WHY DO COLLEGE STUDENTS IMPROVE THEIR LEARNING PERFORMANCE ACROSS TRIALS?

A thesis submitted to the
Kent State University Honors College
in partial fulfillment of the requirements
for University Honors

by
Sarah J. Turner
May, 2012
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ iv

LIST OF TABLES .......................................................................................................... v

ACKNOWLEDGMENTS ................................................................................................. vi

CHAPTER

I. WHY DO COLLEGE STUDENTS IMPROVE THEIR LEARNING PERFORMANCE ACROSS TRIALS?

II. EXPERIMENT 1

   METHOD

   RESULTS AND DISCUSSION

III. EXPERIMENT 2

   METHOD

   RESULTS AND DISCUSSION

IV. GENERAL DISCUSSION

REFERENCES ............................................................................................................. 20-21
LIST OF FIGURES

Figure 1. Experiment 1 Output Order.................................................................22
Figure 2. Experiment 2 Output Order.................................................................23
Figure 3. Experiment 2 Recall.................................................................
Table

1. Experiment 1 Total Points .................................................................................. 25
2. Experiment 1 Recall Percentage ......................................................................... 26
3. Experiment 1 Study Time .................................................................................. 27
4. Experiment 1 Output Order ............................................................................... 28
5. Experiment 2 Total Points ................................................................................ 29
6. Experiment 2 Total Points with Penalties .......................................................... 30
7. Experiment 2 Recall Percentage ........................................................................ 31
8. Experiment 2 Study Time ................................................................................ 32
9. Experiment 2 Output Order ............................................................................... 33
ACKNOWLEDGMENTS

I would like to thank my thesis advisor, Dr. John Dunlosky, for his assistance throughout this process, as well as my thesis committee: Dr. David Riccio, Dr. Christopher Waas, and Dr. Sara Newman. I would also like to thank my parents, Edward and Carol Turner, for their continual support throughout my college experience.
CHAPTER 1

WHY DO COLLEGE STUDENTS IMPROVE THEIR LEARNING PERFORMANCE ACROSS TRIALS?

Students need to learn a lot of material, and certainly, mastering everything is probably not an appropriate goal in many contexts. Some material may be more important than others, such as being more relevant to their majors, or being more likely to appear on an exam. Frequently, materials need to be learned by overlapping deadlines, making time a significant factor. If time is limited, students cannot effectively give equal amounts of study time to all materials. Thus, students need to effectively regulate their learning. Unfortunately, many students do not effectively regulate their learning when they enter college (Credé & Kuncel, 2008), so they must learn how to learn as they proceed toward their degree. In the present study, my main interest is in understanding how self-regulation aids students in improving their learning outcomes with task experience.

Students do improve their learning outcomes in some contexts (Castel, Farb & Craik, 2007; McGillivray & Castel, 2011; Castel, McCabe & Balota, 2009). In a study by Castel et al. (2007), participants studied lists of words paired with either positive or negative point values. For instance, a recalled word could be paired with +16, resulting in the addition of 16 points to the participant’s score, or a recalled word could be paired with -16, resulting in the subtraction of 16 points from their score. The participants were told to maximize their score by remembering
many high-value positive words, with emphasis on the negative impact of remembering high-value negative words. After all six lists, participants had a high probability of recalling the high-value words, and a low probability of recalling the negative-value words. Therefore, participants regulated their behavior in an attempt to maximize their score (Castel et al., 2009; Dunlosky & Ariel, 2011; Dunlosky & Thiede, 1988; McGillivray & Castel, 2011; Ariel & Dunlosky, 2009; Castel et al., 2007).

In another demonstration of improvements in performance with task experience, older adults were presented lists of words paired with point values (McGillivray & Castel, 2011). They were then asked to decide to bet on whether they would later recall these word pairs. If the older adults remembered a word that they had bet on, they received the number of points paired with that word. If they did not recall a word they had bet on, the older adults lost the same number correspondingly. That is, a word could be paired with the value 20, and if bet on and recalled, the participant would receive 20 points. However, if the same number was bet on but not recalled, the participant would lose 20 points. After each list of word pairs, participants were given their total score, followed by the next list. Most important, the participants earned increasingly more points across lists, suggesting that some aspect of their self-regulated learning improved with task experience. Both these studies (Castel et al., 2007; McGillivray & Castel, 2011) demonstrate how participants shape their behavior to improve their performance across lists.

