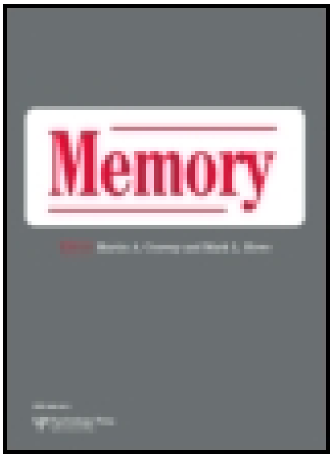

Memory
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pmem20

How do students improve their value-based learning with task experience?

Sarah DeLozier[a] & John Dunlosky[b]

[a] Department of Psychology, Colorado State University, Fort Collins, CO, USA

[b] Department of Psychology, Kent State University, Kent, OH, USA

Published online: 23 Jul 2014.

Memory, 2014
http://dx.doi.org/10.1080/09658211.2014.938083

Routledge
Taylor & Francis Group

How do students improve their value-based learning with task experience?

Sarah DeLozier[1] and John Dunlosky[2]

[1]Department of Psychology, Colorado State University, Fort Collins, CO, USA
[2]Department of Psychology, Kent State University, Kent, OH, USA

(*Received 3 November 2013; accepted 18 June 2014*)

When learning items that vary in reward, students improve their scores (i.e., earned reward) with task experience. In four experiments, we examined whether such improvements arise from better selective encoding of items that would earn more (vs. less) reward. Participants studied and recalled words across multiple study-test trials. On each trial, 12 words were slated with different values (typically from 1 to 12), and participants earned the point value assigned to a given word if it was correctly recalled. In all experiments, participants earned more points across the first two trials. In Experiment 1, participants either self-paced their study or had experimenter-paced study and in Experiment 2, some participants were penalised for each second spent during study. Improvements in points earned were related to increases in overall recall but not to selective encoding. In Experiment 3, some participants were given value-emphasised instructions, yet they did not demonstrate selective encoding. In Experiment 4, we used a larger range of point values, but selective encoding still did not account for the improvement in point scores across lists. These results suggest that metacognitively-driven selective encoding is not necessary to observe improvements in value-based learning.

Keywords: Value-based learning; Selective encoding; Learning how to learn; Metamemory.

Students need to learn a lot of material, and certainly, mastering everything is probably not an appropriate goal in many contexts, especially given limits on the time available for study. Fortunately, some material is typically more important than other material, whether it is more relevant to the students' interests or just more likely to appear on an exam. Thus, students may benefit from prioritising the most important—or most valuable—information as they study. Unfortunately, many students do not effectively regulate their learning when they enter college (Credé & Kuncel, 2008) and in general do not regulate optimally, so they must learn how to learn. In the present study, our main goal was to explore the degree to which students can improve their performance outcomes as they obtain experience studying and recalling lists comprised of words that vary in reward. Will students obtain higher point values across lists as they obtain task

Address correspondence to: Sarah DeLozier, Department of Psychology, Colorado State University, Fort Collins, CO 80523-1876, USA. E-mail: sarah.delozier@colostate.edu

This research was conducted as a senior thesis by the first author at Kent State University and was supported by the James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind, and Behavior Collaborative Award, and by the National Science Foundation Graduate Research Fellowship Program [grant number DGE-1321845], Amendment II. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

experience? And, if they do, will they do so by learning to selectively encode the more highly valued words across lists?

Evidence from prior research demonstrates that students consider value information when they study and perform better as they obtain task experience. First, when people are able to regulate their learning, such as by pacing their study or by selecting a subset of items to study, they tend to prioritise the items that will earn them more points on the upcoming test (Castel, 2008). That is, they spend more time studying (or are more likely to select for study) those items that will be worth the most if they are answered correctly on the test or those items that are more likely to appear on the test (e.g., Ariel, Dunlosky, & Bailey, 2009; Castel, Murayama, Friedman, McGillivray, & Link, 2013; Dunlosky & Thiede, 1998; Soderstrom & McCabe, 2011). Second, and most relevant to our current goal, people also tend to improve their learning outcomes with task experience in some value-based contexts (Castel, Balota, & McCabe, 2009; McGillivray & Castel, 2011). For instance, McGillivray and Castel (2011) had college students study lists of words, with each word on a list paired with a different point value. The students were asked to bet on whether they would later recall each word pair. If they recalled 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, they lost the points. That is, a word could be paired with the value 8, and if bet on and recalled, the participant would receive eight points. However, if the same word was bet on but not recalled, the participant would lose eight 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.

