DISSERTATION

UNDERMINING LEARNING: THE IMPACT OF REWARDS ON LEARNING BEHAVIOR

Submitted by

Hillary Wehe

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2016

Doctoral Committee:

Advisor: Carol Seger

Matthew Rhodes
Bradley Conner
Kim Hoke

ABSTRACT

UNDERMINING LEARNING: THE IMPACT OF REWARDS

ON LEARNING BEHAVIORS

The undermining effect suggests that external rewards can decrease levels of internal motivation. Research exploring student motivation shows that internally motivated students appear to engage longer and in more challenging tasks compared to students focused on external rewards or performance feedback. The current study tested variables that may decrease susceptibility to motivational undermining for learning behaviors. In all studies, students were assigned to either a reward or non-reward condition and completed a word-learning task followed by a final test. Subjects were given the option to choose to re-study the words at two times during the task—pre- (while reward is still achievable) and post-test (after reward is given and no further extrinsic reward is achievable). Across all studies, an undermining effect was expected: Non-reward subjects would spend a greater amount of time reviewing the words during the post-test interval compared to the reward group. Study 1 directly tested the hypothesis by observing whether or not the reward groups behaved differentially at the pre- and post-test choice. Reward subjects spent significantly less time engaging in the task during the post-test review phase, supporting the presence of the undermining effect (t (1,60)=2.06, p = .02, 1-tailed) but a 2 (group: reward x non-reward) x 2 (study time: pre-test x post-test) repeated measures ANOVA comparing the mean study times for the reward and non-reward subjects' pre-test study and post-test review time revealed that the interaction between group and study time did not reach significance (F (1,60) = 3.52, p = .065).

Study 2 was identical to the first study but with the addition of a surprise, 24-hour delayed memory test to examine whether the extra post-test study had beneficial effects on long-term retrieval. Non-reward subjects were hypothesized to recall more items on a delayed memory test compared to reward subjects due to increased study time. A 2 (group: reward x non-reward) x 2 (study time: pre-test x post-test) repeated measures ANOVA was conducted to compare the mean study times for the reward and non-reward subjects' pre-test study and post-test review times. The interaction between group and time spent on task was significant (F (1,241) = 4.24, $p < .05$) but there was not a significant main effect for the between subjects variable of reward on the amount of time spent engaging in the task during the pre- and post-test phases (F (1)= .63, $p = .44$). A 2 (group: reward vs. non-reward group) x 2 (test performance: immediate x delayed) repeated measures ANOVA was conducted to compare the average accuracy between groups on the delayed memory test. There was not a main effect of group on performance (F (1, 110) = .82, $p = .38$) and the interaction between reward group and immediate or delayed test was not significant (F (1,156) = .201, $p = .65$).

Study 3 was similar to the first study but subjects were allowed to choose the material they were learning (i.e., Swahili or Lithuanian words). The element of choice was expected to increase the degree of control and internal motivation students experienced and consequently decrease the effect of undermining between the reward and non-reward group. Specifically, study times between the reward and non-reward group were hypothesized to be equal between groups and higher than then a forced choice condition. A 2 (group: reward x non-reward) x 2 (choice: self-determined x forced-choice) x 2 (study time: pre-test x post-test) x 2 (language: Swahili x Lithuanian) repeated measures ANOVA was conducted. The main effect of choice condition was not significant (F (1,60) = .140, $p = .71$). The main effect of reward was also not significant (*F* (1,60) = .920, *p* = .34) but the interaction between choice and reward on time spent on task was significant (*F* (1,60) =4.11, *p* < .05). A 2 (group: reward x non-reward) x 2 (choice: self-determined x forced-choice) repeated measures ANOVA was conducted to compare performance on an immediate memory test for the self-determined and forced choice group but the effect was non-significant (*F* (1,60) = .67, *p* = .16); in addition, there was not a significant main effect of reward (*p* =.32) nor was there an interaction (*p* = .16).

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Motivation is a broad-reaching construct that encompasses many aspects of behavior. As Graham and Weiner (1996) defined it, "…motivation is the study of why people think and behave as they do" (p. 63). Considering these aspects are important for understanding variability both in individual action choices and dedication towards a goal (Berridge, 2004) but research on this topic is not cohesive across disciplines. Commonly, reward outcomes are assumed to explain human behavior and can be especially prominent in educational and business fields (Schnieder, 1985). Within the memory research, many studies have examined the impact of external rewards, or points, on memory for individual, item-by-item information (Adcock et al., 2006; Castel, 2008; Cohen et al., 2014; Knutson et al. 2000). Generally monetary incentives and points benefit memory for high valued items compared to low, or non-rewarded items. However, other studies indicate that removal of external rewards can be harmful to intrinsic motivation and can cause longer-lasting detriments to intrinsic motivation than? (Deci, Koestner & Ryan, 1999).

The current study seeks to better explore the relationship between rewards and learning. Specifically, I will examine the impact of external rewards on learning, the impact of external rewards on intrinsic motivation, the impact of motivation on self-regulated learning behaviors, and finally, the impact of motivation on broad learning outcomes. Additionally, the current paper discusses a series of experiments that seek to understand how external or internal motivational states interact with students' self-regulated learning behavior.

Extrinsic Reward and Learning

The incentive theory of motivation posits that positive reinforcement (e.g., the addition of a pleasant stimulus or outcome) will encourage behavior to be repeated and strengthen motivation for a task (Berridge, 2004). It therefore seems intuitive that if we want people to perform a task we should provide incentives, and if we want people to improve their performance we should provide stronger incentives. This external offering of rewards is referred to as an extrinsic motivator—a positive enticement that is external to the individual motivating behavior (Cameron & Peirce, 1994). The impact of incentives is of special interest in training and teaching because employers and educators seek to maximize the efficiency and the efficacy of training programs, and research has demonstrated that extrinsic motivators can be contextually beneficial to performance.

Incentives that emphasize important information have positive benefits for directing subjects' attention and improving self-regulatory behaviors. In one study, older and younger adults were provided a list of words with different point values associated with them. Their overall goal during the task was to earn as many "points" as possible through successful recall of the words (Castel, Benjamin & Craik, 2002). Younger adults, overall, recalled more words than older adults. However, in both immediate and delayed recall conditions, older adults were more likely to recall high value words and recalled fewer low value words compared to younger adults. That is, older adults were more selective about what they were encoding. Selectivity of memory refers to allocating attentional resources to specific information and disregarding other pieces (Castel et al., 2002; Castel, 2008). Using external rewards to emphasize information is referred to as "value-directed" learning and has also been used to demonstrate that subjects will independently choose to study items that are allocated a higher point value (Ariel, Dunlosky, & Bailey, 2009; Castel et al., 2013; Soderstrom & McCabe, 2011).

Utilizing value to direct attentional and cognitive resources is common in other fields as well. In a review of perceptual skill learning, Goldstone (1998) identified attentional weighting as a crucial area for skill development. When subjects are able to successfully limit their attention to what is important about a task they are better able to encode that information, which could result from successfully attending to relevant information and ignoring irrelevant information.

Sarter and Gehring (2006) refer to motivated allocation of attention as a 'cognitive incentive' that is directed at goal optimization. Research comparing sustained attention between paid and non-paid subjects indicates that those receiving external incentives exerted more effort to sustain attention in comparison with non-paid subjects who exhibited attentional declines over time on task (Sarter & Gehring, 2006). They argue that motivation is a top-down process that allocates resources to regain or sustain attention. Sarter and Gehring (2006) provide evidence from non-human animal studies indicating that when an externally motivated organism is subjected to additional attentional demands, top-down, frontal cholinergic secretions increase. When rats remain motivated to complete a task, despite increased demands, they will compensate with other "cognitive" or physiological mechanisms in order to complete the task.

Therefore, it is possible that motivation is sustaining and directing attentional resources and the motivational system's energy expenditure supporting attentional direction should parallel that of other physical tasks. This parallel was demonstrated in cognitively demanding tasks in which human attention was needed to maintain successful performance throughout time on a task. In one such study, subject performance was incentivized for either the Stroop task (i.e. a cognitive task) or a hand-grip exercise (i.e., a physical task), activity in the midbrain's reward center, specifically the ventral striatum, correlated with and predicted effort expenditure in both tasks (Schmidt et al., 2012; Meyniel et al., 2013). Additionally, Hubner and Schlosser (2010) found that subjects' accuracy on a flanker task, thought to rely on executive attentional systems, was improved for subjects offered an incentive across a range of forced response times. As time available for response time decreased, both rewarded and non-reward subjects showed a speed-accuracy trade-off, but rewarded subjects remained more accurate. Hubner and Schlosser argued that the motivational context increased preparation for perceptual processing and improved item processing. Similarly, Pessoa and Englemann (2010) showed that stimulus identification in a selective attention task improved in step with incentives. Subjects were able to become more selective as incentives increased. They argued that motivation acts both to improve sensory processing but also to integrate value information and attentional (and, more broadly, cognitive task) information.

The previous results indicate that motivation can be used to enhance or direct the allocation of attentional resources. Similarly, research has shown that incentives or value information can be integrated by other systems in order to improve processing for other functions such as working memory or encoding processes. Many researchers argue that value information is represented in midbrain reward centers (such as the ventral striatum and ventral tegmental area (VTA) which project to task relevant hubs (such as the frontal lobe, parietal lobe, or medial temporal lobe). Adcock et al. (2006) showed that anticipation of a high-reward task compared to a low-reward task recruited more ventral striatum activity and increased functional connectivity between ventral striatum, VTA and hippocampus. The increased functional connectivity predicted the likelihood that rewarded items would subsequently be remembered. Anticipation of high value rewards is shown to recruit reward centers and functional connectivity between reward centers and medial temporal lobe (MTL) regions predict better memory performance (Adcock et al., 2006; Cohen et al. 2011; Gruber et al., 2013). It appears that co-recruitment and connectivity between reward, MTL, and often, frontal executive areas enhances performance.

Taylor et al. (2004) used fMRI to explore motivational effects on working memory performance. They varied working memory load within an object working memory task, which they could separate distinct stages including encoding images, holding them in working memory across a delay, and making mental comparisons of a target stimulus to the stored representations. Subjects had to compare a target stimulus to either three different shapes (high load) or to three identical shapes (low load) held in working memory. They were rewarded either $1 (high reward) or $.01 (low reward) for correct matches and also were rewarded or penalized for correct rejections and false alarms respectively. Taylor et al. found that high-reward trials improved working memory performance compared to low-reward trials and increased neural activity in dorsolateral PFC regions thought to be responsible for working memory performance.

If rewards direct attention to task-relevant information, they should help learners more efficiently focus on specific information. Surprisingly, despite their effects on attention, incentives have not been found to significantly improve memory performance in many memory tasks. For instance, Nilsson (1987) tested whether or not word learning was impacted by external motivation during the study or test phase. Subjects were informed of performance contingent rewards either prior to a study phase, after the study phase but prior to the test, or not at all. There was no difference in memory performance between any of the groups. Murayama and Kuhbandner (2011) found slightly different results in a study examining learning the answers to trivia questions. They cued subjects in a reward group as to the potential reward associated with each item individually at study. The non-reward group was not offered any monetary compensation. Subjects completed an immediate test, a surprise delayed test a week later, and rated the interest level of each question. The presence of reward made little difference to recall except on delayed tests and for material that was rated as uninteresting to the subjects. In these instances, recall was significantly higher for the reward group. Murayama and Kuhbandner concluded that monetary incentives create only a slight advantage for learning. Both of these studies find that reward motivation does not negatively impact learning, but also has, if any, minimally positive benefits (see also Craik & Tulving, 1975; Wehe, Rhodes, & Seger, 2015). Neither study accounted for any changes in learning behaviors elicited by the presence of reward or allowed students to change their study habits in response to rewards.

When subjects try to increase effort in response to incentives it does not always translate into improved performance. For example, Pochon et al. (2002) used fMRI to explore the how the neural substrates for motivation interacted with the neural substrates involved in working-memory tasks. Using an n-back task Pochon et al. studied three varying levels of reward: none, minimal, or high reward. Subjects were given a cue before each trial to indicate whether they were completing a rewarded or non-reward trial. Rewards were based on accuracy. In contrast to results, such as the Taylor et al. (2004) study described previously, Pochon found increased activation for working memory (lateral PFC and lateral frontopolar) areas when subjects were working for rewards but they did *not* find a difference in working memory performance. Despite reporting similar neural and behavioral results between the two studies, both used a different memory task, and additionally Taylor et al. (2004) reported that subjects perceived that they had "tried harder" to do well.

A key difference between Pochon et al. and the other findings could be that subjects in other studies (Locke & Braver, 2010; Taylor et al., 2004) were completing task that had a strategy or 'short-cut' that was intuitive to apply, which increased performance. Another possibility is that that Pochon et al.'s subjects had already maximized their working memory capacity and could not further improve their performance.

If motivation acts to allocate resources that are needed for a task, motivational and incentive effects should reliably result in improved performance when and only when there can be an increased or improved allocation of resources. Developmental studies comparing memory processing and dopaminergic functioning between younger and older adults support this prediction. As Mather and Schoeke (2011) reported, dopamine within the reward and midbrain centers declines substantially across aging. Consequently dopaminergic agonists have a significantly more "motivational" effect on older adults' learning performance compared to younger adults who theoretically are already at their "maximum" functioning in both motivational system and strategy application.

In conclusion, motivating learning is not as simple as increasing monetary incentives in order to increase memory retention. Together, the studies reviewed here imply when resources, or an effective strategy, to improve performance are unavailable that application of incentives will not provide a strong benefit (Craik & Tulving, 1975; Kang & Pashler, 2014; Nilsson, 1987). Researchers argue that in studies where providing value information does lead to memory influences, it is because the point values (as seen in value-directed learning) are inherently creating selective strategies that direct the control of cognitive resources (Castel, 2008; Cohen et al., 2014) at the expense of less valuable information.

Intrinsic Motivation and the Undermining Effect

Consideration of external stimuli alone still cannot fully explain behavior because incentives do not have a constant value. Gottfried, O'Doherty, and Dolan (2003) demonstrated that individual valuation of incentives varies based on internal, motivational states. The researchers conditioned subjects to associate random visual images to delivery of food related and non-food related odors. During the first block of the task subjects were in a fasted state.

Before the second block, subjects were satiated with one of the food options represented by the odor cue in the previous trial. During the second block, both motor reaction times and activity in the amygdala and OFC showed decreased responsiveness to previously pleasant, conditioned cues once subjects were in a satiated state, supporting the notion that our subjective states can devalue previously conditioned stimuli.

Further still, many tasks are pursued only for personal fulfillment, due to intrinsic motivation, in the absence of external incentives. Emphasizing external rewards and performance outcomes may inhibit individuals' motivation to complete a behavior for personal satisfaction, a phenomenon known as motivational undermining (Deci, 1971). Researchers disagree on constructs, and even the existence, of intrinsic motivation. White (1959) argued for a broadened view of motivation, past biological drives that described a limited view for satisfying survival requirements. White posited that biological needs are not all that humans seek to fulfill and reinforcement learning does not explain all of our behaviors. Comparably, Maslow argued that physiological states are important to consider because they can direct what an individual is seeking or is motivated to approach but once those physiological needs, such as hunger, are satisfied, the individual will have different motivations. He provides the example of a starved man who finally has an abundance of food. After the motivation to fulfill that drive is complete, "at once other (and "higher") needs emerge..." (Maslow, 1943, p. 375). Once humans have successfully satisfied physical needs, learned from reinforcement, or have been shaped by their environment, there is still motivation for interactions independent from contingent rewards. Such interactions are not well defined by past theories of motivation that focus on drives and later researchers argued that this residual motivation in our behavior can be described as intrinsic (Deci & Ryan, 1985; White, 1959).

Much of the research conducted on long-term and intrinsic aspects of motivation is in the field of education (Ryan, 2012), and several correlational studies point towards the importance of engagement and interest in improving performance outcomes (Covington, 2000; Henderlong & Lepper, 2002). For example, recent research found that variables related to intrinsic motivation, such as subjects' interest, enjoyment, or valuation of a task, are positively related to high academic achievement; whereas variables of pressure and tension, that are associated with decreased motivation for a task, are positively related to low academic achievement (Malik & Parveen, 2015). The researchers argue that students who are more intrinsically motivated are more likely to engage in and excel at academics, but research has not directly explored how variations in extrinsic and intrinsic motivational states influence learning processes, and how the two forms of motivation may interact. For example, it is very common for schools to rely on reinforcement and extrinsic motivation but researchers have posited that once extrinsic rewards are removed, students' intrinsic motivation to engage in those tasks diminishes (Cameron & Eisenberg, 1994).