In the current study, my aim is to evaluate how students improve their performance in terms of points earned across trials. In particular, I’ll evaluate the possible contribution of two factors – regulation of encoding and regulation of retrieval – to improvements in performance.
Concerning regulation of encoding, perhaps with task experience, students attend more to high value materials and attend less to the low value materials as they study them, which could increase the number of points they earn across trials. Consistent with this possibility, data from Castel et al. (2009) suggest that participants’ selective attention to point values at encoding at least partially accounts for their ability to remember high-valued information. Although plausible, selective encoding may play a smaller role than one might expect. Participants do spend more time on higher value materials (Ariel, Dunlosky & Bailey, 2009; Dunlosky & Theide, 1998), but this selective encoding may not result in increased points across lists because it may not boost memory performance. For instance, consider a study by Dunlosky and Thiede (1998), where participants were allowed to self-pace their time studying materials worth either 8 or 64 points. Although participants spent slightly more time studying the higher value materials in both conditions, the extra study time yielded relatively minimal gains in performance and hence the selective regulation of encoding had a minor impact on points earned. These findings were consistent with the labor-in-vain effect, which describes the phenomena of a large amount of additional study time resulting in little or no gain in recall (Nelson & Leonesio, 1988). Note, however, that Dunlosky and Thiede (1988) used only a single list and a single study-test phase, whereas Castel and colleagues (Castel et al. 2007; Castel et al. 2009; McGillivray & Castel, 2011) showed that participants improved their regulation across lists. Thus, selective encoding may play a larger role in improving performance when participants learn about their performance across trials and through task feedback. Perhaps most important, previous research using multiple lists has not directly measured how participants were regulating their encoding, so we do not know the degree to which regulation of encoding contributed to increased performance.
across trials. To amend these limitations, the current experiment will evaluate the contribution of selective encoding by measuring participants’ self-paced study during a multiple list task.

Concerning regulation of retrieval, students may improve their performance across lists by improving how they output material during each test. For instance, they could output the more weakly learned materials first, leaving time later for the stronger materials they knew they would recall anyway, thereby ensuring points for the more weakly learned materials; or perhaps students will output the higher values first, by virtue of their importance. The order in which searched-for materials are retrieved may demonstrate mental schemas, an idea consistent with the theory of cognitive triaging. The triaging theory states that with unlimited time for recall, materials are output in a pattern that goes from weaker materials to stronger, then back to weaker materials again (Brainerd, Reyna, Howe & Kevershan, 1990). Although previous studies have discussed the average proportions of high and low value words recalled across lists (McGillivray & Castel, 2011), whether changes in proportion recall are due to changes in output order across lists is currently unknown and hence will be investigated in the current study.

To evaluate the potential contribution of encoding and retrieval towards increasing performance across lists, I used methodology from McGillivray et al. (2011), in which participants study high- and low-value words across lists, but with two important changes. First, to analyze encoding, I included a group who paced their study of each item, which permitted measurements of time spent at encoding each word during a given study list. According to the selective-encoding hypothesis, more time should be spent (across lists) learning high value materials than low valued materials, and this focus on high-valued items is expected to increase
with task experience across lists. Second, for regulation of retrieval, we measured the output order. According to Soderstrom & McCabe (2011), participants judge higher valued items as being stronger in memory, thinking they will be more likely to remember them. Thus, perhaps participants triage the higher value (i.e., perceived to be stronger) items, and are more likely to do so as they gain task experience. If participants improve their performance across lists by the use of this kind of triaging, then we expect to see a specific pattern of output: namely, with students recalling lower value items first followed by higher value items.
EXPERIMENT 1

Method

Participants and Design. Eighty undergraduates from Kent State University participated for course credit in their introductory psychology courses. A 2 (Study Group) x 2 (List) x 12 (Point value) mixed design was used, with the first factor being manipulated between-participants and the latter two within participant. Participants were randomly assigned to either the experimenter-paced group (n = 41) or the self-paced study group (n = 39).

