a 50 word list.

Please answer all of the following questions:

a) How many words do you think they would learn during the 1st study trial?

b) How many *more* words do you think they would learn during the 2nd study trial (that were not already learned in the first trial)?

c) How many *more* words do you think they would learn during the 3rd study trial (that were not already learned in the first two trials)?

d) How many *more* words do you think they would learn during the 4th study trial (that were not learned in the first three trials)?

e) How many *more* words do you think they would learn during the 5th study trial (that were not learned in the first four trials)?

f) How many *more* words do you think they would learn during the 6th study trial (that were not learned in the first five trials)?

g) On which study trial (1,2,3,4,5,6) do you think students are learning the most?

Appendix 4.

Survey given in Self-Efficacy Experiments

1) **Next you are going to learn a new list of 50 items, with as much time as you need to study.** How confident are you, on a scale from 0 to 100 percent, that you could learn all **50** items?
(0 = not at all confident, 50 = moderately confident, 100 = completely confident)
____

2) **How confident are you that you will be able to concentrate, and do your best?**
Please answer on a scale from 0 to 100 percent; (0 = not at all confident, 50 = moderately confident, 100 = completely confident)
____

Rate the items on a scale of 1 to 7, with 1 meaning "not at all true of me" and 7 meaning "Very true of me"

1. Compared with other students in this experiment, I expect to do well.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

2. I am certain that I can memorize the items in this experiment.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

3. I expect to do very well in this experiment.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

4. Compared with others in this experiment, I think I'm a good learner.

1 2 3 4 5 6 7

1 = not at all true of me 7 = very true of me

5. I am sure I can do an excellent job on the recall test in this experiment.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

6. I think I will receive a good score in this experiment.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

7. My study skills are excellent compared with others in this experiment.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

8. I know that I will be able to learn the material for this experiment.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

Now we will ask some questions about your study habits for classes:

9. When I study for a test, I try to put together the information from class and from the book.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

10. When I do homework, I try to remember what the teacher said in class so I can answer the questions correctly.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

11. I ask myself questions to make sure I know the material I have been studying.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

12. It is hard for me to decide what the main ideas are in what I read.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

13. When work is hard I either give up or study only the easy parts.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

27. When I study I put important ideas into my own words.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

28. I always try to understand what the teacher is saying even if it doesn't make sense.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

29. When I study for a test I try to remember as many facts as I can.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

30. When studying, I copy my notes over to help me remember material.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

31. I work on practice exercises and answer end of chapter questions even when I don't have to.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

32. Even when study materials are dull and uninteresting, I keep working until I finish.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

33. When I study for a test I practice saying the important facts over and over to myself.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

34. Before I begin studying I think about the things I will need to do to learn.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

35. I use what I have learned from old homework assignments and the textbook to do new assignments.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

36. I often find that I have been reading for class but don't know what it is all about.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

37. I find that when the teacher is talking I think of other things and don't really listen to what is being said.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

38. When I am studying a topic, I try to make everything fit together.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

39. When I'm reading I stop once in a while and go over what I have read.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

40. When I read material for a class, I say the words over and over to myself to help me remember.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

41. I outline the chapters in my book to help me study.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

42. I work hard to get a good grade even when I don't like a class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

43. When reading I try to connect the things I am reading about with what I already know.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

In terms of this experiment,

1. How well can you study when there are other interesting things to do?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

2. How well can you concentrate on the material?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

3. How well can you remember information presented?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

4. How well can you avoid getting distracted?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

5. How well can you motivate yourself to study the items?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

Please answer ALL the following questions using the following scale:

+4 = very strong agreement
+3 = strong agreement
+2 = moderate agreement
+1 = slight agreement
0 = neither agreement nor disagreement
-1 = slight disagreement
-2 = moderate disagreement
-3 = strong disagreement
-4 = very strong disagreement

1. I would prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun.
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.
6. I find satisfaction in deliberating hard and for long hours.
7. I only think as hard as I have to.
8. I prefer to think about small, daily projects to long-term ones.
9. I like tasks that require little thought once I've learned them.
10. The idea of relying on thought to make my way to the top appeals to me.
11. I really enjoy a task that involves coming up with new solutions to problems.
12. Learning new ways to think doesn't excite me very much.
13. I prefer my life to be filled with puzzles that I must solve.
14. The notion of thinking abstractly is appealing to me.
15. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.
16. I feel relief rather than satisfaction after completing a task that required a lot of mental effort.
17. It's enough for me that something gets the job done; I don't care how or why it works.
18. I usually end up deliberating about issues even when they do not affect me personally.

Brief Mood Introspection Scale (BMIS)

by John D. Mayer

INSTRUCTIONS: Circle the response on the scale below that indicates how well each adjective or phrase describes your present mood.