Extrinsic motivators have been shown to compete with and lower inherent motivation for tasks, as shown in the undermining effect. The undermining effect is commonly demonstrated through a free-choice paradigm (Deci et al., 1999). Subjects in a reward condition are provided an incentive to participate in a task for a certain period of time and then frequency of task engagement is observed during a free period once the reward is removed. The amount of time spent engaging in the experimental task is commonly used as a measure of intrinsic motivation. Often, the removal of a reward will significantly impact intrinsic motivation and reduce the time spent on a task compared to a group that did not receive the incentive (Deci et al., 1999).

Deci and colleague's early development of the undermining paradigm used a seven-piece puzzle kit named the Soma puzzle (Deci & Cascio, 1972). Soma puzzle pieces can be combined together in order to create larger shapes. Images of configurations are provided that the subjects need to create with the puzzle pieces. In these original studies exploring intrinsic motivation, subjects would complete three sessions of Soma puzzle completion. One group was paid for every puzzle correctly replicated but otherwise was identical to the control group. Each session would last for an hour followed by 8 minutes for "free-choice" time. Time spent engaging with the puzzle task during the "free choice" period was regarded as resulting from intrinsic motivation because subjects chose to participate of their own volition (i.e., there was no reward offered). Initial results indicated that pay for performance decreased free choice time spent engaging in the task compared to control subjects. The researchers argued that external rewards reinforced behavior and increased reliance on the rewards, leading to decreased intrinsic motivation.

In subsequent extensions of this first study, the researchers demonstrated that other forms of feedback could either be detrimental or beneficial to intrinsic motivation. For example, one study conducted by Deci and Cascio (1972) manipulated negative feedback and threat of punishment. For the experimental group, the researchers increased the difficulty level of the puzzle to ensure that subjects frequently failed. This failure was operationally defined as negative feedback and because it was based on subjects' ability to complete the task, it was considered to be "self-administered". With threat of punishment, subjects were given ten minutes to solve the puzzle and told that if they did not an annoying buzzer would sound for 1 second. The control group engaged in the task during the free choice period for nearly two more minutes than the group that received negative feedback (i.e. the group that was able to complete fewer puzzles). The control group also engaged in the task for 90 seconds longer than group that was threatened with punishment; however, it was not significantly longer.

In other extensions, providing external rewards in the form of positive, verbal feedback increased intrinsic levels of motivation. This was operationalized by time spent solving a puzzle during a free-choice period. Internalizing verbal rewards, Deci argues, may be comparable to internal feelings of satisfaction (Deci, 1971; Deci & Cascio, 1972). In these studies, reward and feedback were contingent on performance. Specifically, payment was not received unless subjects met the set criteria and, likewise, feedback was provided that directly reflected performance. These studies laid a foundation for several that have utilized a free-choice paradigm in order to define the parameters of reward and feedback that impact internal motivation.

Deci et al. (1999) conducted a meta-analysis of undermining effect studies that included a total of 128 studies comparing a reward group with a non-reward group. All studies examined the effect of a reward on the amount of time or number of times subjects participated in the experimental task during an optional practice period. Their primary analysis of 101 papers investigating the impact of rewards on subjects' voluntary behavior during a free-choice period yielded a mean effect size of $d = -.24$. Based on their review of studies examining the relationship of extrinsic and intrinsic motivation, the researchers concluded that rewards, across a variety of tasks, significantly impacted intrinsic motivation to perform the task once the reward was removed.

In their review they also discussed how personality variables and type of reward affected motivation. Deci et al. (1999) referred to their theory of how individual differences interact with perceptions of reward as cognitive evaluation theory (CET). CET states that people seek different levels of control in a situation. If an individual prefers to be in control of a situation then external reward or feedback guiding their actions might be interpreted more negatively. Consequently, the emphasis a person places on *environmental* measures of performance may determine their susceptibility to the motivational undermining by external reward. In similar form, CET remained a popular explanation for individual differences in perceived ability and motivation to continue self-improvement (Elliot & Dweck, 1988; Miele, Finn, & Molden, 2011) and will be discussed more in depth in regards to individual achievement goals.

In addition to perceptions of reward, Deci et al. (1999) explored how the application of different types of reward could affect the degree of internal motivation by either signaling increased regulation through reward structure or, alternatively, fostering positive benefits through encouragement. The first type of reward described was task-noncontingent. Subjects are rewarded simply for being present during the task and neither accurately completing the task nor finishing the task is required for reward. Task- noncontingent rewards conceivably apply the least amount of control and resultant undermining of internal motivation because pressure to participate in the task is limited and no feedback is provided about performance. Another reward type was engagement-contingent reward, which required active task participation. Directing subjects but not providing feedback on behavior was predicted to negatively impact internal motivation because it asserted increased amounts of control. Similar to engagement-contingent, completion-contingent rewards provide more external direction for activity because subjects only receive a reward if they finish the task; however, they also potentially receive positive confirmation if finishing the task signals an accomplishment. The final established reward was performance-contingent rewards, threatening the highest impact to internal motivation but conversely has the most beneficial effect on performance overall. Subjects are only provided a reward if performance is at, or above, a set criterion. If feedback is negative or supervisory, then it increases the likelihood of undermining intrinsic reward. If performance feedback is informative and positive, then it increases the chances of subjects experiencing positive effects on motivation.

Deci et al. (1999) concluded that increasing environmental control negatively affects internal motivation. He argues that internal motivation is a healthy, natural drive that encourages improvement and leads to self-sufficiency. Therefore, although it is important to receive constructive feedback for our goals, a balance between control and positive encouragement needs to be achieved. When task completion is entirely dependent on external rewards, motivation for that task is externally regulated; however, the broader theory of self-determination theory (SDT), similar to locus of control, identifies stages in which externalization of rewards can be accepted and internalized by the individual. Internalizing the rewards increases internal meaning and value. There are three classifications and "levels" through which external contingencies can be internalized: introjection, identified regulation, and integrated regulation. Introjection mirrors others' values and is a shallow form of internalization in which self-worth is now invested and contingent on the external outcome. Personal investment is increased but motivation is still at risk of environmental effects. Identified regulation is internalization and comprehension of the importance of a reward but the importance may nonetheless be at odds with personal goals. For example, an individual may understand that their current job is important for them to earn a living but it is not fulfilling their ultimate career goals. Integrated regulation implies a full acceptance of value or importance and coincides with personal values—they have been internalized to be harmonious. Receiving positive or constructive feedback can result in gains for intrinsic motivation. When a reward is internalized and results in an increase in motivation it is referred to as “crowding-in”.

Recent interest in the undermining effect was renewed through exploration of the neural mechanism responsible for guiding motivation. Murayama et al. (2010) tested whether subjects offered money for each successful completion of a task would show decreased motivation to perform the experimental task during a free period once the monetary feedback was removed. They used functional magnetic resonance imaging (fMRI) to investigate whether or not the processing of extrinsic and intrinsic reward affects neural activity in similar ways in conditions of intact and undermined motivation. Murayama et al. emphasized that the task used for exploring effects of motivational undermining needed to be personally interesting to create a valid result.

Murayama and colleagues had a reward and non-reward group complete two sessions of a stop-watch and watch-stop task. For the stop-watch task, a cue indicated when subjects needed to stop the watch. For example, they were cued to stop the watch at 5 seconds. The trial was successful if the watch was stopped within 50 milliseconds of the cue time. The watch-stop task was a passive version of the stop-watch task, merely requiring subjects to press a button when the watch stopped. During the first session, the reward group earned money for each successful trial but the non-reward group was offered a flat rate of money in exchange for task participation. Because the non-reward group received no feedback or reward based on their performance, their behavior during the subsequent experimental task was thought to be internally motivated. Both groups performed the stop-watch or watch-stop tasks while fMRI data was collected. In between the sessions, each participant was told they had free time to perform any activities available, including the experimental stop-watch task. The experimenters recorded how often each participant chose to engage in the stop-watch task during that time, and used that as a measurement of internal motivation. No money was earned for the second session. The absence of reward during the second session was intended to undermine motivation.

As expected, the behavioral results revealed a significant difference between the amount of time the reward group spent on the task during the free time period compared to the non-reward group. Subjects who were not provided any monetary compensation engaged more frequently in the task between sessions. Neurally, the fMRI results showed interesting interactions between the reward and non-reward group and the first and second session. In the first session the reward group activated the lateral prefrontal cortex, midbrain, and anterior striatum significantly more than the non-reward group; however, activity in all these areas decreased in the reward group once monetary incentives were withdrawn resulting in greater activity in those regions in the non-reward group. Midbrain and striatal regions were thought to be active because of their role in encoding and processing rewards, and frontal activity may reflect preparation for future tasks (Murayama et al., 2010).

Albrecht et al. (2014) attempted to replicate Murayama et al.'s previous fMRI findings of the undermining effect using both monetary and verbal rewards to additionally explore the neural correlates of the motivational crowding-in effect of verbal rewards. The researchers had subjects complete three phases of a perceptual discrimination task. Subjects were in a reward, non-reward or control group. During the first phase, all subjects completed the task with no feedback to gather baseline information. In a second phase, subjects experienced one of three conditions: performance-contingent rewards (in the reward group), verbal feedback (in the non-reward group), or no changes (in the control group). During the third phase, all verbal and monetary rewards were removed. The authors predicted the addition of monetary incentives would cause an increase of blood flow in the striatum and dopaminergic midbrain regions but that this activity would decrease by the third phase, consistent with previous investigations of the undermining effect. Additionally, they speculated that an increase in the same areas with the addition of verbal feedback would be indicative of subjective internalization of rewards, consistent with explanations of SDT and internalizing rewards.

They found that the addition of both monetary and verbal rewards in the second block of the task increased neural activity within the striatum and midbrain regions. During the third block, the monetary and verbal incentives were removed. For the monetary reward group, neural activity did not decrease compare to a control group, which would replicate the undermining effect. For the verbal rewards group, higher activity was sustained relative to a control condition—indicating that feedback had an intrinsically positive value. In addition to increased striatal activity, only the verbal reward group continued to improve their performance across all trials of the task. Albrecht posited that the performance improvement was a reflection that the internalization of feedback has subjective value. These results complement Murayama et al. (2010)'s study by demonstrating the sustained ability of subjects to remain engaged in a task while only receiving cognitive feedback for performance. Activity is increased in response to verbal rewards because they are incentivizing—meaning that they encourage pursuit and motivational direction to achieve them and can foster motivational crowding-in.

Motivation and Self-Regulated Learning

Learning is an active process and an individual's engagement in the process can greatly facilitate the amount and quality of information retained. Self-regulated learning is an approach to learning that involves variables of interest, application of cognitive resources, and the ability to properly monitor and gauge understanding (Clayton et al., 2010). Kornell and Bjork (2007) discuss several relevant decisions that an individual has to make while studying and two, in particular, are of interest to this review. The first is a deliberate, conscious decision of which items to study. The second, how long to study an item. Previous research exploring the effect of reward on decision to study and total time spent studying suggest that rewards may impact both of these aspects of self-regulated learning (Wehe, Rhodes, & Seger, 2015).

Inherent in the ability to successfully self-regulate learning is the ability to monitor knowledge and apply effective strategies to maximize learning (Bjork, Dunlosky & Kornell, 2013; Kornell & Bjork, 2007). One method used to examine self-regulated study behavior involves providing subjects with information to learn and allowing them to choose which items to re-study. One such study considered subjects' self-reported confidence and their decision to restudy items (Dunlosky & Hertzog, 1997). Dunlosky and Hertzog provided subjects with word pairs to learn. Later they were presented with half of the pair (i.e., a retrieval cue) and asked to judge how likely they would be to correctly recall the correct paired associate later. Subjects' judgments were highly accurate, and their judgments and recall ability were highly correlated with their restudy choices. Subjects were more likely to choose an item for restudy when they were not confident they would be able to recall that item on a future test. These results are supportive of a discrepancy reduction model for self-regulated approaches to learning; students apply a strategy that is focused on decreasing the distance between what they do and do not know.

Son and Metcalfe (2000) demonstrated that the relationship between item difficulty and time spent studying depends on the total time available for studying. They found that when subjects had ample time to study material, they were more likely to devote time to the more difficult items but when they were under time pressure they devoted more time to easier items.

These and other results indicate that students are more likely to apply a strategy for learning that is consistent with demands and goals of the task. For instance, students may ignore item difficulty when given more specific information that allows them to strategically maximize a score on an upcoming test. Ariel et al. (2009) weighted to-be-remembered items with varying reward values. Subjects were more likely to devote time restudying high reward information compared to low reward information. Castel et al. (2013) similarly found that both younger and older subjects strategically chose to allocate more study time to information associated with high point values than low point values. Ariel et al. (2009) and Castel et al. (2013) demonstrate that subjects attribute importance to high value information based on personal agendas and strategies they develop. These results are consistent with an agenda based theory of regulation.

Similarly, Thiede and Dunlosky (1999) argued that learners alter their strategy according to the level of effort that is required to meet their goals. The researchers manipulated the performance goal for subjects in order to determine how the goal interacts with subjects' strategy. Previous results examined the outcome of self-regulated strategies when the goal was chosen by the subject. They found that when subjects have an easier task goal they focus on learning information that most efficiently allows them to reach their goal and consciously regulate application of cognitive resources in order to maximize efficiency. The above studies demonstrate how the integration of external reward information into learning strategies can direct allocation of resources in order to minimize effort expended and maximize efficiency in reaching a learning goal. Students use prior knowledge and ability to determine the likelihood that items can be remembered (Metcalfe, 2002; Metcalfe & Kornell, 2005;Son & Metcalfe, 2000); however, a large amount of research indicates that task constraints, goals, and reward information is incorporated in order to determine students' agendas and direction of learning behavior.

Intrinsic Motivation and Overall Learning Outcomes

Motivational orientations are important for instructors to consider because motivation appears to be necessary for applying effective cognitive strategies, improving learning outcomes, and facilitating positive affective responses. Among undergraduate engineering students, high intrinsic motivation and task value were positively correlated with the likelihood that students would cognitively elaborate on class material (Stolk & Harari, 2014). Intrinsic motivation, task value, and self-efficacy were also correlated with rates of critical thinking and problem-solving skills. Metallidou and Vlachou (2007) found that elements of motivation including self-efficacy and competence were correlated with effective strategy application in both math and languages courses for elementary students. Grolick and Ryan (1987) measured differences in elementary school students' rote and conceptual knowledge when in a controlling or non-controlling learning environment. The researchers found that subjective interest for the material was negatively correlated with the level of control exerted over the task; however, direction and control led to short-term, increased rote learning (i.e., repetition of direct facts). Students in the non-controlling environment reported more subjective interest in the task, and showed greater long-term retention of rote information and increased conceptual knowledge compared to students in the controlling environment. These results indicate that consideration of students' motivational states is important for improving learning outcomes.

In addition to performance outcomes, White and Fantone (2010) measured satisfaction differences between traditional classroom grading and less restrictive, pass-fail systems. They reported that medical students reported higher satisfaction with less restrictive classroom design. Students being graded pass-fail also reported having more time to independently explore academic material and work on improving their personal happiness. In addition to affective changes, students maintained their high rates of performance and did not differ significantly in grade or instructor-observed performance. Overall, White and Fantone's qualitative observation supports the notion that instructors should aim to facilitate intrinsic motivation in order to improve students' self-regulated learning and motivation to engage in tasks outside of class.

For all of the following statistical analyses, the significance level was set to $p = 0.05$ unless otherwise noted (e.g., if family wise error corrections are conducted). If the statistical analysis showed the significance level was below 0.05, results will be discussed. If they did not fall below the cut-off, the null hypothesis was not rejected and the results were not discussed.

CHAPTER 2: STUDY 1 INTRODUCTION

The purpose of Study 1 was to explore the impact of rewards on subjects' study behavior before and after the removal of a performance based incentive. Previous research demonstrated an undermining effect on students' intrinsic motivation to engage in self-regulated learning behavior (Wehe et al., 2015). Intrinsic motivation was operationally defined as the amount of time students persisted on a learning task. Students given an external incentive were less likely to spend time voluntarily studying the experimental materials after the external incentive was removed in comparison with control subjects who never received external incentives. This effect is widely interpreted as reflecting a decrease in intrinsic motivation in the external incentive group (Tang & Hall, 1995; Deci et al., 1999; Murayama et al., 2010; Hagger & Chatzisarantis, 2011). The question still remains whether or not studying behavior is facilitated while working for a reward. In other words, do students that are learning in order to achieve an external reward exert more effort to learn material compared to students who are not learning for a reward? Study 1 examined whether students would voluntarily choose to study material while an external incentive was still achievable and after the external incentive was disbursed and no longer available. The first hypothesis was that while the external incentive was achievable, subjects in the reward group would voluntarily choose to study the experimental word pair associates more frequently, and for a longer duration, compared to subjects in the non-reward group. The second hypothesis was that the direction of this behavior would change after the distribution of external rewards; subjects in the reward condition would voluntarily choose to review the experimental word pair associates less frequently and for a shorter duration, compared to non-reward subjects.