Materials. Seventy-two word-value paired associates were used. The stimulus words were commonly occurring nouns between four and five letters, used in McGillivray et al. (2011), and were randomly assigned to two lists of 12 words. Words within each list were randomly paired with a unique integer from 1 – 12 inclusive. LiveCode was used to program the experiment and to present all stimuli on Windows-based computers.

Procedure. Participants were told to gain as many points as possible and that remembering all materials on a list would probably not be possible; therefore, they should be selective in order to maximize their points. Participants were told that each point represented how much each word was worth, if correctly recalled. For example, if the word-value pair to be recalled was “chair – 6,” then if a participant correctly recalled the word “chair,” he or she would receive 6 points. Word-value pairs were shown one at a time for both groups. Participants in the experimenter-paced study group were instructed that they would view each word-value pair for 2
s a word, after which the next word-value pair would be presented. Participants in the self-paced study group were instructed to study each word-value pair as long as they chose before continuing onto the next pair. Participants from both groups were given their choice of a distracter task to work on (Sudoku or a crossword puzzle), for 60 s following each list. Immediately after each distracter task, participants were given a self-paced recall test, where they were asked to type all the words they could recall from the list pairs. Participants were then informed of their total score but not the results for each word. The second list began immediately after recall of the first one. Output order and total time spent on each word pair was recorded. Words with simple typing errors (e.g., pliot instead of pilot), plurals (e.g., berries instead of berry), and simple spelling errors (e.g., metel instead of metal, plumb instead of plum) were scored as correct answers.

RESULTS AND DISCUSSION

Points Earned

Given that a main goal was to explain why learners’ performance improves with task experience in this value-based learning task, we first need to establish improvements across lists. Accordingly, for each participant, we computed the number of points earned per list for each student by awarding points assigned to each word on a list that was correctly recalled. Means across participants’ points are presented in Figure 1.

To analyze these values, a 2 (Study Group: experimenter paced vs. self paced) X 2 (List: 1 or 2) analysis of variance (ANOVA) was conducted. A main effect occurred for study group, $F(1,78) = 23.07$, $MSE = 13948.67$, indicating that point values were higher for the self-paced than experimenter-paced group. Most important, a main effect also occurred for list, $F(1,78) =$
15.24, $MSE = 2235.30$. Follow-up t-tests revealed that participants earned significantly more points on list 2 than on list 1, $t(79) = 3.91$. The Study group X List interaction was not significant, $F(1,78) = .34$, $MSE = 49.43$, $p = .56$.

In summary, participants in both groups did improve earn more points across lists. The most critical remaining analyses will be focused on variables that may explain this improvement, such as increases in recall across lists. For each variable, we also present two different analyses: the first is collapsed across all point values and is meant to highlight any differences between groups. The second analysis presents the dependent variable as a function of point values, so as to provide a fine-grained analyses of whether participants focused more on higher value items (vs. lower value items) on list 2 than on list 1.

Recall Performance

One reason why scores (Figure 1) may have improved is that participants recalled more items on the second list than on the first. To evaluate this possibility, mean recall performance was computed for both lists. For the experimenter-paced group, recall percentage was .21 (SEM = .03) on trial 1 and .30 (SEM = .03) on trial 2. For the self-paced group, recall percentage was .44 (SEM = .05) on list 1 and .55 (SEM = .04) on list 2. As evident from inspection of these values, a 2 (Study Group) X 2 (List) ANOVA revealed a main effect for study group, $F(1,78) = 24.55$, $MSE = 2.32$, indicating that recall was higher for the self-paced than the experimenter-paced group. A main effect was also found for list, $F(1, 78) = 18.28$, $MSE = .39$, indicating that recall was greater for list 2 than for list 1. The Study Group X List interaction was not significant, $F(1,78) = .02$, $MSE < .001$, $p = .88$. Thus, one reason why learners may have earned more points on the second trial is that they simply recalled more items.
Of course, another reason why learners earned more points could be that as compared to list 1, on list 2 they also recalled more of the items with the higher point values. To evaluate this possibility, recall performance is plotted as a function of point value in Figure 2. Given the outcomes of the prior analyses of recall collapsed across points, it is not surprising that a 2 (Study Group) X 2 (List) X 12 (Point) ANOVA revealed a main effect for group, \(F(1,78) = 24.55, MSE = 27.78\). Most important, a main effect was also found for list, \(F(1, 78) = 18.28, MSE = 4.71\). Participants recalled more items on list 2 than on list 1. No main effect was found for points, \(F(1, 78) = .70, MSE = .11, p = .41\), and none of the interactions were significant, all Fs < 1.3. Participants did not differentially recall higher point values more often than lower point values.