(definitely do not feel) (do not feel) (slightly feel) (definitely feel)

XX X V VV

Lively	XX X V VV	Drowsy	XX X V VV
Happy	XX X V VV	Grouchy	XX X V VV
Sad	XX X V VV	Peppy	XX X V VV
Tired	XX X V VV	Nervous	XX X V VV
Caring	XX X V VV	Calm	XX X V VV
Content	XX X V VV	Loving	XX X V VV
Gloomy	XX X V VV	Fed up	XX X V VV
Jittery	XX X V VV	Active	XX X V VV

Overall, my mood is:

Very Unpleasant	Very Pleasant

-10 –9 –8 –7 –6 –5 –4 –3 –2 –1 0 1 2 3 4 5 6 7 8 9 10

Appendix 5. Class Survey

Please rate these questions, in terms of this class.
Rate the items on a scale of 1 to 7, with 1 meaning "not at all true of me" and 7 meaning "Very true of me"

2. I prefer classwork that is challenging so I can learn new things.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

3. Compared with other students in this class, I expect to do well.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

4. I am so nervous during a test that I cannot remember facts I have learned.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

5. It is important for me to learn what is being taught in this class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

6. I like what I am learning in this class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

6. I am certain that I can understand the ideas taught in this course.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

7. I think I will be able to use what I learn in this class in other classes.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

8. I expect to do very well in this class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

9. Compared with others in this class, I think I'm a good student.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

10. Even when I do poorly on a test I try to learn from my mistakes.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

11. I am sure I can do an excellent job on the exams for this class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

12. I have an uneasy, upset feeling when I take a test.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

13. I think I will receive a good grade in this class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

14. I think that what we are learning in this class is interesting.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

15. Understanding this subject is important to me.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

16. My study skills are excellent compared with others in this class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

17. I think what I am learning in this class us useful for me to know.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

18. Compared with other students in this class I think I know a great deal about the subject.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

19. I know that I will be able to learn the material for this class

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

20 I worry a great deal about tests.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

21 When I take a test I think about how poorly I am doing.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

22. When I study for a test, I try to put together the information from class and from the book.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

23. When I do homework (studying or working on the study guide), I try to remember what the teacher said in class so I can answer the questions correctly.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

24. I ask myself questions to make sure I know the material I have been studying.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

25. It is hard for me to decide what the main ideas are in what I read.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

26. When work is hard I either give up or study only the easy parts.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

27. When I study I put important ideas into my own words.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

27. I always try to understand what the teacher is saying even if it doesn't make sense.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

28. When I study for a test I try to remember as many facts as I can.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

29. When studying, I copy my notes over to help me remember material.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

29. I work on practice exercises and answer end of chapter questions even when I don't have to.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

30. Even when study materials are dull and uninteresting, I keep working until I finish.

1 2 3 4 5 6 7

1 = not at all true of me 7 = very true of me

31. When I study for a test I practice saying the important facts over and over to myself.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

32. Before I begin studying I think about the things I will need to do to learn.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

33. I use what I have learned from old homework assignments and the textbook to do new assignments.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

34. I often find that I have been reading for class but don't know what it is all about.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

35. I find that when the teacher is talking I think of other things and don't really listen to what is being said.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

36. When I am studying a topic, I try to make everything fit together.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

37. When I'm reading I stop once in a while and go over what I have read.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

38. When I read material for this class, I say the words over and over to myself to help me remember.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

39. I outline the chapters in my book to help me study.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

40. I work hard to get a good grade even when I don't like a class.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

41. When reading I try to connect the things I am reading about with what I already know.

1 2 3 4 5 6 7
1 = not at all true of me 7 = very true of me

For the next several questions, indicate how much you agree with each statement:
42. You have a certain amount of intelligence and you cannot do much to change it.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

43. Difficulties and challenges prevent you from developing your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

44. The effort you exert improves your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

45. If you fail in a task, you question your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

46. Criticism from others can help develop your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

47. You can develop your intelligence if you really try.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

48. Good performance in a task is a way of showing others that you are intelligent.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

49. When you exert a lot of effort, you show that you are not intelligent.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

50. When you learn news things, your basic intelligence improves.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

51. If you fail in a task, you still trust your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

52. Performing a task successfully can help develop your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

53. Your abilities are determined by how intelligent you are.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

54. Good preparation before performing a task is a way to develop your intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

55. You are born with a fixed amount of intelligence.

1 2 3 4 5 6 7
1 =Strongly Disagree 7 = Strongly Agree

In terms of this course,

6. How well can you finish your homework on time?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

7. How well can you study when there are other interesting things to do?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

8. How well can you concentrate on your schoolwork?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

9. How well can you remember information presented in class and in your textbook?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

10. How well can you arrange a place to study at home where you won't get distracted?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

11. How well can you motivate yourself to do schoolwork?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

12. How well can you participate in class discussions?

1 2 3 4 5 6 7
1 = not at all well 7 = very well

13. What grade are you trying to achieve in this course (Honestly)?

14. What grade do you honestly expect to get in this course?

15. How many classes are you taking this summer?

16. How busy do you estimate that you are this summer, with classes, work, and other commitments?

1 2 3 4 5 6 7
1 = not at all busy 7 = very busy

17. What is your major?