Study 1 Method

Subjects

Subjects were recruited from the Psychology Department's research pool at Colorado State University. The pool consists of all undergraduate students in both PSY100 and PSY250 classes. The students are required to participate in research studies for the Psychology department and are compensated with class credit.

G-Power (Faul, Erdfelder, Lang, & Buchner, 2007) was used to estimate an *a priori* effect size of .3 and reported that a total of 40 subjects would need to be enrolled; however, 69 subjects participated in the study. Of these, a total of 7 subjects' data was excluded due to a failure to follow directions or computer error. 3 of the 7 reported that they forgot which keys to press to control the program, 1 subject spoke English as a second language and reported struggling to understand directions, and the remaining 3 pressed escape to terminate the program early. Subjects were randomly assigned into the reward group (n=31) or the non-reward group (n=31). For the statistical analyses testing the primary hypotheses all remaining 62 subjects were included and the a priori effect size was reached (d= -.31, F = 3.52).

Design and Materials

A 2 (group: reward x non-reward) x 2 (study time: pre-test x post-test) mixed design was used. Subjects in each group studied 30 Swahili-English word pairs. The word pairs were selected from a corpus developed by Nelson and Dunlosky (1994), who provide normative ratings for each pair for learning difficulty, likelihood of remembrance, and how closely the Swahili words match their English counterpart. Word pairs were selected from three difficulty levels, low, medium and hard, in order to examine the effect on performance of the difficulty level of stimuli. 10 word pairs were randomly selected from each difficulty level to create a stimuli list. Difficulty level was determined by proportion of correct paired associates recalled, as reported in Nelson and Dunlosky (1994). Stimuli in the low group had recall percentages above 58%; the middle group ranged from 25-58%, and the hard group was lower than 25%. The English words from non-studied word pairs were used as distractor items during the multiple-choice test and were matched for difficulty level. If there were not enough word pairs to use as test and distractor items, words were obtained from another corpus of normed stimuli and ratings of accessibility were used to match items on difficulty level (Nelson, McEvoy & Schreiber, 1998). PsychoPy (Peirce, 2009) was used to present stimuli during all phases of the experiment.

Procedure

Both groups were exposed to one study block of Swahili-English word pairs. All 30 of the word pairs were presented once in a random order. Each word was presented for 3 seconds. Before the session they were instructed to learn as many words as possible because they would be tested on them after the study session. Subjects in the reward group were additionally informed that if they answered at least 18 items correctly they would be entered into a drawing at the end of the semester for a gift card.

Following the study block, subjects completed math problems as a brief distractor task for 90 seconds. Subjects were shown multiplication, addition, or subtraction problems one at a time. The problem remained on the screen for 6 seconds and subjects used the keyboard to enter their solution after the problem disappeared. After the distractor task was the study choice phase. In this phase, subjects in both groups could choose to see the words again before taking the final test. Subjects were first given a prompt indicating they should press “y” to continue to study or “n” to continue to the final test. If subjects selected “y” they were further instructed to press “y” to see a new word or press “n” at any time to discontinue studying (See Figure 1.).

If subjects indicated that they wanted to study the words, subjects controlled the rate and number of words presented. The program presented the Swahili word as a cue and subjects would press ‘enter’ to see the English translation. After the subjects were done studying, or if they indicated that they did not want to study, they completed a multiple-choice memory test over all of the studied words. In the multiple-choice test, a Swahili word was presented at the top of the computer screen. Four English words labeled with letters A-D were listed below. Subjects selected their answer by pressing the corresponding letter on the computer keyboard. Distractor words were taken from unstudied word pairs as described above.

Once finished, the computer displayed the participant’s total number of correct answers and informed the reward group subjects whether or not they had qualified for the drawing (score of 18 or greater). After they reviewed their score, subjects could choose to see the words a third time during the review choice phase. The procedure was the same as the study choice phase. After subjects indicated they wanted to review the words for a final time, they were presented with the Swahili word and controlled the presentation rate of the English word and could end the presentation of words at any time. After the subjects were done reviewing or if they indicated that they did not want to review the words, they were provided with a completion screen and a debriefing form.

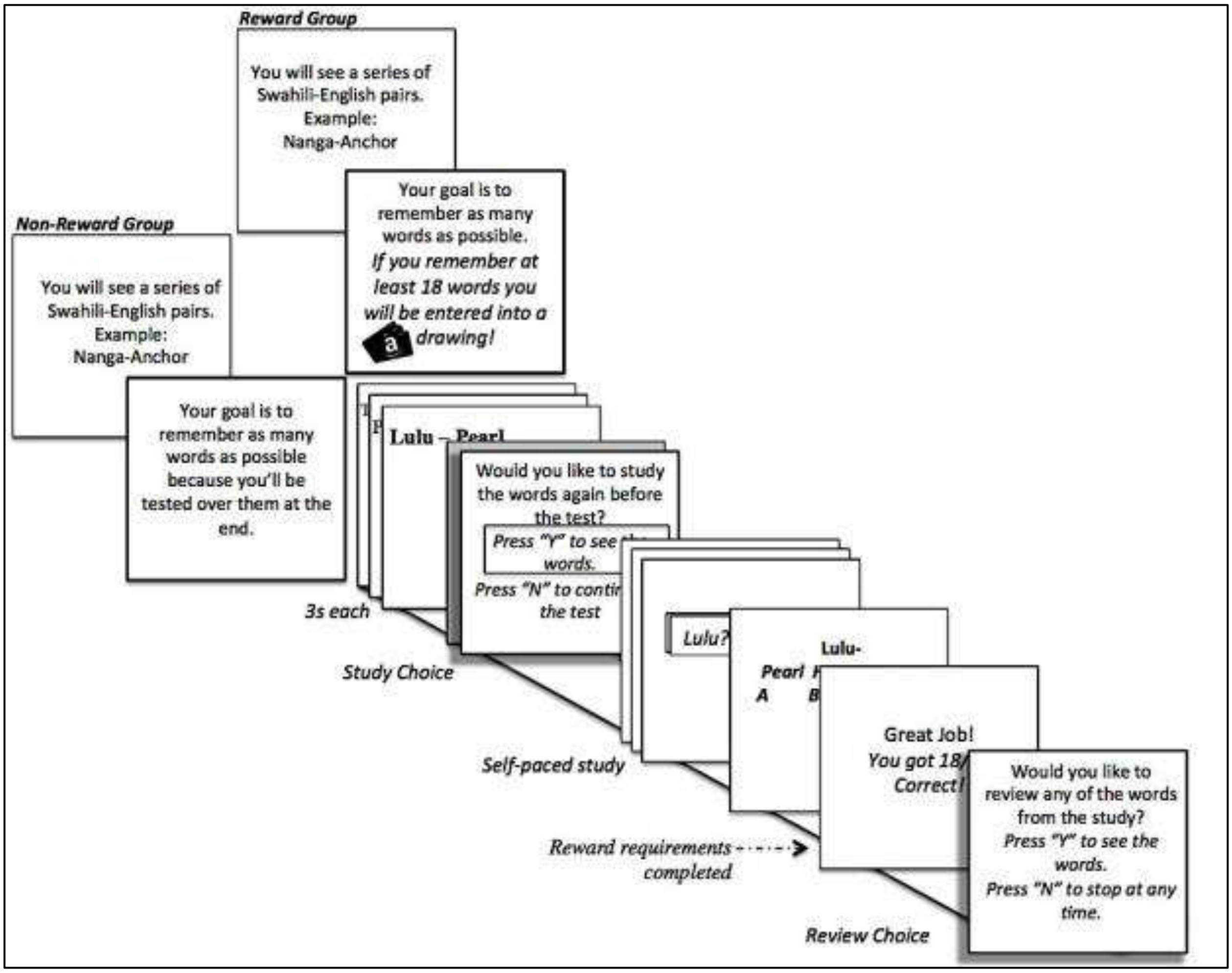


Figure 1. A depiction of the experimental procedure. Subjects in reward group are offered a performance-contingent reward. All subjects are provided a self-paced study choice and review choice.

Results and Discussion

I predicted that study during the study choice (before test) and review choice (after test) phases would differ by group and interact with the presence of the external reward. Specifically, I hypothesized that subjects in the reward group would choose to study during the study choice phase (before test) for a longer duration and more frequently than subjects in the non-reward group. I predicted that in the review choice phase (after test) the reward group would choose to review the material less frequently and for a shorter duration than the non-reward group. This latter pattern is characteristic of the motivational undermining effect.

Both the time that subjects persisted during the study and review choice and the frequency of subjects who made a study or review choice were recorded (see Figure 2 and 3). First, in order to test directly for the presence of an undermining effect, an independent-samples t-test was conducted to compare the average time spent voluntarily engaging in the task when incentives were no longer attainable during the review phase between the reward (n=31, M=5, SD= 22) and non-reward group (n=31,M=16, SD= 25). Reward subjects did choose to review the word pairs for significantly less time than the non-reward subjects (t (1,60)=2.06, p = .02, 1-tailed), indicating that reward subjects did experience a lower motivation to engage in the task during the review phase compared to non-reward subjects.

The mean time for both study choice and review choice periods were compared between groups. A 2 (group: reward x non-reward) x 2 (study time: pre-test x post-test) repeated measures ANOVA was conducted to compare the mean study times for the reward subjects' pre-test study time (M=20, SD= 27) and post-test review time (M=5, SD= 22) to the non-reward group subjects' pre-test study time (M=16, SD= 22) and post-test review time (M=16, SD= 25). There was a significant main effect for the within subjects variable of amount of time spent studying between the pre-test study and post-test review phase (F (1,62)=4.39, $p < .05$) but the interaction was not significant (F (1,60) = 3.52, p = .065). There was not a significant main effect for the between subjects variable of reward on the amount of time spent engaging in the task during the pre- and post-test phases (F (1,61)=.62, p = .42).

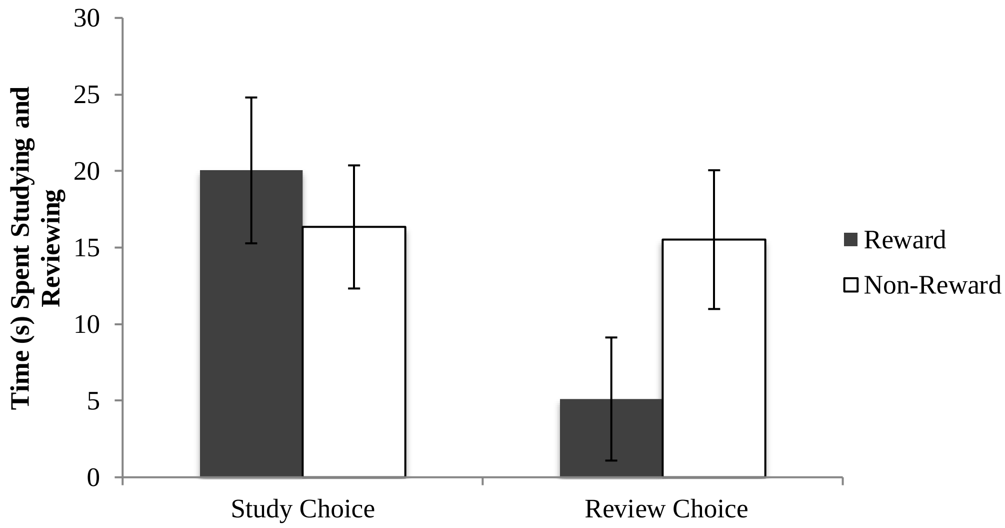


Figure 2. The proportion of subjects in both groups choosing to see the task word pairs before and but not after completing the test (i.e., Study- No Review).

A non-parametric Chi-square test could not be completed in order to test the hypothesis that the frequency of subjects choosing to engage in self-regulated learning would be higher in the reward group during the pre-test study phase but lower during the post-test review phase when compared to the non-reward group. The sample size did not allow for each cell in the test to be greater or equal to 10 (Cochran, 1952; 1954).

Additional analyses were completed to analyze performance differences between groups for accuracy (see Figure 4) and reaction time (see Figure 5) on the multiple-choice test. A 2 (group: reward x non-reward) x 3 (difficulty level: easy x medium x hard) ANOVA was conducted to compare the average accuracy between groups. Response accuracy for the reward group for easy (M=.85, SD= .15), medium (M=.77, SD= .25), and difficult (M=.73, SD= .22) words was compared to the response accuracy for the non-reward group for easy (M=.89, SD= .15), medium (M=.84, SD= .16), and difficult (M=.80, SD= .19). There was a significant main effect for reward group (F (1,30) = 4.85, p < .05) and a significant main effect for difficulty level on accuracy (F(1,30)=4.8, p < .01). A post hoc Tukey test showed that the reward subjects' accuracy for easy items and the non-reward subjects' accuracy for hard items differed significantly at $p < .05$; no other between or within group variable differed significantly on the post hoc tests.

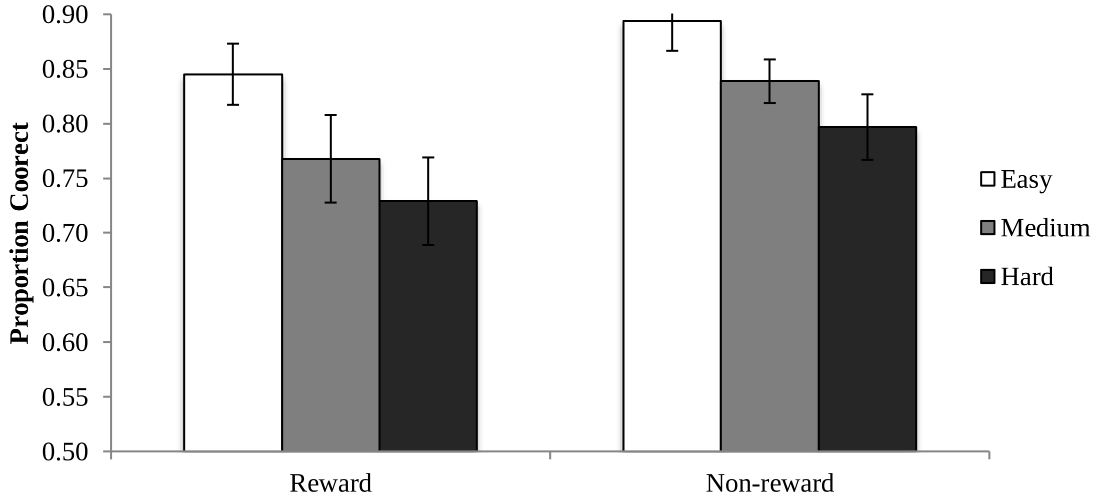


Figure 4. The proportion correction between reward and non-reward group for difficulty of test items.

A 2 (group: reward x non-reward) x 3 (difficulty level: easy x medium x hard) ANOVA was conducted to compare the mean reaction times during the final memory test. The reward group's reaction time, measured in seconds, to easy (*M*=4.1, *SD*= 1.3), medium (*M*=4.7, *SD*= 1.7), and difficult (*M*=5.3, *SD*= 2.1) words was compared to the reaction time for the non-reward group for easy (*M*= 4.0, *SD*= 1.2), medium (*M*=4.7, *SD*= 1.8), and difficult (*M*=5.3, *SD*= 1.8). There was not a significant main effect for reward group ($F(1,30) = .22$, p =.63) but there was a significant main effect for difficulty level on response time ($F(1,30)=9.49$, $p < .001$).

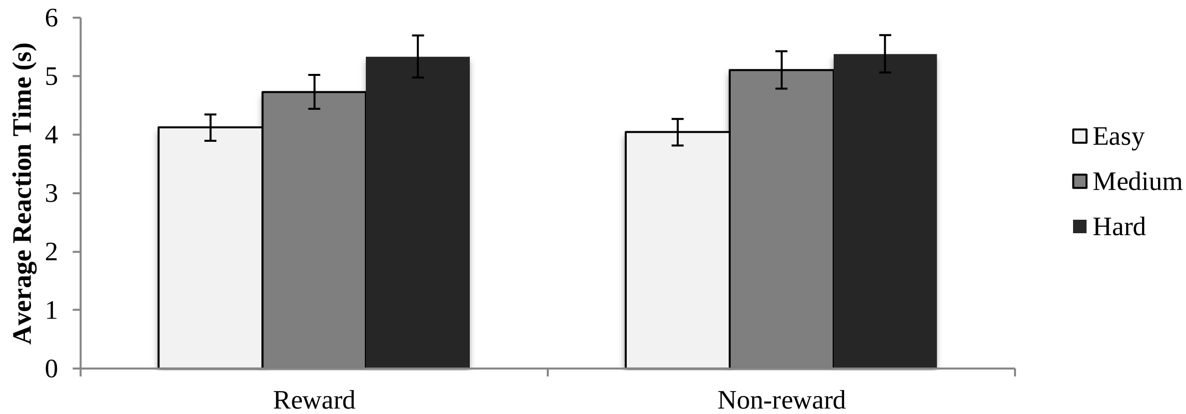


Figure 5. Subjects' average response time for test items based on difficulty level.