In the current study, our main goal was to evaluate how students improve their performance in terms of points earned across trials, with a special focus on the contribution of metacognitively-guided selective encoding. In particular, each word on a 12-word list was slated with a different integer from 1 to 12. The word and its value were presented for study, and after all words were presented for study, participants attempted to recall words from the list. This procedure was repeated one or more times with a new list of words. To foreshadow, participants earned more points across trials, which allowed us to evaluate the *encoding-selectivity hypothesis* that states people will selectively encode higher-value words over lower-value ones. As important, for selective encoding to explain increases in points across trials, participants should focus more on higher-value words (i.e., enhance their selective encoding) *across* trials. To evaluate this hypothesis, we had some participants pace the study of each word–value pair, with a main question being, Will participants spend more time studying higher-value words (than lower-value ones) and will any emphasis on higher-value words increase across lists?

Although plausible, selective encoding may play a smaller role than one might expect when people pace their study of items. People sometimes spend more time on higher-value items (Ariel et al., 2009; Dunlosky & Thiede, 1998; for an exception, see Le Ny, Denhière, & Le Taillanter, 1972). However, such selective encoding may not always result in increased points across lists because it may not boost memory performance. For instance, in research by Dunlosky and Thiede (1998), participants were allowed to self-pace their study of words worth either 8 or 64 points. Although participants spent more time studying the higher-value words, the extra study time (which was massed) yielded relatively minimal gains in performance, and hence selective encoding had a minor impact on points earned. These findings were consistent with the labour-in-vain effect, which describes the phenomenon of additional study time resulting in little or no gain in recall (Nelson & Leonesio, 1988). Note, however, that Dunlosky and Thiede (1998) used only a single list and a single study-test phase, whereas McGillivray and Castel (2011; see also Castel, Farb, & Craik, 2007; Castel et al., 2009) showed that participants improved their performance across multiple study-test phases.

Selective encoding may play a larger role in improving performance when participants gain task experience and receive feedback about how they are performing across lists. In fact, a recently published study provides evidence relevant to this point. Castel et al. (2013) presented participants with an array that included point values from 1 to 30. When participants selected a point value from the array, they were then presented with a word (slated with that value) and could study it as long as they wanted. They could select words for study in any order and study each one as long as they wanted, at least until the 2-minute time period elapsed. Participants then attempted to recall the words, and this procedure was repeated with new words. They spent more time studying higher-value words (over lower-value ones) and this propensity to focus on higher-value words increased across lists. These data support the encoding-selectivity hypothesis.

Despite this support for the hypothesis, such selective encoding may be less prevalent in the current context in which each word–value pair is presented *individually* for self-paced study. First, selective study of higher-value words presumably arises because learners are goal oriented and develop an agenda to focus on the higher-value words. Importantly, such agenda-regulated study is more prevalent when items are presented in an array (as used by Castel et al., 2013, discussed above) than when they are presented individually for study, because the array triggers agenda development as well as supports the effective execution of any agenda that is developed (Dunlosky & Ariel, 2011; Dunlosky & Thiede, 2004). Second, when items are presented individually for study, participants may be less likely to develop an agenda to focus on higher-value items, and if so, their study time would be less influenced by value and instead be driven more by item difficulty. The idea here is that self-paced study of individual items may often be data driven (Koriat, Ackerman, Adiv, Lockl, & Schneider, 2014), and it is this data-driven aspect of self-paced study that can undermine agenda development and its execution during the self-paced study of individual items. Put differently, selective encoding of higher-value words is a metacognitively driven enterprise that involves agenda development and execution that may be minimised during self-paced study of individual items. Exploring this issue is important because this learning context (self-paced study of individual items) is common and may represent a critical boundary condition to when explicit metacognitive control will contribute to task performance.

EXPERIMENT 1

In summary, important and open issues include whether learners will spend more time studying higher-value than lower-value items when they are presently individually for study, and most important, whether any preferential study of higher-value items will increase with task experience. To explore these issues in Experiment 1, we evaluated the contribution of selective encoding by measuring participants' self-paced study during a multiple list task in which each to-be-remember word was slated with a different point value. According to the selective-encoding hypothesis, participants will use more time studying higher-value than lower-value words, and this focus on higher-value words will increase across lists. In addition to this self-paced group, we also included a group in which the presentation of words for study was experimenter-paced, which more closely matched the methods that have been used to explore value-based learning across multiple study-test trials (for a review, see Castel, 2008).