**Study Time**

The question remains whether part of the improvement in points earned is also due to an increased focus on higher valued points. Mean study times for list 1 (M = 11.15, SEM = 1.73) and list 2 (M = 10.28, SEM = 1.74) were computed for the self paced group. A paired samples t-test revealed no main effect for list, \(t(60) = .45, p = .66\).

As evident from inspection (Figure 3), participants did not appear to spend relatively more time on the most valuable items on list 1 as compared to list 2. Consistent with this observation, no main effect was found for list, \(F(1, 38) = 1.20, MSE = 177.12, p = .28\), or for points, \(F(1, 38) = .03, MSE = 1.44, p = .88\). Students did not focus their study time on the higher point items any differently on list 1 than on list 2, as indicated by the lack of a list X points interaction, \(F(1, 38) = .07, MSE = 3.79 p = .79\). Thus, differential encoding, or an increased
focus on higher valued points, did not appear to be the reason for students' improvement across trials.

*Output Order*

Finally, another possibility for why learners’ scores improved across trials is that they were more likely to triage the high valued items and hence could output more of the lower valued items before proceeding to the higher value ones. The analysis of recall as a function of point value suggests that this hypothesis will be disconfirmed, because recall increased similarly across trials for all point values, as seen in Figure 4. Nevertheless, I present output order for completeness.

Most participants recalled less than 50% of the items presented and thereby would be dropped from analysis of variance. Therefore, to sidestep this problem and ensure statistical power, I collapsed output order across low, medium, or high (points 1-4, 8-11, and 9-12) point values. A 2 (group) X 2 (lists) X 3 (points: low, medium, high) ANOVA was conducted. A main effect was found for group, $F(1, 32) = 11.57, MSE = 70.86$, because participants in the self-paced group had more time to study, they recalled more items than those in the experimenter-paced group. No main effect was found for list, $F(1, 33) = 2.22, p = .15$. There was no difference in the output order between list 1 and list 2 for either group. There were no significant interactions, all $Fs < 1.8$. 
EXPERIMENT 2

It was surprising that regulation of encoding did not play a larger part in the improvements in points earned from list 1 to list 2, considering that it appeared to play a role in McGillivray and Castel (2011). However, their work used a betting paradigm, which introduced negative consequences in addition to incentives, which presumably enhanced students’ regulation. In our study, there was not a lot of pressure to regulate. With unlimited study time, students could use whatever strategies they thought best to recall all the items, and hence may not have needed to differentially allocate more time to the more highly valued items.

To evaluate this possibility, I used a penalty group in Experiment 2. In this penalty group, participants could study as long as they chose, but they lost 1 point for each second they spent studying. If they spend more than a few seconds studying lower value items, they would actually lose points even if they recalled them. In this case, students may better regulate their encoding by using more time studying the higher value items and little time for the lower valued items. If students learn to use this strategy more effectively across lists, then it would provide some evidence that regulation of encoding contributes to increases in points earned. Besides the addition of this penalty group, I also used 6 study-test trials, so as to evaluate whether students continue to improve their scores with even more task experience.

Method

Participants. Sixty-one undergraduates from Kent State University participated for course credit in introductory psychology; they were randomly assigned to the no penalty group \( n = 30 \) or the penalty group \( n = 31 \).
**Materials and procedure.** Materials were the same as those used in Experiment 1.

Procedures in Experiment 2 were identical to those of Experiment 1, with few exceptions. First, the penalty group was additionally instructed that they would lose 1 point for each second of study. After reading the instructions, participants were required to answer a brief test on the instructions before beginning to study the materials. For example, test questions included “Will you be permitted to study the material as long as you want?” If students answered a question incorrectly, they were redirected back to the instruction screen, where they could find the correct answer before continuing. By this means, we ensured that students understood what was expected of them.