Summary: The within subjects effect of reward did not have a significant interaction between time spent studying and reviewing. The time that subjects spent voluntarily choosing to review, once performance-incentives were no longer achievable, was consistent with an undermining effect and supported the hypothesis that reward would cause an undermining effect. The non-reward group appeared to perform slightly better than the non-reward group on the memory test.

CHAPTER 3: STUDY 2 INTRODUCTION

The purpose of Study 2 was to extend Study 1 to include an assessment of long-term memory performance. The results of Study 1 were congruent with past research on the undermining effect and demonstrated that subjects completing a learning task for a performance-contingent reward were less likely to voluntarily spend time studying the words once the reward was removed. The subjects in the non-reward condition were more likely to review the material during the post-test review choice phase. Study 2 examined whether or not this increased study time during the post-test review phase provides a benefit to long-term memory performance. Subjects in Study 2 completed the same task as in Study 1, with the addition of a delayed memory test. The predictions regarding subjects' voluntary choice to study and review material remain the same as in Study 1. Due to the anticipated increased time that non-reward subjects will spend on the task during the review phase, it was further hypothesized that non-reward subjects would remember significantly more words during a delayed free recall test.

Study 2 Method

Subjects

Subjects were recruited from the Psychology Department's research pool at Colorado State University. The pool consists of the same students described in Study 1.

G-Power (Faul, Erdfelder, Lang, & Buchner, 2007) was used to estimate an *a priori* effect size of .3 and reported that a total of 196 subjects were needed but 181 participated in the lab and 158 completed the delayed test. In the laboratory, subjects were randomly assigned to the reward group (n=96) or the non-reward group (n=85). There was no interaction effect ($F(1,240) = .542$, $p = .462$) found for time of study between the two experiments and therefore, subjects from Study 1 (*n*=62) were combined with this analysis to increase the sample size for both the reward (*n*=127) and non-reward groups (*n*=116). The *a priori* effect size was still not met (d= .12 F (1, 241) = .630, η^2 = .003).

Design and Materials

A 2 (group: reward x non-reward) x 2 (test performance: immediate x delayed) mixed design was used. The materials were identical to those used in Study 1 with the addition of online software used for the delayed memory test. Subjects completed the delayed memory test 24 hours after they finished their in-person portion. A link to the test and surveys was emailed to them and their responses were collected using the online survey software, Qualtrics (Qualtrics, 2009). To measure each subject's engagement in the task the Intrinsic Motivation Inventory (IMI) was completed after the delayed memory test (see Appendix A). Four subscales were used to measure how interested subjects were in the task, perceived competency level, how much effort subjects put into the task, and whether or not they felt they had a choice to complete the task. Each subscale included 5-7 questions that are averaged to determine relative score for each factor.

Procedure

The procedure for Study 2 was identical to that of Study 1 with the exception that subjects completed a series of questionnaires and an additional, surprise memory test 24 hours after they left the laboratory. Students were informed when they signed up for the study that they were completing a portion of their research credit in-person and another portion at home. Before the session a research assistant explained that they would complete a word-learning task in the laboratory and additionally complete a series of personality surveys on their own to receive complete credit. Subjects were not informed that there would be a second test completed online.

As in Study 1, subjects were instructed to learn as many words as possible because they would be tested on them after the study session. Subjects in the reward group were additionally informed that if they answered at least 18 items correctly they would be entered into a drawing at the end of the semester for a gift card. Subjects completed the test of the word pairs, shown their score, and then given the opportunity to review the words again.

Unlike Study 1, after the subjects were done reviewing, or if they indicated that they did not want to review the words, a completion screen and a reminder to check their email for more follow-up surveys appeared. Exactly 24 hours after the completion of the in-person task, subjects were emailed a link to a cued-recall test for all 30 Swahili-English word pairs and a series of personality surveys. The subjects completed the test for the words prior to completing the personality questionnaires. For this study I specifically interested in performance differences after a delay. Therefore, a cued recall test was used to provide a more sensitive measurement of memory performance by limiting the number of retrieval cues that subjects were given at test (Yonelinas, 2002). Subjects were shown the Swahili word and given up to two letters of the English target word. The second letter of the target word was given to differentiate words that overlap in their first letter. A second cue letter was provided for the longer word in order to limit the amount of feature information provided to subjects.

Study 2 Results and Discussion

In order to test the hypothesis that reward subjects would experience decreased motivation to engage in self-regulated behavior after the removal of a reward, both the time that subjects viewed words during the study and review choice phases and the frequency of subjects making a study or review choice were recorded. Consistent with the undermining effect, I predicted that the non-reward group would engage in the experimental task during the review choice phase longer than the reward group. Due to the directional prediction of this hypothesis, a 1-tailed independent-samples t-test was conducted comparing the mean review times for the reward (n=127, M=9, SD= 20) and non-reward group (n=116, M=13, SD= 22). The effect of reward on the amount of time spent reviewing did not reach significance (t (1,241)=1.56, $p = .06$, 1-tailed).

I further hypothesized that the amount of time the reward subjects spent performing the self-regulated learning behavior during the post-test review time would decrease significantly compared to the pre-test study phase. A 2 (group: reward vs. non-reward) x 2 (study time: pre-test x post-test) repeated measures ANOVA was conducted to compare the mean study times for the reward subjects' pre-test study choice time (n=127, M=25, SD= 54) and post-test review choice time (n=127, M=9, SD= 17) to the non-reward group subjects' pre-test study time (n=116, M=16, SD= 29) and post-test review time (n=116, M=13, SD= 22) (see Figure 6). There was a significant main effect for the within subjects variable of amount of time spent studying between the pre-test study phase (n=243, M=21, SD= 44) and post-test review phase (n=243, M=11, SD= 20) (F (1,241) = 9.73, $p < .01$). There was not a significant main effect for the between subjects variable of reward on the amount of time spent engaging in the task across both the pre- and post-test phases (F (1)= .63, $p = .44$) but the interaction between reward and time spent on task was significant (F (1,241) = 4.24, $p < .05$). Paired-samples t-tests were conducted to compare the change in average study and review times for the reward and non-reward groups. There was a significant difference between study and review time for the reward group (t (1,126)=3.18, $p < .01$) but not for the non-reward group (t (1,115)=.97, $p = .33$).

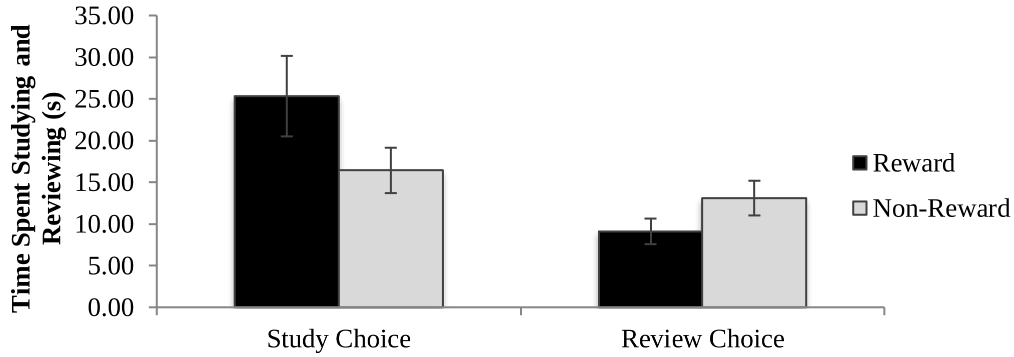


Figure 6. The average amount of time spent in self-regulated behavior during the pre-test study and post-test review phase.

Early investigations of the undermining effect proposed that the difference in subjects' time spent on the task before and after the reward is removed can be used as a comparison of motivational changes between the reward and non-reward subjects (Deci, 1971). Since it is a repeated measure, each subject's time spent reviewing can be compared to her baseline measure of time spent studying. Often, in the free-choice paradigm there is a period before rewards are offered wherein subjects' behavior is observed in order to act as a true baseline; however, in a classroom setting instructors do not often have access to students' prior self-regulated learning behavior. The interest of this exploration is focused on the change of self-regulated learning once incentives are removed. Therefore, the time that subjects spent during the first study phase was subtracted from the time they spent engaging in self-regulated behavior during the second review phase (See Table 1). An independent-samples t-test was conducted to compare the difference in time spent studying and reviewing between the reward (M=-16, SD = 49) and non-reward group (M=-3, SD=57). There was a significant difference between the groups (t (1,241) = 2.05, α_{FW} < .025, 1-tailed), indicating that the subjects in the reward group experienced a larger decrease in time spent reviewing compared to their time spent studying relative to subjects in the non-reward group.

Table 1.
Mean Number of Seconds Spent Studying and Reviewing and Differences Between Times

Group	Study Choice (S)	Review Choice (R)	Review – Study Time
Reward Group	25.3	9.1	-16.2
Non-reward Group	16.4	13.1	-3.3
		Reward (R-S) – Non-reward (R-S)	-12. 9*

Note. As outlined in Deci (1971), the higher the score, the higher the motivation. Asterisks denote significant differences at the $\alpha_{FW} < .025$.

Similar to the above analysis, the within subjects variable of self-regulated behavior allows subjects to be paired on their decision to engage in the task during the study choice phase and their choice again during the review choice phase. Due to this repeated measure, two 2 (Study Choice) x 2 (Review Choice) McNemar tests were completed in order to test the hypothesis that the frequency of subjects choosing to engage in self-regulated learning during the pre-test study choice was higher in the reward group but that these subjects would choose less frequently to re-engage in self-regulated learning behavior during the post-test review phase compared to the non-reward group (Table 2). The number of subjects that voluntarily chose to study or review was compared, irrespective of the amount of time subjects spent viewing the words. Table 2 shows the frequency and proportion of subjects that voluntarily chose to engage in self-regulated learning behavior during both the pre-test study and post-test review phase; only during the pre-test study phase and not the review phase, and those that chose to only engage in either the pre-test or post-test review phases. Initially, the paired test was conducted within each group. There was no statistically significant difference in the proportion of non-reward group subjects' choosing to voluntarily participate in the task during the pre-test or post-test phase (p=1). There was a significant difference between the proportion of reward group subjects choosing to study during the pre-test study phase and the post-test review phase ($p < .05$). As shown in Table 2, the non-reward group had a higher proportion of subjects choose to both study and review. Similarly, the non-reward group also had fewer subjects who initially choose to study and later declined to review the word pairs than such subjects in the reward group.

Table 2.
Paired proportions of subjects during pre-test study and post-test review phases.

			Review Choice Phase							McNemar p
			Did		Did Not		Total			
			n	%	n	%	n	%		
Reward	Study Choice Phase	Did	21	16.5	36	**28.3**	57	44.8		
		Did Not	20	15.7	50	39.3	70	55.1		
		Total	41	32.2	86	67.7	127	100.0	1.0-3.11	.04
Non-reward	Study Choice Phase	Did	21	18.1	23	**19.8**	44	37.9		
		Did Not	23	19.8	49	42.2	72	62.0		
		Total	44	37.9	72	62.0	116	100.0	.56-1.78	1.0

Note. P-values are given for two-tailed results. The discordant cells that would indicate undermined motivation are bolded.

As stated in Study 1, it could be argued that voluntarily choosing to view the word pairs during both the study and review choice phases demonstrates sustained motivation for the task; whereas later choosing not to review may indicate a decrease in task motivation. Therefore, the difference in study-no review choices between reward and non-reward groups is shown in Figure 7. The proportional difference in study-no review choice is 8.5% but the estimated proportions for both the reward and non-reward groups overlap, indicating that the proportional difference is not significant.

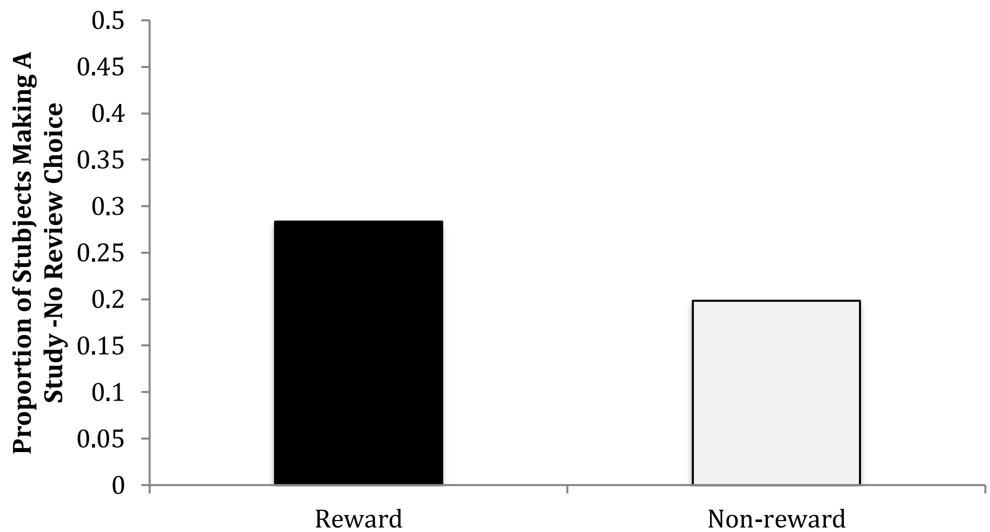


Figure 7. The proportion of subjects in both groups choosing to see the task word pairs before and but not after completing the test (i.e., Study- No Review).

In order to test for differences in memory performance, a 2 (group: reward vs. non-reward group) x 2 (test performance: immediate x delayed) repeated measures ANOVA was conducted to compare the average accuracy between groups on the memory tests (see Figure 8). In order to detect potential outliers, subjects' scores were converted into standard scores for both the immediate and delayed tests. Scores that were -/+ 3 deviations from the mean were removed. One subject was excluded from analysis due to performing substantially lower (z = -3.28) on the immediate test and two were excluded due to performing substantially higher (z = 3.02 - 3.17) on the delayed memory test compared to other subjects. It should also be noted that all subjects did not respond to the test within 24 hours, however, when the analyses were restricted to only subjects completing the test after 24 hours there were no significant differences between groups (F (1, 110) = .82, p = .38) and therefore, all responses are included in the following analyses.

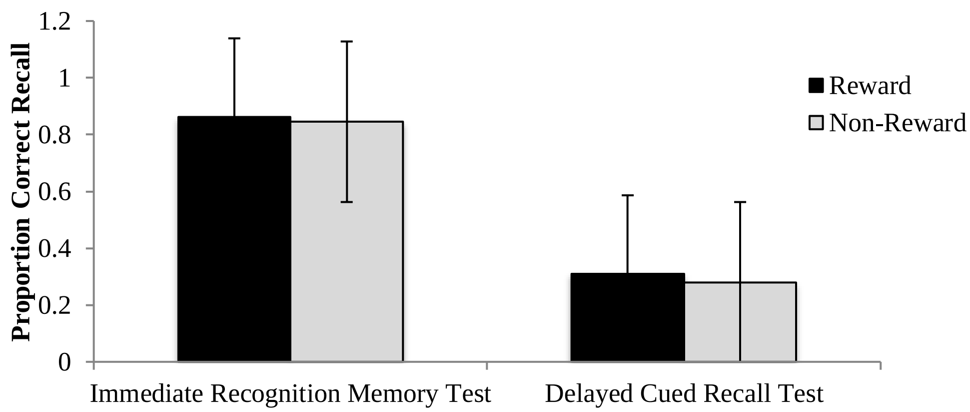


Figure 8. The proportion correct between reward and non-reward groups for immediate and delayed tests.

For the immediate memory test, it was hypothesized that the reward (*n*=85, *M*= .86, *SD*= .16) and non-reward groups (*n*=73, *M* = .85, *SD*= .14) would perform similarly on the immediate memory test; however, for the 24-hour delayed, cued recall test, it was hypothesized that the non-reward group (*n*=73, *M*=.28, *SD*= .2) would perform significantly better compared to the reward group (*n*=85, *M*=.31, *SD*= .3). There was a significant main effect for the within subjects variable of immediate (*M* = .85, *SD* = .15) and delayed (*M* = .29, *SD* = .21) test performance (F (1,156) = 1755, $p < .001$). The interaction between reward group and test was not significant (F (1,156) = .201, $p = .65$) and there was no significant main effect for the between subjects variable of reward on immediate or delayed test performance (F (1,156)= .82, $p = .37$).