Method

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

Materials and procedure. Twenty-four word–value pairs were used. The stimulus words were commonly occurring nouns of four or five letters (from McGillivray & Castel, 2011) and were randomly assigned to 2 lists of 12 words. Words within each list were randomly paired with a unique integer from 1 to 12. LiveCode was used to programme the experiment and to present all stimuli on Windows-based computers. Participants were told to gain as many points as possible, and 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 six 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 seconds 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-pair as long as they chose before continuing onto the next pair. In particular, instructions for the self-paced study group were as follows:

> In this experiment, you will be asked to study a series of words and recall them; each word has a different point value assigned to it, and your goal is to gain as many points as possible. After clicking "Continue Experiment," you will be presented with the words and their values. After 12 word-value pairs have been presented, you will work on a crossword puzzle for a predetermined length of time, and you will then be asked to recall as many words as you can. For every word you recall correctly, you will receive the point value originally paired with that word. After you study and recall a list, the computer will tell you how many points you earned, and then, you will move on to the next list. Additional points will be given for each completed word or number in the puzzle. This entire procedure will continue for 2 lists, so do your best to maximize your score.
> Importantly, you may study each word-value pair for as long as you wish before continuing to the next pair. You will be shown the word-value pair with a button underneath it called "Next Item." When you're finished studying a particular word-value pair, simply click on the "Next Item" button to move to the next item. Your object is to maximize your score. You will not receive any points or any penalties for incorrectly answering or for failing to recall a word.

For 60 seconds following each list, participants from both groups were given their choice of a distracter task to work on (Sudoku or a crossword puzzle). 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 immediately preceding list. Participants were then informed of their total score. Presentation of the second list began immediately after recall was completed for the first one.

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 needed to establish improvements in points earned across lists. Accordingly, for each participant, we computed the number of points earned by awarding points assigned to each word that was correctly recalled. Recall was hand scored; besides correctly spelled target words, words with typing errors (e.g., pliot instead of pilot), plurals (e.g., berries instead of berry) and spelling errors (e.g., metel instead of metal, plumb instead of plum) were scored as correct.

For the experimenter-paced group, mean total points earned were 17.07 (*SE* = 2.29) on list 1 and 23.44 (*SE* = 2.13) on list 2. For the self-paced group, mean total points earned were 34.82 (*SE* = 4.03) on list 1 and 43.62 (*SE* = 3.57) on list 2. To analyze these values, a 2 (study group: experimenter-paced vs. self-paced) × 2 (list: 1 or 2) analysis of variance (ANOVA) was conducted. A main effect occurred for study group, $F(1, 78) = 23.61$, $MSE = 14{,}373.08$, $p < .001$, $\eta^2 = .232$, indicating that points earned were higher for the self-paced than experimenter-paced group. Most important, a main effect was found for list, $F(1, 78) = 15.44$, $MSE = 2297.04$, $p < .001$, $\eta^2 = .17$. A follow-up *t*-test collapsed across group, as the list × study group interaction was not significant, $F(1, 78) = .40$, $p = .53$, revealed that participants earned significantly more points on list 2 than on list 1, $t(79) = 3.93$, $p < .001$, Cohen's $d = .35$.

Given that participants earned more points across lists, the critical analyses will now focus on variables that may explain this improvement, such as increases in selective encoding or recall across lists. For each variable, we present two analyses. The first analysis is collapsed across all point values and is meant to highlight any differences between groups. The second analysis directly assesses selective encoding by presenting 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 words (vs. lower-value words) on list 2 than on list 1.