**RESULTS AND DISCUSSION**

**Points Earned**

We analyzed the number of points earned in two ways, with no penalty for time spent studying (as in Experiment 1) and with a penalty for each second studying. The former analysis is most appropriate for the no penalty group, and the latter analysis is most appropriate for the penalty group. Nonetheless, we present both measures for each group for completeness.

*Total points earned without penalty.* As shown in Figure 5, results from the no penalty group replicated Experiment 1, with participants earning more points on list 2 than on list 1. The penalty group also appeared to earn more points across lists. Consistent with these observations, a 2 (Penalty group) X 6 (List) ANOVA revealed a main effect of penalty group, $F(1, 59) = 13.92$, $MSE = 15548.13$, indicating that more points were earned by the no penalty group. A main effect of list also occurred, $F(1, 59) = 15.95$, $MSE = 3455.76$. To follow-up on the main effect of list, we compared points earned on adjacent lists, collapsed across groups (because the
interaction was not significant, $F[1, 59] = .06, MSE = 13.15, p = .81$). Points earned increased from list 1 to list 2, $t(60) = 1.95, p = .03$ (one-tailed test, given replication of gains in Experiment 1). All of the other contrasts were not significant, $t < 2.0$.

**Total points earned with penalty.** As shown in Figure 6, results from the penalty group replicated Experiment 1, with participants earning more points on list 2 than on list 1. Consistent with these observations, a 2 (Penalty group) X 6 (List) ANOVA for adjusted totals (participants' totals after the penalties had been subtracted) revealed a main effect of penalty group, $F(1, 59) = 9.86, MSE = 63712.05$, indicating that more points were earned by the no penalty group. A main effect of list also occurred, $F(1, 59) = 22.05, MSE = 34094.15$. To follow-up on the main effect of list, we compared points earned on adjacent lists, collapsed across groups (because the interaction was not significant, $F[1, 59] = .03, MSE = 45.25, p = .87$). Points earned increased from list 1 to list 2, $t(60) = 1.66, p = .05$ (one-tailed test, given replication of Experiment 1), from list 2 to list 3, $t(60) = 2.95, p < .02$, and from list 4 to list 5, $t(60) = 2.17, p = .04$. All of the other contrasts were not significant, $t < 1.0$. However, it is noteworthy that although not all improvements reached significance, adjusted totals for both groups consistently improved across all lists.

**Recall Performance**

As we considered in Experiment 1, one reason why scores (Figures 5 & 6) may have improved is that participants recalled more items on list 2 than on list 1. To evaluate this possibility, mean recall performance was computed for just these lists. For the no-penalty group, recall percentage for trials 1-6 (respectively) were .39 (SEM = .05), .51 (.05), .47 (.05), .50 (.05), .48 (.04), and .52 (.05). These results replicate and extend the results from the self-paced group
in Experiment 1. For the penalty group, recall percentage for trials 1-6 (respectively) were .27 (SEM = .04), .30 (.03), .36 (.04), .33 (.03), .33 (.03), and .37 (.03). As evident from inspection of these values, a 2 (Study Group) X 6 (List) ANOVA revealed a main effect for study group, $F(1,59) = 12.06, \text{MSE} = 2.17$, indicating that recall was higher for the no-penalty than the penalty group. A main effect was also found for list, $F(1, 59) = 7.51, \text{MSE} = .25$, indicating that recall was greater for list 2 than for list 1. The Study Group X List interaction was not significant, $F(1,59) < .01, \text{MSE} < .001, p = .95$. Thus, it appears that one reason why learners performed better on successive lists in terms of points earned is that they simply recalled more items.