Additional analyses were completed to analyze performance differences between groups on accuracy for the immediate and delayed tests. A 2 (group: reward x non-reward) x 2 (test time: immediate vs. delayed) x 3 (difficulty level: easy x medium x hard) repeated measures ANOVA was conducted to compare the average accuracy between groups. Immediate test response accuracy for the reward group for easy (*M*=.88, *SD*= .16), medium (*M*=.87, *SD*= .19), and hard (*M*=.82, *SD*= .22) words was compared to the response accuracy for the non-reward group for easy (*M*=.89, *SD*= .13), medium (*M*=.85, *SD*= .17), and hard (*M*=.78, *SD*= .19).

Delayed test response accuracy for the reward group for easy (*M*= .46, *SD*= .28), medium (*M*=.29, *SD*= .23), and hard (*M*=.17, *SD*= .18) words was compared to the delayed test response accuracy for the non-reward group for easy (*M*=.43, *SD*= .28), medium (*M*=.26, *SD*= .23), and hard (*M*=.14, *SD*= .17). There was a significant main effect for difficulty level (F (1,5)=745.27, p < .001) but no significant main effect for reward group (F (1,5) = .444, p = .82) on response accuracy.

Multiple linear regression analyses were then performed to examine the relationship between time spent voluntarily participating in the task and subjects' immediate and delayed scores, adjusting for reward group. A significant regression equation for study time and reward group was found to predict subjects' immediate tests score (F (2,155) = 9.44, p < .001). The effect of reward was not significantly different than zero (b=.005, *SE*= .024, t= .299, p = .82) but the effect of study time was significantly different from zero (b=.001, *SE* < .001, t= 4.28, p < .001). Subjects' predicted test proportion is equal to .829 + .005**Group* + .001**Study Time,* where group is coded as 1 = reward, 0 = non-reward, and study time is measured in seconds. Subjects' scores increased .1% for each second studied and subjects in the reward group scored .5% higher than subjects in the non-reward group (see Figure 9). This model predicts about 11% of the variance associated with immediate test accuracy (R^2= .109).

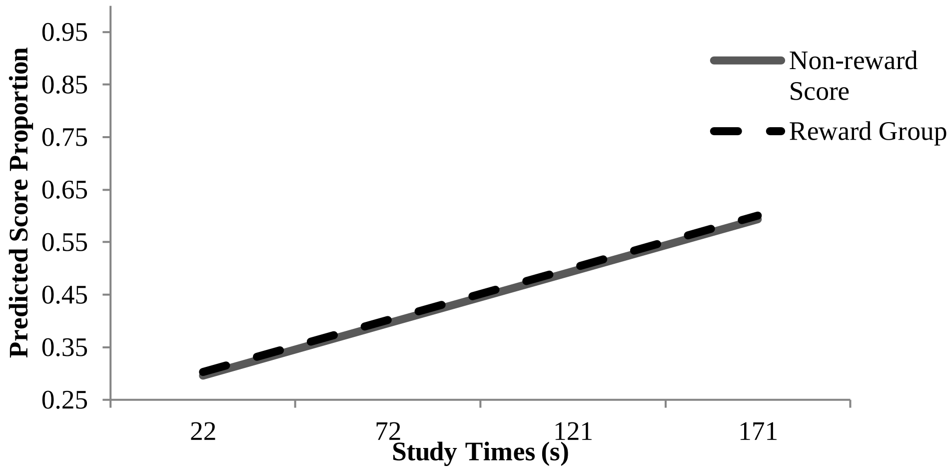


Figure 9. Projected study times and immediate test accuracy. Projected study times are calculated from: the average study time, and 1, 2, or 3 *SD*s above the mean.

In regards to the relationship between time spent voluntarily participating in the task and delayed score, a multiple linear regression was calculated to predict subjects' delayed score based on time spent reviewing and reward group but this did not produce a significant regression equation (F (1,155) = .523, p = .594). A multiple linear regression was then calculated based on study time, review time, and reward, which produced a significant regression equation (F (1,154) = 11.45, p < .001). The effect of reward (b=.01, SE =.031, t= .334, p = .74) and review time (b=.001, SE=.001, t= .334, p = .74) were not were not significantly different than zero but the effect of study time was significantly different from zero (b= .002, SE <.001, t= 5.75, p < .001). Subjects' predicted delayed test proportion is equal to .244 + .01**Group* + .002**Study Time + .001*Review Time,* where group is coded as 1 = reward; 0 = non-reward; and study and review times are measured in seconds. Subjects' scores increased .2% for each second studied, .1% for each second reviewing, and subjects in the reward group scored 1% higher than subjects in the non-reward group (see Figure 9). This model predicts about 18% of the variance associated with immediate test accuracy (R^2= .182).

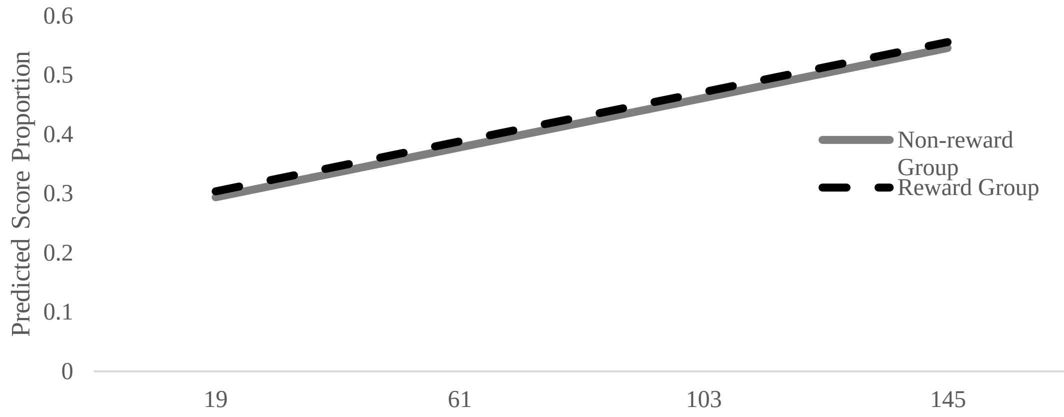


Figure 10. Projected study times and delayed test accuracy. Projected study times are calculated from: no study, the average study time, and 1, 2, or 3 *SD*s above the mean. Review time is adjusted based on subjects' average review time (11.9s).

A 2 (group: reward x non-reward) x 5 (IMI subscale: interest x competence x pressure x value x effort) ANOVA was also conducted in order to find any differences between the Intrinsic Motivation Inventory (IMI) questionnaire between the reward (n = 85) and non-reward groups (n = 73) that was completed online, during the 24-hour delayed portion. The results are displayed in Table 3 and show that there was not a significant main effect of reward on any of the IMI subscales (F (1,160) = 1.27, p = .28). Subjects in the reward group did tend to report higher rates of exerted effort expended on the task and, overall, effort was slightly, positively correlated with time spent studying during the pre-test phase (r (161) = .21, p < .01). Similarly, interest (r (161) = .18, p < .05), choice (r (161) = .18, p < .05), and value (r (161) = .20, p < .05) were positively correlated with study time across all subjects. Although this does not support a difference between groups, it does indicate an increased level of initial task engagement to be related to motivational variables; however, time spent reviewing was not significantly correlated with any IMI variables. It is also important to reiterate that subjects completed this questionnaire at a delay and, therefore, their subjective ratings about the task may have shifted.

Table 3.
Mean scores of the Intrinsic Motivation Inventory (IMI) Questionnaire

	Reward		Non-reward		F-Value
	M	*SD*	*M*	*SD*	
IMI Subscales					
Interest/Enjoyment	4.53	1.38	4.51	1.14	.01
Perceived Competence	4.57	1.38	4.48	1.14	.18
Effort	4.90	1.15	4.59	1.17	2.78
Perceived Choice	4.18	1.24	4.28	1.16	.276
Value	4.60	1.30	4.32	1.14	2.09

Note. F-values are provided for the main effect of reward. Asterisks denote significance at the $p < .05$.

Summary: Two main results were predicted. First, an undermining effect was expected, as was found in Study 1. The non-reward subjects were predicted to remain intrinsically motivated to participate in the task, which was operationally defined as spending a greater amount of time voluntarily reviewing the word pair associates after the immediate memory test. In order to test this, a between-subjects analysis was completed to compare average time spent reviewing between the reward and non-reward group. There did not appear to be a significant difference between the groups' review times. The effect of this contrast was small (d=.19). Using effect size and sample size, G*Power was used to conduct a post hoc power analysis. This revealed that power was 43%. It is possible that the study was too underpowered to detect a significant difference between the groups.

These results appear to support the hypothesis that subjects will increase self-regulated learning behavior when an incentive is present but decrease this behavior when the incentive is no longer present compared to subjects never offered a reward. This was demonstrated by the significant interaction between reward groups and time spent on task and the significant within subjects decrease in time spent studying compared to reviewing in the reward group. The effect for the interaction was small (d=.26) and the observed power was 53.6%. The low power indicates that there may have been too much error and variability in the study to detect a meaningful difference.

The decreased learning behavior after reward removal is not unexpected since previous research demonstrates that rewards will increase activity in pursuit of that reward (Crespi, 1942; Deci, 1971, Eisenberg, Pierce & Cameron, 1994) but the subsequent decreased behavior may contrast the goals of instructors who are attempting to facilitate student engagement and who wish to promote self-regulated learning behaviors independent of the pursuit of external rewards. Study time was significantly related to final test outcomes but ensuring that subjects are engaging in the most beneficial learning strategies overtime might increase the level of retention between the immediate and delayed test.

Second, the non-reward group was predicted to benefit from this additional study time when their memory was assessed on the delayed memory test. Due to the time spent studying, the reward group spent more time overall engaging in the task compared to the non-reward group and the non-reward group did not spend a substantially longer time engaging in review compared to the reward group. Therefore, any motivational benefit that the non-reward group potentially received did not demonstrate a long-term benefits to memory performance. Both the reward and non-reward subjects performed comparably during the immediate and delayed tests, indicating both that there did not appear to be a short-term benefit of performance for reward group subjects and that there was not a long-term benefit for non-reward subjects.

Limitations of this current study include that it may be underpowered to detect meaningful differences. Furthermore, the type of assessment used was not adequate for measuring differences between the groups. For instance, the average subject earned 80% correct on the immediate test indicating that ceiling effects were present for most subjects. The distractor items for the multiple-choice test were not previously studied items and may have allowed subjects to respond accurately by simply recognizing a studied word among non-studied words. It is possible that a more sensitive test involving both studied and non-studied items would allow for greater detection of memory differences and any possible advantages of increased study time. In contrast, it is possible that the delayed test was too difficult. The average subject only answered 29% of the items correctly, indicating that floor effects were most likely present. Similar to the previous limitation, if the follow-up test allowed for a greater range of subject responses an advantage of increased review time may be measurable. The difference in test type was intended to allow for a more sensitive assessment of memory at a delay; however, the difference in tests at immediate and delayed intervals may also be a confounding variable. Study 3 attempted to address these concerns with by adoption a more sensitive recognition test that was used for both the immediate and delayed memory assessments.

Additionally, it is also possible that other measures of motivation may have been more sensitive than choice to study during the free choice interval, which may have been seen by the subjects as a time-pressured laboratory task rather than a learning opportunity. Specifically relevant to the current study, offering subjects the ability to participate in a free trial of Rosetta Stone may be better indicator of intrinsic motivation for language learning. Similarly, measuring participation of reward and non-reward subjects' during a delayed follow-up session could better indicate long-term intrinsic motivation for the task.

CHAPTER 4: STUDY 3 INTRODUCTION

The purpose of Study 3 was to investigate the interaction of choice and reward on subjects' study behavior before and after the removal of performance-based incentives. Studies 1 and 2 tested the impact of reward on study behaviors. As summarized previously, research supports the notion that when external rewards are perceived as regulatory and dictate task participation, rewards can be detrimental to subjects' voluntary choice to participate in a learning task. However, other research has identified mitigating factors that can decrease the harmful effects of external rewards and attenuate the undermining effect. One possible mitigating factor is control over the learning task via subject choice. For example, Muraryama et al. (2013) had subjects complete a reaction time task on a computer with graphics resembling a stopwatch. On some trials, subjects were given a choice of which graphic image, or style, of stopwatch they wanted to use but on other trials the computer chose the watch for subjects. The type of watch was irrelevant to task completion. The researchers found that performance was improved and feedback was more likely to be perceived as positive during the subject-choice compared to the forced-choice trials. Study 3 was similarly designed to test if introducing choice into a learning task increases interest and decreases the detrimental undermining effect of rewards. Half of the subjects in Study 3 were randomly assigned to a self-determined condition in which they chose what to learn (Swahili or Lithuanian words) and the other half were randomly assigned to a forced choice condition in which the language to be learned was selected for them. Within both conditions, subjects were again randomly assigned to a reward or non-reward group.

It was hypothesized that subjects assigned to the forced choice condition would perform similarly to subjects in Study 1. Specifically, reward subjects would demonstrate an undermining effect and choose to spend significantly less time engaging in the task during the post-test review phase. In contrast to previous results, it was further hypothesized that both reward and non-reward subjects assigned to the self-determined content condition would voluntarily choose to view the task material for longer durations than subjects in the forced-choice condition. Of particular interest, the reward subjects in the self-determined condition were expected to voluntarily choose to review the words for a significantly longer duration of time during the review phase compared to reward subjects in the forced choice condition, which would reflect increased intrinsic motivation resulting from self-determined learning.

Study 3 Method

Subjects

Subjects were recruited from the Psychology Department's research pool at Colorado State University. The pool consists of the same students as described previously.

In order to detect the differences between the four conditions of reward and choice for the two within subject measurements of study and review choice, *a priori* effect size of .4 was used in G-Power (Faul, Erdfelder, Lang, & Buchner, 2007) to determine that a minimum of 36 subjects total needed to be enrolled; 64 subjects were actually recruited. Subjects were randomly assigned into the self-determined condition for reward group (n=22) and the non-reward group (n =19) or the forced-choice condition for reward group (n =24) and non-reward group (n =27). The *a priori* effect size was still met (d= .43, F (1, 60) = 2.79, η^2 = .044).

Design and Materials

A 2 (group: reward vs. non-reward) x 2 (choice: self-determined vs. forced-choice task) x 2 (study time: pre-test x post-test) mixed design was conducted. Subjects in the self-determined choice group chose between learning Swahili or Lithuanian words; subjects in the forced-choice group were randomly assigned to language.

The Lithuanian-English study word pairs were selected from a corpus developed by Grimaldi et al. (2010) to complement and extend the paired associates provided by Nelson and Dunlosky (1994). This corpus also included normative ratings for each pair for learning difficulty, likelihood of remembrance, and how closely the Lithuanian words matched their English counterpart. Word pairs were selected from three difficulty levels, low, medium or hard, in order to analyze performance based on difficulty level of stimuli and maintain consistency between stimuli sets. Difficulty level was developed using the same method as described in Study 1.

The study lists consisted of 42 items. 6 items were randomly selected to be presented first and last in the study list. These items were never included in the final test in order to control for primacy and recency effects in memory. Of the remaining 36 studied items, only 24 were included in the initial test. Items from each difficulty level were selected for the study and test items but there was not an equal distribution across all levels due to the limited numbers of stimuli available for each of the normed lists.

Subjects completed an associate recognition test that consisted of 36 items. The test lists were comprised of studied and non-studied items. There were 12 studied items that were intact pairs, 12 studied items that were conjunction pairs, and 12 non-studied, new word pairs. Both the intact and conjunction pairs consisted of 4 easy, medium, and hard word pairs but the new pairs consisted of entirely all-medium difficulty word pairs. The intact pairs were the same items that subjects studied during the initial study phase. The conjunction pairs consisted of a previously studied Swahili or Lithuanian word randomly re-paired with an incorrect, but previously studied, English translation. The only contingency for the re-pairing was that the word pairs were rearranged within their respective difficulty level (e.g., English translations from an easy difficulty level were not paired with a foreign word from a medium difficulty level). The new word pairs were never studied items that were randomly selected from the corpus of normed word pairs for the respective languages. During the test phase, subjects responded “S” to items that were the same, “D” to different, conjunctive pairs, and “N” to new, non-studied word pairs. The study and test lists were originally counterbalanced into 8 different tests in order to allow for each word pair to appear as intact, conjunction, or new; however, 3 word pairs were duplicated across 3 of the study and test lists which required them to be dropped from the final memory test analyses. For example, the proportion of items correctly identified for list 3 was determined out of 30 possible remaining questions and for list 5 out of 33. All other test versions did not have errors that interfered with testing. Due to the replication error, after it was detected a single randomized list was created for the remaining subjects. Subjects 1 – 48 were assigned to 1 of the 8 original lists and subjects 49 – 64 were assigned to the randomly generated study and test list.

The Intrinsic Motivation Inventory (IMI) described in Study 2 was used again as a manipulation check and to measure each subjects’ reported engagement in the task (see Appendix B). In addition to the inventory, specific questions pertaining to the choice of the stimuli were included. Subjects were asked if they had a preference of which language they selected/was selected for them and whether or not they were satisfied with that decision. This was completed as the last portion of the study and was followed by a debriefing of the experiment.

Procedure

The procedure for Study 3 was similar to that described in Study 1 with the exception of task choice and test type. Subjects in the self-determined choice condition chose the language of words to study whereas subjects in the forced-choice condition were randomly assigned to study Lithuanian-English or Swahili-English word pairs. Before the selection, subjects were shown a map of where the languages are typically spoken and pictures depicting the culture in order to facilitate interest in the task and make choices more personally desirable (see Figure 11).

Both groups were exposed to one study block of the word pairs. All 30 of the word pairs were presented once in a random order. Each word was presented for 3 seconds. Before the session they were instructed to learn as many words as possible because they would be tested on them after the study session. Half of the subjects in each condition were randomly assigned to the reward group and were additionally informed that if they answered at least 18 items correctly they would be entered into a drawing at the end of the semester for a gift card.

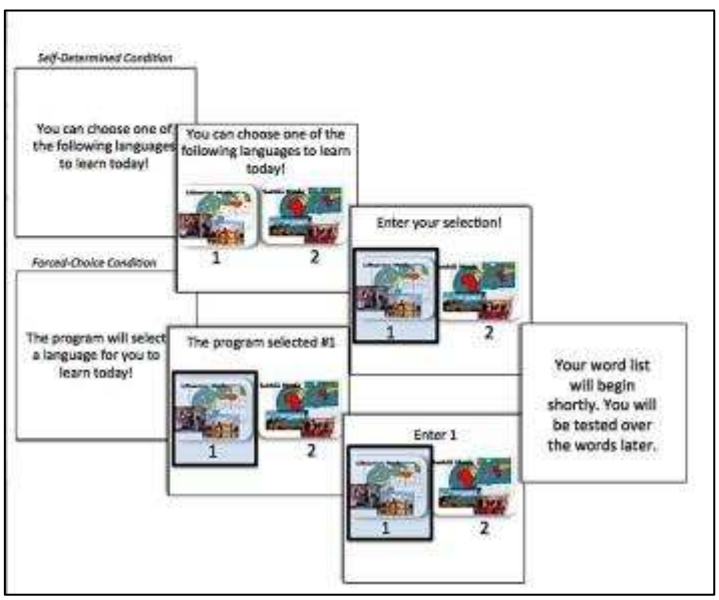


Figure 11. A depiction of the experimental procedure. Subjects in the self-determined condition can select between an Lithuanian or Swahili word list and the computer selects the word list for subjects in the forced-choice condition.

Study 3 Results and Discussion

I hypothesized that the amount of time the reward subjects spent viewing the words during post-test review choice would decrease significantly compared to the pre-test study choice only in the forced-choice condition, replicating an undermining effect in the forced choice condition. The reward subjects assigned to the self-determined choice condition were hypothesized to spend significantly more time engaging in self-regulated learning behavior in the post-test review phase compared to the forced choice condition. A 2 (group: reward x non-reward) x 2 (choice: self-determined x forced-choice) x 2 (study time: pre-test x post-test) x 2 (language: Swahili x Lithuanian) repeated measures ANOVA was conducted in order to compare the mean participation times for each group. The main effect of language on subjects being assigned to, or choosing, either Swahili (n = 33) or Lithuanian (n = 31) words did not have a significant effect on the dependent variables (F (1,56) = 1.52, p = .22) and was removed from the model to focus on the main variables of interest. Within the self-determined condition, reward subjects' pre-test study (n=14, M=36, SD= 54) and post-test review time (n=14, M=28, SD= 34) was compared to the non-reward group subjects' pre-test study (n=13, M=79, SD= 70) and post-test review time (n=13, M=42, SD= 57); see Figure 12. Within the forced-choice condition, reward subjects' pre-test study (n=19, M=70, SD=80) and post-test review time (n=19, M=22, SD=36) was compared to the non-reward group subjects' pre-test study (n=18, M=40, SD=61) and post-test review time (n=18, M=37, SD=34).

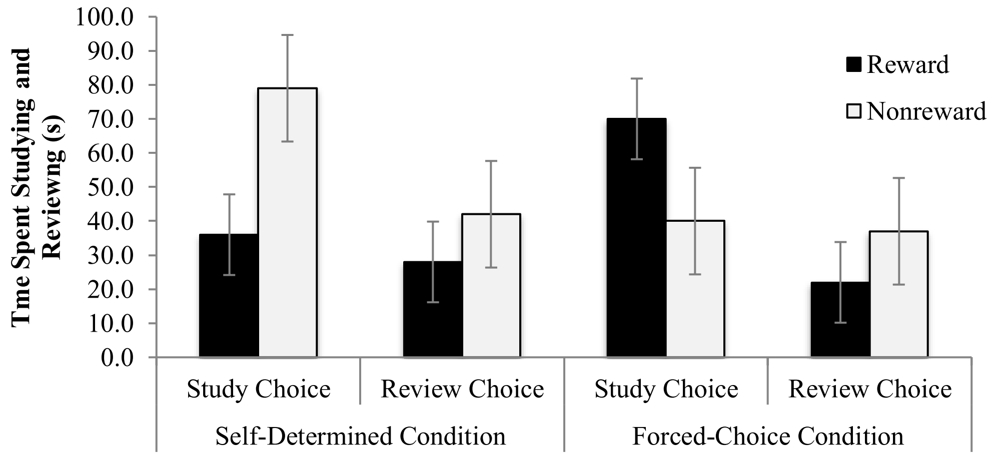


Figure 12. The average amount of time spent voluntarily performing the task in the study and review choice phase.

There was a significant main effect for the within subjects variable of amount of time spent studying between the pre-test study phase (*n*=64, *M*=56, *SD*= 68) and post-test review phase (*n*=64, *M*=32, *SD*=40) (F (1,60) = 6.95, $p < .05$). There was no main effect of choice condition (F (1,60) = .140, $p = .71$) on the amount of time that subjects in the self-determined condition spent studying during the pre-test study (*n*=27, *M*=57, *SD*= 65) and post-test review phase (*n*=27, *M*=35, *SD*= 46) and time that subjects in the forced-choice condition spent studying during the pre-test study (*n*=37, *M*=55, *SD*= 72) and post-test review phase (*n*=37, *M*=29, *SD*= 35) (see Figure 13).

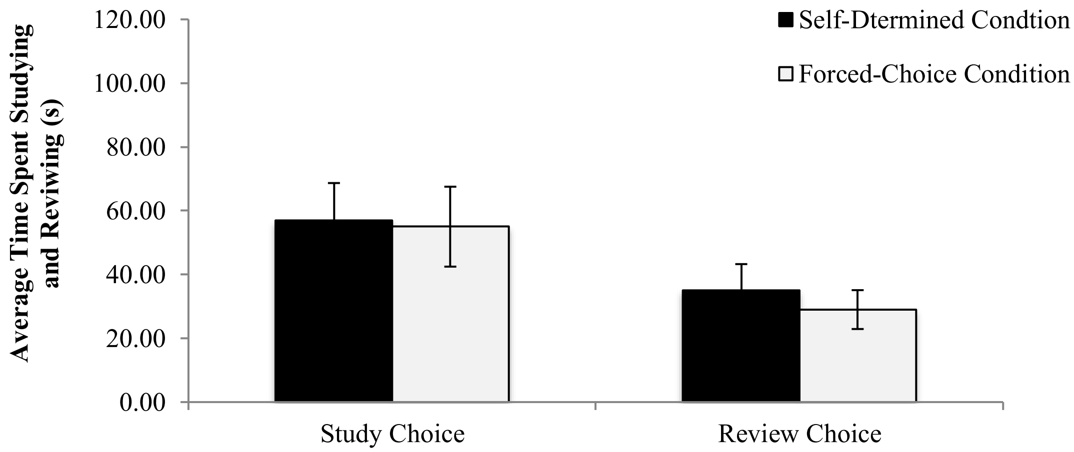


Figure 13. The average amount of time spent voluntarily performing the task in the study and review choice phase between choice condition.

There was not a main effect of reward group (F (1,60) = .920, p = .34) on the amount of time that subjects in the reward group spent studying during the pre-test study (n=33, M=55, SD= 71) and post-test review phase (n=33, M=25, SD= 34) and subjects in the non-reward group spent studying during the pre-test study (n=31, M=56, SD= 67) and post-test review phase (n=31, M=39, SD= 45).

The interaction between reward and time spent on task was not significant (F (1,60) = .150, p = .70). There was not a significant interaction for the between subjects variable of self-determined or forced-choice condition on the amount of time spent engaging in the task during the pre- and post-test phases (F (1,60)= .025, p = .87) and the simple main effect of reward and choice condition was not significant (F (1,60)= 2.79, p = .10). The double interaction between reward and choice condition was significant (F (1,60)=4.11, $p < .05$). The interaction was most likely driven by the increased study activity by the non-reward subjects in the self-determined condition and reward subjects in the forced choice condition. Although it was not significant, the difference between time spent studying for the reward (n = 14, M = 36, SD = 54) and non-reward subjects (n = 13, M = 79, SD = 70) in the self-determined condition was 43 seconds (F(1,25) =

3.25, p = .08). The difference between time spent studying for the reward (n = 19, M = 70, SD = 80) and non-reward subjects (n = 18, M = 40, SD = 60) in the self-determined condition was 30 seconds (F(1,35) = 1.61, p = .21).

As in Study 2, subjects' time spent engaging in the task during the pre-test study phase can be compared to the time they spent engaging in the task during the review phase. The difference between the time spent reviewing and studying for reward and non-reward subjects was compared separately for the self-determined and forced choice conditions in order to explore whether the reward and non-reward subjects performed similarly to those in Study 2 on this measure. A 2 (group: reward x non-reward) x 2 (condition: self-determined x forced choice) factorial ANOVA and planned, post-hoc independent samples t-tests were conducted to compare the difference in time spent studying and time spent reviewing between self-determined and forced choice condition for reward and non-reward subjects.

The main effect of choice (F (1,60) = .25, p = .87) was not significant, the main effect of reward was not significant (F (1,60) = 1.50, p = .70), but there was an interaction between the reward and choice condition (F (1,60) = 4.11, p < .05). The planned post hoc comparisons were conducted in order to explore the interaction effect (See Table 4). There was not a significant difference between the reward and non-reward groups (t (1,62) = 1.46, p = .07,1-tailed α_{FW} > .016) and there was not a significant difference within the self-determined choice condition (t (1,31) = -.43, p = .33,1-tailed α_{FW} > .016) or the forced choice conditions between the reward and non-reward groups (t (1,29) = -.22, p = .41,1-tailed α_{FW} > .016), indicating that the subjects in the reward group did not spend a significantly different amount of time reviewing compared to time studying relative to subjects in the non-reward group.

Table 4.
Mean Number of Seconds Spent Studying and Reviewing Within Subjects

		Study Choice (S)	Review Choice (R)	Review – Study Time
Self-Determined Condition (SDC)	Reward	35.90	28.25	-7.65
	Non-reward	79.12	41.52	-37.60
Forced-Choice Condition (FCC)	Reward	69.78	22.23	-47.54
	Non-reward	40.10	36.63	-3.46
		Reward (R-S) – Non-reward (R-S)		-12.84
		SDC Reward (R-S) – SDC Non-reward (R-S)		-29.95
		FCC Reward (R-S) – FCC Non-reward (R-S)		-44.08

Note. The higher the score, the higher the motivation. Significance is provided for 1-tailed tests and asterisks denote significant differences at $\alpha_{FW} < .016$.

Additionally, the within subjects variable of study and review behavior allows subjects to be paired on their decision to engage in the task during the study choice phase and their choice again during the review phase; however, unlike in Study 2, a series of 2 (Study Choice) x 2 (Review Choice) McNemar tests were not completed because the sample size did not meet the necessary minimum requirements to conduct a chi-square test (Cochran 1952, 1954).

A 2 (group: reward x non-reward) x 2 (choice: self-determined x forced-choice) repeated measures ANOVA was conducted to compare the average accuracy between conditions on the memory test (see Figure 15). It was hypothesized that subjects in the self-determined content condition would perform significantly better than those in the forced-choice content condition. There was not a significant difference ($F(1,6)=.971$ $p = .32$) for the performance of self-determined choice subjects ($n = 27$, M= .66, SD= .16) and forced choice subjects (n = 37, M= .63, SD= .17), the effect was non-significant (F (1,60) = .67, $p = .16$). There was not a significant main effect of reward ($F(1,6)=.971$ $p = .32$) nor was there an interaction ($p = .16$).

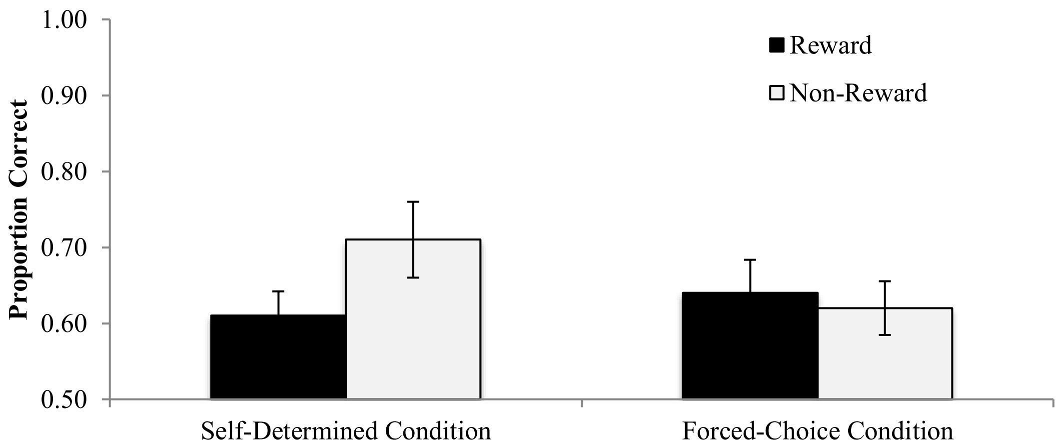


Figure 15. Accuracy in self-determined and forced choice content conditions.

The proportion of subjects responding to the test stimuli as same, or intact, is displayed in Table 6. The purpose of the associative recognition test was to detect if the reward and choice conditions facilitated different criteria for subject responses. It was chosen to allow for analyses between studied items to detect how well subjects in each condition connected the foreign word to the correct translation and between recognition of studied and non-studied items. In other words, the subjects' ability to discriminate between correct and incorrect studied items and non-studied items can be measured.

Table 6.
Proportion of Test Stimuli Called Intact by Choice Condition and Reward Group

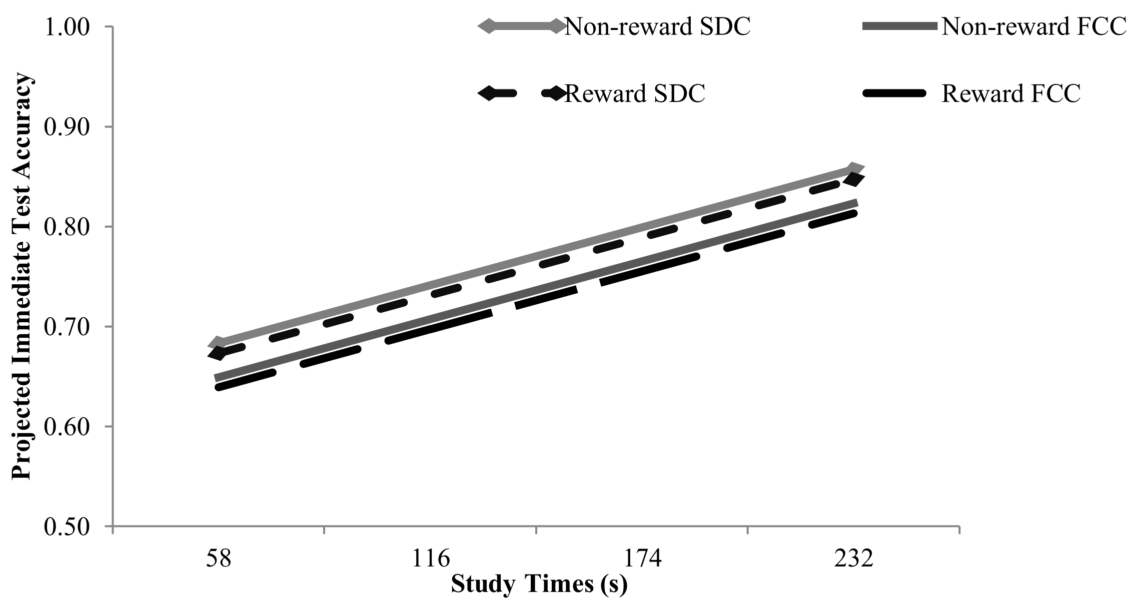
		Intact		Conjunction		New		d'	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Self-Determined	Reward	.79	.19	.40	.22	.18	.18	1.20	.69
	Non-reward	.76	.17	.24	.19	.09	.11	1.62	1.02
Forced-Choice	Reward	.75	.23	.30	.25	.19	.22	1.44	1.24
	Non-reward	.76	.21	.42	.17	.13	.16	1.00	.94

Note. d' is shown for intact and conjunction pairs.