Of course, another reason why learners earned more points could be that as compared to previous lists, they also recalled more of the items with the higher point values. Because most of the differences occurred between lists 1 and 2, all further analyses were conducted considering only these two groups. If desired, values for all six lists may be found in Table 3. To evaluate this possibility, recall performance is plotted as a function of point value in Figure 7. A 2 (Study Group) X 2 (List) X 12 (Point) ANOVA was conducted. A main effect was found for group, $F(1,59) = 12.06, \text{MSE} = 26.00$. Participants in the no penalty condition recalled more items than participants in the penalty group. Most important, a main effect was also found for list, $F(1, 59) = 7.51, \text{MSE} = 3.05$. Participants recalled more items on list 2 than on list 1. A main effect was found for points, $F(1, 59) = 10.38, \text{MSE} = 4.09$. Follow-up t-tests were collapsed across low, medium and high point values for simplification (values 1-4, 5-8, and 9-12), and revealed differences between the no penalty group's recall of low-value vs. high-value words, $t(29) = 1.73$, and between their recall of medium-value vs. high-value words, $t(29) = 1.65$. Participants recalled more low-value words (M = .42, SEM = .06) than medium-value words (M = .33, SEM
and significantly more high-value words (\(M = .44, \text{SEM} = .07\)) than medium-value words (\(M = .33, \text{SEM} = .06\)). No differences were found between the number of points recalled for any other comparisons in either group, all \(ts < 1.0\). The interaction of list X point was significant, \(F(1, 59) = 10.19, MSE = 2.66\). No other interactions were significant, all \(Fs < 1.8\). Participants did not recall higher valued items more often than the lower valued items on either list, regardless of their group.

**Study Time**

The question remains as to whether part of the improvement in points earned is also due to an increased focus on higher valued points. A main effect was found for group, \(F(1, 59) = 16.82, MSE = 987.56\). Participants in the no penalty group spent more time studying than participants in the penalty group. A main effect was found for list, \(F(1, 59) = 10.64, MSE = 110.01\). No interaction was found for list X group, \(F(1, 59) = .07, MSE = .74, p = .79\). Differences in participants’ study time across lists did not differ significantly for either group.

As evident from inspection (Figure 8), participants did not appear to spend relatively more time on the most valuable items across lists. A main effect was found for group, \(F(1, 59) = 16.82, MSE = 11850.66\). Participants in the no penalty group spent more time studying than participants in the penalty group. A main effect was found for list, \(F(1, 59) = 10.64, MSE = 1320.07\). Follow up T-tests were run to determine where these differences lay. Participants in the penalty group studied list 1 longer than list 2, \(t(30) = 2.75\). Participants in the no penalty group did not study list 2 significantly longer than list 1, \(t(29) = .84\). No main effect was found for points, \(F(1, 59) = 3.08, MSE = 45.11, p = .09\). Unexpectedly, the interaction for point X group was not significant, \(F(1, 59) = 2.67, MSE = 39.09, p = .11\). All other interactions were also not
significant, $Fs < .08$. Thus, in contrast to expectations, participants in the penalty group did not regulate their study time towards the high-value words any more than participants who were being penalized for each second spent studying.

*Output Order*

As in the first experiment, most participants recalled less than 50% of the items presented and these participants would be dropped from the analysis of variance. Therefore, analysis for output order was again collapsed into three categories (low, medium and high point values). All values may be found in Table 2. A main effect was found for group, $F(1, 28) = 7.89, MSE = 31.04$. As may be expected due to the construction of the study, participants in the no penalty group recalled more points than participants in the penalty group, and therefore more values were present to be analyzed than in the penalty group, thereby causing a difference in their output order - participants in the no penalty group often recalled seven or more items, and participants in the penalty group recalled much fewer. No main effect was found for list, $F(1, 28) = 2.43, MSE = 5.25, p = .13$. Participants in either group did not output their recalled items any differently between lists 1 and 2. The interaction between group X list was not significant, $F(1, 28) = 2.91, MSE = 6.31$. 
GENERAL DISCUSSION

The current study used value to determine why students improve across trials. The results show that students produced gains in performance or reward across trials. Students' goals in this study were to maximize their score, which could be done either by recalling many words paired with low-point values, or by recalling a few high-point values. Improvements in either, or through a combination of these, would result in improvement across trials. Previous research suggested that students improved because of value-based encoding, or increased focus on higher valued items during study (e.g., Castel et. al 2009, McGillivray & Castel 2011). Another possibility is that value-based retrieval contributed to increase in points earned across lists.

Evidence from two experiments suggests that neither value-based encoding nor retrieval contribute to the gains in performance. I included a self-paced study group into previous methods (e.g., McGillivray & Castel 2011), with the expectation that students would self-regulate themselves to spend differential amounts of time on the higher valued than the lower valued items, in order to increase their reward. However, students did not differentiate. In Experiment 2, participants were penalized for each second they spent studying, which was expected to pressure them to focus on the higher valued, and more rewarding, word-value pairs. However, even in Experiment 2, students did not differentiate between high- and low- point word-value pairs. Value-based retrieval did not play a significant role in students' improvement across trials, either. High-value words were not recalled any more frequently than low-value words in either experiment.
After considering these findings, it was necessary to look elsewhere for an explanation of students' improvements in reward across trials. Most important, time pressure played a significant role in students' improved performance in Experiment 2. When students were penalized for the time spent studying, they spent less time studying without dropping in performance, thereby becoming more efficient. This aspect of self-regulation did contribute to gains. However, time pressure cannot explain all improvements, as people did not speed up (or get penalized for being slow) in Experiment 1, where gains were also seen. Another possible explanation is that students simply learned how to learn. Across trials, students may simply have realized what was expected of them, and what was effective or ineffective in their attempts to maximize their score. For instance, perhaps students switched to more effective learning strategies across lists.

These findings do not conceptually replicate the value-based improvements in self-regulation reported by McGillivray and Castel (2011), which this study was based upon. Rather, this study tends to replicate research by Dunlosky and Theide (1998), showing that students don't effectively self-regulate their value-based learning. Therefore, it is important to note the differences between this study and its predecessor. A very important difference between our studies are the instructions. McGillivray and Castel (2011) gave students specific instructions on exactly how they could improve their scores, which resulted in large improvements across trials. However, in the present experiments, students were expected to learn how to improve on their own, as we were interested in spontaneous improvement. Therefore, students in this experiment were told that remembering all items would not be possible, so they should develop strategies in order to improve their score. These differences mean that McGillivray and Castel (2001) studied whether students improved across trials when told how to do so, which his results supported. However, we measured whether students spontaneously improved across trials. We discovered that
students do earn more points across trials even when not given specific instructions about how to regulate their encoding. These improvements were not maximal, but it is important to note that these findings are indicative of how students learn in real-world situations.
REFERENCES


Table 1. Average output order as a function of group. High value is the average output for items with values 9-12.

<table>
<thead>
<tr>
<th>Group</th>
<th>Experimenter-paced</th>
<th>Self-paced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>1</td>
<td>2.29</td>
<td>2.57</td>
</tr>
<tr>
<td>2</td>
<td>2.66</td>
<td>2.56</td>
</tr>
</tbody>
</table>
Table 2. Average output order as a function of group. High value is the average output for items with values 9-12.

<table>
<thead>
<tr>
<th>Group</th>
<th>No Penalty</th>
<th></th>
<th>Penalty</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>М</td>
<td>М</td>
<td>М</td>
</tr>
<tr>
<td>List</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.33</td>
<td>4.13</td>
<td>3.45</td>
</tr>
<tr>
<td>2</td>
<td>3.56</td>
<td>4.03</td>
<td>4.35</td>
</tr>
</tbody>
</table>
Table 3. Average recall percentage as a function of group. High value is the average recall for items with values 9-12.
Figure 1. Average total points recalled as a function of group for lists 1 and 2. Error bars represent standard error of the mean.
Figure 2. Recall percentage for each point value, as a function of group for lists 1 and 2. Error bars represent standard error of the mean.
Figure 3. Time spent studying as a function of points, for the self-paced study group. Error bars represent standard error of the mean.
Figure 4. Average output order as a function of point. Error bars represent standard error of the mean.
Figure 5. Average total points recalled as a function of group for lists 1 through 6. Error bars represent standard error of the mean.
Figure 6. Average total points after penalties, as a function of group. Error bars represent standard error of the mean.
Figure 7. Percentage recalled as a function of point. Error bars represent the standard error of the mean.
Figure 8. Time spent studying as a function of point. Error bars represent the standard error of the mean.
Figure 9. Average output order as a function of point. Error bars represent standard error of the mean.