A 2 (group: reward x non-reward) x 2 (condition: self-determined x forced choice) x 2 (item type: intact x conjunction) factorial ANOVA was conducted in order to determine if there was a difference in response tendencies between the groups. There was no main effect for subjects in either reward group (F (2,59) = .09, p = .91) or choice condition (F (2,59) = .37, p = .69) on discrimination between identification of intact and rearranged test stimuli. There was an interaction between reward group and choice condition (F (2,59) =3.57, p < .05). Between subjects comparisons showed that the interaction stemmed from differences in response to conjunction, or rearranged, pairs. Post-hoc comparisons further showed that non-reward subjects in the self-determined conditioned (n=13, M=.24, SD=.19) were less likely to incorrectly identify conjunction pairs as intact pairs than non-reward subjects in the forced choice condition (n=18, M=.42, SD=.17) (F (1,29) =7.59, p = .01, α_{FW} < .025). This trend was reversed in the forced choice condition. Reward subjects in the forced choice condition appeared to be less likely to incorrectly identify conjunction pairs as intact but the trend was not significant (F (1,31) =1.79, p = .19, α_{FW} > .025).

The ability for subjects to distinguish between previously studied pairs was further analyzed by calculating subjects' discriminability index (d') for the intact and conjunction pairs. The discriminability index did not significantly differ based on reward group (F (1,60) = .002, p = .96), choice condition (F (1,60) = .577, p = .45), or an interaction between the group and condition (F (1,60) = 2.77, p = .10), indicating that overall sensitivity between subjects' choices did not differ by reward or choice condition.

As in Study 2, a multiple linear regression analysis was performed to examine the relationship between time spent studying and subjects' immediate test scores, adjusting for reward group and choice condition. The results support a significant relationship between the predicting variables and test score (F (3, 88)= 2.87, p < .05). The effects of reward group (b=-.031, SE =.03, t= -.785, p = .43) and choice condition (b=.031, SE =.04,t= .78, p = .44) were not significantly different from zero but the effect of study time was significant (*b*=.001, *SE* <.001, *t*= 2.70, *p* < .01). The regression model (ŷ = .572 - .031**Reward* + .031*S*elf-determined Choice* + .001**Study Time)* predicts a significant relationship between study time and immediate test proportion and it is predicted that non-rewarded subjects will receive a **.**10% increase in final accuracy for every second spent studying. Receiving a reward incentive is predicted to result in a 3.1% decrease but subjects offered choice in which language to study are expected to receive a 3.1% increase in accuracy (see Figure 16). This model predicts about 13% of the variance



associated with immediate test accuracy (R^2= .126).

Figure 16. Projected study times and immediate test accuracy for reward groups and self-determined (SDC) or forced-choice condition (FCC). Projected study times are calculated from: no study, the average study time, and 1, 2, and 3 *SD*s above the mean.

A 2 (group: reward x non-reward) x 2 (choice: self-determined x forced) x 5 (IMI subscale: interest x competence x pressure x value x effort) factorial ANOVA was conducted to compare Intrinsic Motivational Inventory scores between groups (Table 7). As shown in the table, there were no significant interactions on the dependent variables that were meant to measure motivational variations between the groups. Unlike Study 2, subjects immediately completed the IMI. The results are similar to those in Study 2 and suggest that the IMI results are consistent when taken immediately following or after a delay. This may indicate that the motivational variables are not enough to influence subjective ratings of task motivation or that the scale and paradigm are not converging on the same measures (Peirce et al., 2003)

Additionally, a 2 (group: reward x non-reward) x 2 (choice: self-determined x forced) x 2 (choice questions: preference x satisfaction) ANOVA was conducted to compare preferences between all conditions. These questions were intended to test if the subjects in the self-determined choice condition would report feeling a higher initial preference for the material and a higher rate of final satisfaction. Subjects in the self-determined condition (M = 4.33, SD = 2.20) reported a higher preference for a possible language compared to subjects in the forced choice condition (M = 2.67, SD = 1.88) and the interaction between reward group and choice condition was significant (F (1,60) = 10.07, $p < .01$).

The main effect of choice condition did not appear to significantly impact satisfaction for the language learned (F (1,60) = .13 p = .71). There were no significant interactions between reward group and choice for on satisfaction and there was no significant main effect of reward on subjects' response to their preference or satisfaction of language choice (F (1,60) = .01, p = .60). The interaction between reward and choice on satisfaction was also not significant (F (1,60) = .44, p = .64). Reward subjects in the self-determined condition reported being more satisfied than their non-reward counterparts but reward subjects in the forced choice condition reported being less satisfied than the non-reward subjects in the same condition. Increased preference in the self-determined condition indicates that the manipulation of choice may have increased initial task engagement, encouraging subject participation.

Table 7.
Mean Scores for the Intrinsic Motivational Inventory (IMI) and Language Preference Questions for All Groups

	Self-Determined Condition				Forced-Choice Condition				F-Value
	Reward		Non-reward		Reward		Non-reward		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
IMI Subscales									
Interest/Enjoyment	4.00	.96	4.17	1.08	4.15	.85	4.29	.83	.002
Perceived Competence	4.22	1.01	4.47	1.09	4.41	1.16	4.61	.96	.008
Effort	4.20	1.38	4.52	1.03	4.32	1.11	4.84	1.19	.106
Perceived Choice	4.55	.78	4.44	.61	4.41	.56	4.54	.57	.556
Value	4.67	1.01	4.67	.83	4.72	.95	4.90	1.00	.139
Choice of Language Questions									
I had a preference of language	4.6	3.34	4.00	2.08	2.63	1.86	2.77	1.96	.501
I was happy with my language	5.50	.94	5.00	1.29	5.42	1.46	5.33	1.64	.342

Note. F-values are provided for the interaction between reward and choice condition. Asterisks denote significance at the $p < .05$.

Summary: Three main results were predicted. First, a main effect for the choice condition was expected for time spent voluntarily engaging in the task. Subjects in the self-determined choice condition were expected to show enhanced motivation relative to the forced-choice condition, which was operationally defined as spending a greater amount of time voluntarily viewing the word pair associates during both the study and review choice. On average, subjects in the self-determined condition (*n*=27, *M*=57, *SD*= 65) and subjects in the forced choice condition (*n*=37, *M*=55 *SD*= 72) studied for nearly the same amount of time. Similarly, review time for subjects in the self-determined condition (*n*=27, *M*=35, *SD*= 46) resembled subjects in the forced choice condition (*n*=37, *M*=29 *SD*= 35), providing no substantial support for the possible benefit of increased autonomy on motivation. The effect size for reward was found to be small (*d*=.24) and the effect size of choice condition was found to be inconsequential (*d*=.09). The post hoc observed power was 15% for the effect of reward and

6.6% for the effect of choice condition. It is possible that this study was too underpowered to detect a difference between the variables.

Second, choice condition and reward were expected to interact during the review phase, such that the reward group would show less undermining in the self-determined choice condition than in the forced-choice condition. There was an interaction between reward group and choice condition found and there was a substantial difference to provide directional support to the hypothesis; however, there was still was not a significant decline in between the time spent reviewing in the forced choice condition for reward and non-reward groups to demonstrate an undermining effect. As with the above variables, the interaction had only a small effect size (d=.43) and the post hoc observed power for the interaction was 37%.

Lastly, the increased time spent on the task was expected to result in superior performance in the self-determined choice condition for an immediate memory test. There was no difference found between the reward group or choice conditions test performance. Reward and non-reward subjects in both conditions also appeared to use similar levels of discriminability when identifying intact and rearranged previously studied word pairs but there was a significant difference in errors committed between self-determined and forced choice condition non-reward subjects when responding to rearranged pairs. This could potentially indicate that the choice manipulation altered subjects' approach to the task. Harackiewicz (1979) found that subjects offered a performance or task-contingent reward for completing puzzles remembered less incidental, or irrelevant, task information (e.g., the title of the puzzles) compared to their non-rewarded peers. She argued that it was possible that reward subjects were focused more narrowly on the purpose of the task. It's possible that the forced choice condition interacted with caused a similar impact on the non-reward subjects but it is difficult to determine if non-reward subjects in the forced choice condition approached studying or testing differently than non-reward subjects in the self-determined condition. However, accuracy and discriminability rates between the studied items was not significant, indicating that there is little evidence that the experimental manipulations altered subjects' overall approach to the task or criteria for responding to test items.

Study 3 was intended to extend previous research by demonstrating limiting factors of the undermining effect for a learning task and to support other research demonstrating that incorporating choice into tasks combats the detrimental influence external rewards can have on task performance (Murayama et al., 2014). There does not appear to be strong enough results from this study to draw final conclusions concerning increased autonomy and learning behaviors.

CHAPTER 5: GENERAL DISCUSSION

This study was an exploration and extension of the undermining literature to educationally relevant research. It was aimed at expanding previous literature that demonstrated that student-learning behaviors were susceptible to the undermining effect (Wehe et al., 2015). Previous research on the undermining effect has broadly shown that rewards decrease intrinsic motivation to perform a task. Overall, the undermining effect was replicated in Study 1. Specifically, students offered a reward for better performance chose to review information less during their free time than students not offered a reward. Additionally, rewards tended to increase study time before an upcoming test. Study 2 was designed to explore whether different motivational states would alter student behavior and, consequently, learning outcomes. The trend of Study 2 was in the direction of the hypotheses; there was support to indicate that subjects offered an incentive studied longer for an upcoming test but experienced a decline in motivation to continue engaging in the task once the incentive was no longer achievable. Initially, it was hypothesized that sustained motivation to continue participating in the task would provide the greatest benefit to learning but the current results indicated the effort spent studying and engaging in the task prior to a test was a better predictor of long-term memory performance.

Study 3 examined whether increased self-choice would mitigate undermining. Subjects were assigned to either a self-determined choice condition in which they were allowed to choose the language to be learned or a forced choice condition wherein the computer chose the language. Reward and choice did significantly interact with time subjects voluntarily spent on task; however, this interaction appears to be attributable to time subjects spent engaging in the task during the pre-test study phase. There was not sufficient support to indicate that an undermining effect had occurred. Subjects assigned to a self-determined choice condition reported having a higher preference for the material they were learning.

Further Considerations

The main emphasis for the current experiments was to examine the effects of motivation on subjects' self-regulated behavior and subsequent learning. There is little support that motivation, specifically monetary incentives, directly influences learning (Kang & Pashler, 2014; Nilsson, 1987) however motivation is widely thought to indirectly improve learning and is factored into academic performance. For example, Steinmayr and Spinath (2009) found that motivational variables (e.g., achievement goals, perceived task value, etc.) were valuable when predicting 11^{th} and 12^{th} grade high school students' academic success as measured by their yearly grade point average. The researchers controlled for students' prior academic success, which explained the most variance in their general performance (73.8%). When prior success was controlled, motivational variables such as goals for high or low achievement (7.6%) or learning and performance goals (4.0%) continued to explain variance. These contributions are arguably more substantial because they were comparable to the amount explained by intelligence in their model (7.9%). Duckworth and Seligman (2005) further found that 8^{th} grade students' self-discipline (i.e., impulsivity measured by subjective, teacher, and parental surveys) better predicted general academic performance relative to measures of intelligence and those students high in self-discipline performed better than their low self-disciplined peers. Students' beliefs about their ability and learning potential should be factored into their likeliness to be better-disciplined students. Blackwell, Trzesniewski, and Dweck (2007) demonstrated that students' possessing a belief that intelligence was fluid and developable, not innate or fixed, performed better when exposed to increasingly challenging math curriculum.

Again, this current study examined how reward and externally motivated states may impact students' spontaneous, self-regulated behaviors (i.e., not prompted by an instructor). In regards to students' independent learning behaviors, Dunlosky et al. (2013) highlight many common techniques that students can apply to their studying in order to improve learning outcomes. Several of these techniques (e.g., elaborating material and self-testing) require students to put more effort into expanding and cognitively processing material (Dunlosky et al., 2013; Mosovitch & Craik, 1976). Dunlosky et al. (2013) further highlights that student characteristics are important to consider when promoting different techniques, including state-based characteristics such as motivation. From the current set of studies it is apparent that rewards interact with subjects' initiation and discontinuation of self-regulated learning behaviors but more research is needed to understand how motivation influences subjects' effort and cognitive strategies.

For example, previous research has indicated that motivational states interact with students' likelihood to engage in complex cognitive strategies (Stolk & Harari, 2014). Additionally, subjects' self-imposed and task-imposed goals have been shown to interact with self-regulation strategy (Thiede & Dunlosky, 1999; Ariel et al., 2009). The popular discrepancy reduction model for self-regulated learning strategies posits that students attempt to reduce the level of unknown information by primarily focusing on information that is harder to learn; however, research has demonstrated that task goals, or agendas, can lead to the adoption of different strategies. Thiede and Dunlosky demonstrated that subjects attempted to maximize efficiency when attempting to reach a set criterion. Instead of focusing on reducing the discrepancy in knowledge, subjects focused on learning the least number of words necessary to reach a goal. In this way, agendas decrease the likelihood that subjects depend on difficulty level to determine learning strategy

(Ariel et al., 2009;Thiede & Dunlosky, 1999). The interaction between learning strategy and reward structure is important to consider. The results from Study 2 do indicate that reward interacts with subjects' decisions to begin or end self-regulated learning behaviors but more research needs to be conducted to measure differences in approach to tasks and willingness to apply more effortful strategies.

Study 1 and 2 tested the impact of a performance-contingent reward on subjects' voluntary behavior but there is evidence that not all rewards or schedules of rewards are harmful to motivation (Eisenberger, Pierce & Cameron, 1999). Study 3 tested the impact of choice in order to examine if providing subjects with increased autonomy lessened the impact of a performance-contingent reward. Within Deci and colleagues' cognitive evaluation framework of intrinsic motivation, it is posited that if rewards are not perceived as controlling and provide information concerning how to improve, or master, a situation rewards can facilitate and enhance interest and motivation for a task (Deci et al., 1999). It was posited that creating a situation where subjects felt increased choice, they would view the reward as less controlling.

Harackiewicz, Manderlink, and Sansome (1984) demonstrate how the context and structure of reward is important to consider. The researchers utilized a free-choice paradigm in order to dissociate the impact of rewards and performance evaluations on intrinsic motivation. Their main operational definition of intrinsic motivation was time that subjects spent voluntarily performing the experimental task of pinball. Subjects were divided into performance-contingent reward, performance evaluation only, or evaluation and unanticipated reward groups. They found that subjects expecting only an evaluation of performance experienced motivational undermining relative to a group that received a reward for their performance. Harackiewicz et al. (1984) further found that when subjects received an unanticipated reward based on their performance, their intrinsic motivation was increased relative to peers that anticipated the reward contingent on a performance evaluation. In regards to task performance, a performance-contingent reward was found to improve performance for one group of subjects but all reward groups consistently reported that they were more concerned about performance compared to the control group that did not anticipate a reward.

Additionally, Peirce et al. (2003) demonstrated that performance-contingent rewards bolster motivation when rewards are distributed for completing increasingly difficult tasks. The researchers required subjects to complete a series of visual recognition puzzles; half received a reward for accurate completion of each puzzle and half were not given an incentive. Half of the subjects in each group were assigned to either a constant-challenge group and completed the same number of puzzles per session. Researchers assigned the other half to a progressive-challenge group and subjects completed an increasing number of puzzles per session. Peirce et al. measured the amount of time that subjects in each group spent voluntarily choosing to complete more puzzles during a free-choice period. The results demonstrated the reward group assigned to a progressive-challenge task spent significantly more of their free time engaging in the puzzle compared to the other groups. The researchers reported that subjects in the progressive-challenge group did not describe feeling controlled or pressured to complete the task. Therefore, it is possible that if rewards do not induce feelings of pressure or negative emotional responses, they can be beneficial to enhancing intrinsic motivation.

These results manipulating reward structure (Harackiewicz et al.,1984; Pierce et al.,1993) support the results described in Study 3 arguing for the importance of autonomy and choice in facilitating intrinsic motivation (4et al., 2013). Considering the results from Study 3, these are important considerations for subjects' self-regulated learning behaviors. Taken together, the results of Study 3 indicate that preference and choice regarding content can increase task engagement and improve immediate learning outcomes.

These results indicate that motivation is an important factor concerning students' self-regulating learning behavior. It is evident that motivation influences the basic decision of when to engage and disengage from learning behaviors but further research is needed to test other aspects of learning behavior and also to test factors that enhance or decrease motivation. Regarding the application of the current results to educational settings, I think it is ultimately to important to consider prescriptions from Hattie (2009), he recommends that effects for education should be above d = .40. All others below this point should only be considered in light of the possible burdens or consequences they may pose to instructors and/or students.

The current studies demonstrated that variables associated with subjective accounts of intrinsic motivation were correlated with task participation and that subjective ratings of preference were increased in groups offered higher autonomy. Subjects also consistently chose to study more during a pre-test phase and subjects offered a reward tended to decrease the amount of the time they spent reviewing during a post-test phase. Reward increased self-regulated learning behaviors in preparation for an upcoming exam, but may decrease motivation to continue engaging in learning behaviors. Therefore, it is arguably still important for instructors to incorporate factors into their classroom that can enhance intrinsic motivation but due to the small effects of undermining on self-regulated learning behaviors it is important for future researchers and instructors to consider the impact changes to classroom structure may create in regards to benefits gained for students compared to possible costs to instructors.

Limitations

Although Study 2 was designed to test the benefits of increased task time on long-term memory performance, it was limited by characteristics of the memory tests. The immediate memory test was too easy and showed ceiling effects; in addition, subjects could perform this test using a simple recognition strategy rather than retrieving the paired associate because the distractors all derived from non-studied words. In contrast, the 24-hour delay test appeared too difficult and likely showed floor effects. Cued recall was used for the 24-hour delay test in an attempt to avoid ceiling effects and control for subjects being re-tested in the same ways as in the immediate memory test; however, based on the poor performance across all subjects, it appeared that this choice was ineffective. This was addressed in Study 3 by using a more difficult, and therefore more sensitive, associative recognition test that did exhibit more variance within subjects. Additionally, Study 3 would be strengthened with the addition of a 24-hour delayed memory test. Without a delayed test it was not possible to determine if the amount of time subjects studied during the post-test review phase had any long-term benefits on memory performance.

In addition, Study 3 had a small sample size and there was an unequal distribution of subjects into reward groups and choice conditions. Additionally, there was low to moderate power for observing an interaction (d=.42), and low power (37%). It is possible that this limited the ability to detect a clearer relationship between the variables of choice and reward on subjects' behavior.

Future research testing undermining and subject' self-regulating learning behavior should also consider adding a baseline to measure subjects' behavior. One paradigm used to test the undermining effect uses three free-choice periods. The first period no rewards are introduced and subjects' initial behavior is measured to gain a baseline of behavior. For the second session, an incentive is introduced for experimental group that is later removed for the third period. The current studies were limited because the relatively small set of stimuli used did not easily allow for multiple study and review sessions. Multiple sessions would more accurately allow for a within subjects analysis of behavior. Since the undermining effect is primarily concerned with fluctuations in motivation this analysis would be preferred. Multiple testing sessions would also allow for a more accurate measure of memory performance in addition to behavioral choices.

Future Directions

The current studies focused on the changes in voluntary choices to study and persist at a learning task and future directions should focus on two main areas: exploring the impact of reward on subjects' decision to choose more effortful and helpful study strategies and isolating the impact of the reward on subjects' behavior more thoroughly. More specifically, future studies should explore the impact of rewards on subjects' decisions about what they want to re-study and more closely examine how reward interacts with other components of students' self-regulated learning behaviors in addition to the initial decision to begin studying. Dunlosky et al. (2013) highlight many techniques that students can apply to their studying in order to improve learning outcomes. Several of these techniques (e.g., elaborating material and self-testing) require students to put more effort into expanding and cognitively processing material (Dunlosky et al., 2013; Mosovitch & Craik, 1976). In order to begin this exploration, the focus should be on an initial study phase and exposure to the words and focus on subjects' initial decisions of what they want to re-study on an item-by-item basis. This procedure would differ from the current studies because those subjects only made a global decision to review all the words from the experiment after they were studied and would allow for subjects to decide at the time of study if they would want to see a specific word pair again later. Research has further shown that students' application of self-testing strategies to gauge their understanding is positively correlated with high academic achievement (Hartwig & Dunlosky, 2012). Therefore, it would also be helpful if subjects in a reward or non-reward group were given different options of how they want to review items (e.g. re-read or be tested over an item). This would allow for a beginning exploration in strategy choice between reward and non-reward groups.

Additionally, subjects' self-imposed and task-imposed goals have been shown to interact with self-regulation strategy (Ariel et al., 2009;Thiede & Dunlosky, 1999). Subjects in a non-reward group may not be focused on a performance criterion, whereas, the reward group may narrow their focus apply employ a different study strategy (e.g., only focus on easy items or repeat to themselves only ten items) in order to maximize efficiency in order to reach their goal and therefore are expected to choose to study easier items more often than hard items. It could be possible that performance-contingent rewards alter subjects' study strategy in comparison to the non-reward group subjects but the performance goal and the reward incentive are confounded in the current studies in order to more accurately generalize to traditional educational grading schemes.

Future directions should both explore the differences in subjects' decisions of what to study and isolate the impact of reward on students' strategies by separating the performance criteria from the external reward. Randomly assigning subjects to a reward or non-reward condition and additionally randomly assigning them to a performance-criteria or no-criteria condition could accomplish this. Subjects would only need to complete the initial study phase and the procedure would not need additional study and review phases in order to focus on the item-by-item decisions made in an externally compared to internally motivated state. Effects of external motivation on learning have not previously been demonstrated and therefore it is not hypothesized that external rewards alone would result in accuracy differences, but that strategies would be predicted to change as a result of rewards. It would be expected that subjects choices of which items and how to study would differ by condition.

In short, future research in this area should continue to examine how reward interacts with subjects' self-regulated learning behavior but should also expand to better understand the choices subjects make when externally motivated. The current results show that subjects increased their self-regulated learning behavior before an upcoming test compared to a post-test phase; it would be beneficial to identify how strategies and goals for approaching learning change as a result of external incentives. This additional research may help account from the discrepancy between studies such as the current study in which reward did not improve memory performance, and research described previously that did find a benefit of rewards to memory performance (Nilsson, 1987; Pochon et al., 2002) and may help educators better know how to inform students to independently apply effective learning strategies.

REFERENCES

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron*, *50*(3), 507-517.

Albrecht, K., Abeler, J., Weber, B., & Falk, A. (2014). The brain correlates of the effects of monetary and verbal rewards on intrinsic motivation. *Frontiers in neuroscience*, *8*.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology General*, *138*(3), 432.

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & behavior*, *81*(2), 179-209.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, *64*, 417-444.

Blackwell, L.S., Trzesniewski, K.H., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263, doi: 0009-3920/2007/7801-0014.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational research*, *64*(3), 363-423.

Castel, A. D., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*(7), 1078-1085.

Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. *The psychology of learning and motivation*, *48*, 225-270.

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & cognition*, *41*(1), 28-35.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, *104*(3), 268.

Crespi, L. P. (1942). Quantitative variation of incentive and performance in the white rat. *The American Journal of Psychology*, *55*(4), 467-517.

Cochran, W. G. (1952). The χ2 test of goodness of fit. *The Annals of Mathematical Statistics*, 315-345.

Cochran, W. G. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics*, *10*(4), 417-451.

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual review of psychology*, *51*(1), 171-200.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of personality and Social Psychology*, *18*(1), 105.

Deci, E. L., & Cascio, W. F. (1972). Changes in Intrinsic Motivation as a Function of Negative Feedback and Threats.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). The undermining effect is a reality after all Extrinsic rewards, task interest, and self-determination: Reply to Eisenberger, Pierce, and

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109-134.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological science*, *16*(12), 939-944.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52*(4), P178-P186.

Eisenberger, R., Pierce, W. D., & Cameron, J. (1999). Effects of reward on intrinsic motivation—Negative, neutral, and positive: Comment on Deci, Koestner, and Ryan (1999).

Elliott, E. S., & Dweck, C. S. (1988). Goals: an approach to motivation and achievement. *Journal of personality and social psychology*, *54*(1), 5.

Faul, F., Erdfelder, E., Lang, A. G. , & Buchner, A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, *49*(1), 585-612.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*(5636), 1104-1107.

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. *Handbook of educational psychology*, *4*, 63-84.

Gruber, M. J., Watrous, A. J., Ekstrom, A. D., Ranganath, C., & Otten, L. J. (2013). Expected reward modulates encoding-related theta activity before an event. *Neuroimage, 64*, 68-74.

Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: an experimental and individual difference investigation. *Journal of personality and social psychology*, *52*(5), 890.

Hagger, M. S., & Chatzisarantis, N. L. (2011). Causality orientations moderate the undermining effect of rewards on intrinsic motivation. *Journal of Experimental Social Psychology*, *47*(2), 485-489.

Harackiewicz, J. M., Abrahams, S., & Wageman, R. (1987). Performance evaluation and intrinsic motivation: The effects of evaluative focus, rewards, and achievement orientation. *Journal of Personality and Social Psychology*, *53*(6), 1015.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement?. *Psychonomic Bulletin & Review*, *19*(1), 126-134.

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. *Tertiary assessment & higher education student outcomes: Policy, practice & research*, 259-275.

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation:a review and synthesis. *Psychological bulletin*, *128*(5), 774.

Hübner, R., & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review*, *17*(6), 821-826.

Kang, S. H., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation?. *Journal of Applied Research in Memory and Cognition*, *3*(3), 183-188.

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

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci*, *21*(16), RC159.

Locke, H. S., & Braver, T. S. (2010). Motivational influences on cognitive control: a cognitive neuroscience perspective. *Self control in society, mind, and brain*, 114-140.

Malik, M., & Parveen, N. (2015). Intrinsic Motivation of the High and Low Academic Achievers. *American Journal of Educational Research*, *3*(11), 1481-1483.

Mather, M., & Schoeke, A. (2011). Positive outcomes enhance incidental learning for both younger and older adults. *Frontiers in neuroscience*, *5*(129), 1-10.

Metallidou, P., & Vlachou, A. (2007). Motivational beliefs, cognitive engagement, and achievement in language and mathematics in elementary school children. *International journal of psychology*, *42*(1), 2-15.

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

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

Meyniel, F., Sergent, C., Rigoux, L., Daunizeau, J., & Pessiglione, M. (2013). Neurocomputational account of how the human brain decides when to have a break. *Proceedings of the National Academy of Sciences*, *110*(7), 2641-2646.

Miele, D. B., Finn, B., & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science*, *22*(3), 320-324.

Moscovitch, M., & Craik, F. I. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior*, *15*(4), 447-458.

Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences*, *107*(49), 20911-20916.

Murayama, K., & Kuhbandner, C. (2011). Money enhances memory consolidation–But only for boring material. *Cognition*, *119*(1), 120-124.

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General*, *143*(1), 15.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur?. *Journal of Memory and Language*, *33*(4), 545.

Nilsson, L. G. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research*, *49*(2-3), 183-188.

Pessoa, L., & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Frontiers in neuroscience*, *4*, 17.

Pierce, W. D., Cameron, J., Banko, K. M., & So, S. (2003). Positive effects of rewards and performance standards on intrinsic motivation. *The Psychological Record*, *53*(4), 561.

Pochon, J. B., Levy, R., Fossati, P., Lehericy, S., Poline, J. B., Pillon, B., ... & Dubois, B. (2002). The neural system that bridges reward and cognition in humans: an fMRI study. *Proceedings of the National Academy of Sciences*, *99*(8), 5669-5674.

Peirce JW (2009) Generating stimuli for neuroscience using PsychoPy. Front.Neuroinform.2:10., doi:10.3389/neuro.11.010.2008

Ryan, R. M. (Ed.). (2012). Part two: general theories of motivation. In *The Oxford handbook of human motivation*. (pp. 13-85). New York, NY: Oxford University Press.

Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: the neurobiology of attentional effort. *Brain research reviews*, *51*(2), 145-160.

Schmidt, L., Lebreton, M., Cléry-Melin, M. L., Daunizeau, J., & Pessiglione, M. (2012). Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biol*, *10*(2), e1001266.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *27*(3), 285-300.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1236.

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.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*(1), 80-90.

Stolk, J., & Harari, J. (2014). Student motivations as predictors of high-level cognitions in project-based classrooms. *Active Learning in Higher Education*, *15*(3), 231-247.

Tang, S. H., & Hall, V. C. (1995). The overjustification effect: A meta-analysis. *Applied Cognitive Psychology*, *9*(5), 365-404.

Taylor, S. F., Welsh, R. C., Wager, T. D., Phan, K. L., Fitzgerald, K. D., & Gehring, W. J. (2004). A functional neuroimaging study of motivation and executive function. *Neuroimage*, *21*(3), 1045-1054.

Wehe, H. S., Rhodes, M. G., & Seger, C. A. (2015). Evidence for the negative impact of reward on self-regulated learning. *The Quarterly Journal of Experimental Psychology*, *68*(11), 2125-2130.

White, R. W. (1959). Motivation reconsidered: the concept of competence. *Psychological review*, *66*(5), 297.

White, C. B., & Fantone, J. C. (2010). Pass–fail grading: laying the foundation for self-regulated learning. *Advances in health sciences education*, *15*(4), 469-477.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441-517.

APPENDIX A

Intrinsic Motivation Inventory (IMI)

Each subscale is rated by participants on a 7-point scale (1 indicates that the subject strongly disagrees and 7 indicates that the subject strongly agrees).

Interest/Enjoyment

1. I enjoyed doing this activity very much.
2. This activity was fun to do.
3. I thought this activity was kind of boring.
4. This activity did not hold my attention at all.
5. I would describe this activity as very interesting.
6. I thought this activity was quite enjoyable.
7. While I was doing this activity, I was thinking about how much I enjoyed it.

Perceived Competence

1. I think I am pretty good at this activity.
2. I think I did pretty well at this activity, compared to other students.
3. After working at this activity for a while, I felt pretty competent.
4. I am satisfied with my performance at this task.
5. I was pretty skilled at this activity.
6. This was an activity that I couldn’t do very well.

Effort

1. I put a lot of effort into this.
2. I didn’t try very hard to do well at this activity.
3. I tried very hard on this activity.
4. It was important to me to do well at this task.
5. I didn’t put much energy into this task.

Perceived Choice

1. I believe I had some choice about doing this activity.
2. I felt like it was not my own choice to do this task.
3. I didn’t really have a choice about doing this task.
4. I felt like I had to do this.
5. I did this activity because I had to.
6. I did this activity because I had no choice.
7. I did this activity because I wanted to.

APPENDIX B

Intrinsic Motivation Inventory (IMI) Adapted for Conditions

Each subscale is rated by participants on a 7-point scale (1 indicates that the subject strongly disagrees and 7 indicates that the subject strongly agrees).

Interest/Enjoyment
1. I enjoyed doing this activity very much.
2. This activity was fun to do.
3. I thought this activity was kind of boring.
4. This activity did not hold my attention at all.
5. I would describe this activity as very interesting.
6. I thought this activity was quite enjoyable.
7. While I was doing this activity, I was thinking about how much I enjoyed it.

Perceived Competence
1. I think I am pretty good at this activity.
2. I think I did pretty well at this activity, compared to other students.
3. After working at this activity for a while, I felt pretty competent.
4. I am satisfied with my performance at this task.
5. I was pretty skilled at this activity.
6. This was an activity that I couldn't do very well.

Effort
1. I put a lot of effort into this.
2. I didn't try very hard to do well at this activity.
3. I tried very hard on this activity.
4. It was important to me to do well at this task.
5. I didn't put much energy into this task.

Perceived Choice
1. I believe I had some choice about doing this activity.
2. I felt like it was not my own choice to do this task.
3. I didn't really have a choice about doing this task.
4. I felt like I had to do this.
5. I did this activity because I had to.
6. I did this activity because I had no choice.
7. I did this activity because I wanted to.

    Choice of Language For Forced-Choice Condition

    8. I wanted to learn the words that were selected for me.
    9. It didn't matter to me which language was chosen.

    Choice of Language for Self-Determined Choice Condition

    8. I was happy with the words I chose to learn.
    9. I felt a strong preference for the language I chose to learn.