Recall performance. One reason why scores may have improved is that participants recalled more words on the second list than on the first. To evaluate this possibility, we analyzed mean recall performance. For the experimenter-paced group, the mean proportion of correct recall was .21 (*SE* = .03) on trial 1 and .30 (*SE* = .03) on trial 2. For the self-paced group, it was .44 (*SE* = .05) on list 1 and .55 (*SE* = .04) on list 2. As evident from inspection of these values, a 2 (study group) × 2 (list) ANOVA revealed a main effect for study group, $F(1, 78) = 24.55$, $MSE = 2.32$, $p < .001$, $\eta^2 = .239$, indicating that recall was higher for the self-paced than the experimenter-paced group. Most important, a main effect was also found for list, $F(1, 78) = 18.28$, $MSE = .39$, $p < .001$, $\eta^2 = .190$, indicating that recall was greater for list 2 than for list 1. The list × study group interaction was not significant, $F(1, 78) = .02$, $p = .88$. Thus, one reason why learners earned more points on the second trial is that they simply recalled more words on the second list than on the first one, $t(79) = 4.30$, $p < .001$, $d = .37$.

Another explanation for the improved performance on the second list is that participants also recalled more words with the higher point values for list 2 than for list 1. To evaluate this possibility, recall performance was examined as a function of point value (Table 1), which were collapsed into three bins (low points: 1–4; medium: 5–8; and high: 9–12) to reduce noise in the analyses (as per Castel et al., 2007). Given the outcomes of the analyses of recall collapsed across points, it is not surprising that a 2 (study group) × 2 (list) × 3 (points: low, medium, high) ANOVA revealed that recall was significantly higher for the self-paced than the experimenter-paced group, $F(1, 78) = 24.55$, $MSE = 6.95$, $p < .001$, $\eta^2 = .239$. A main effect was also found for list, $F(1, 78) = 18.28$, $MSE = 1.18$, $p < .001$, $\eta^2 = .190$, indicating that participants recalled more words on list 2 than on list 1. However, no main effect was found for points, $F(2, 156) = .41$, $p = .66$, and none of the interactions were significant, all Fs < 1.38. Thus, participants did not differentially recall words with higher point values more often than those with lower point values for either list.

Study time. A question remains as to whether part of the improvement in points earned for the self-paced group is also due to an increased focus on higher-value than lower-value words, as predicted by the encoding-selectivity hypothesis. Mean study times (in seconds) were not different for words from list 1 ($M = 11.15$, $SE = 1.73$) and list 2 ($M = 10.28$, $SE = 1.74$), $t(38) = 1.09$, $p = .28$, $d = .08$. As evident from inspecting Table 2, participants did not appear to spend relatively more time on the most valuable words on list 2 as compared to list 1. Consistent with this observation, no main effect was found for either list, $F(1, 38) = 1.20$, $p = .28$, or points, $F(2, 76) = 0.17$,

TABLE 1
Average recall proportions for Experiments 1, 2, 3 and 4 as a function of group

Value	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
List	*Experimenter-paced*			*Self-paced*		
1	0.19 (.03)	0.18 (.03)	0.25 (.04)	0.46 (.06)	0.42 (.06)	0.46 (.05)
2	0.30 (.04)	0.30 (.04)	0.30 (.03)	0.52 (.04)	0.57 (.05)	0.55 (.06)
	Penalty			*No penalty*		
1	0.28 (.05)	0.27 (.05)	0.27 (.04)	0.42 (.06)	0.33 (.06)	0.44 (.07)
2	0.31 (.05)	0.31 (.04)	0.29 (.05)	0.50 (.06)	0.49 (.07)	0.54 (.06)
3	0.33 (.05)	0.39 (.06)	0.35 (.05)	0.45 (.06)	0.45 (.06)	0.52 (.06)
4	0.28 (.04)	0.31 (.04)	0.39 (.05)	0.48 (.06)	0.53 (.06)	0.50 (.06)
5	0.29 (.05)	0.34 (.04)	0.36 (.06)	0.37 (.06)	0.52 (.04)	0.55 (.06)
6	0.32 (.05)	0.35 (.06)	0.44 (.05)	0.40 (.06)	0.56 (.06)	0.63 (.06)
	Standard			*Value-emphasised*		
1	0.46 (.05)	0.43 (.05)	0.37 (.05)	0.47 (.07)	0.41 (.06)	0.45 (.06)
2	0.41 (.05)	0.46 (.05)	0.48 (.05)	0.54 (.06)	0.55 (.05)	0.49 (.06)
3	0.57 (.05)	0.54 (.06)	0.51 (.05)	0.51 (.05)	0.54 (.06)	0.60 (.05)
4	0.60 (.05)	0.58 (.05)	0.60 (.05)	0.56 (.05)	0.58 (.05)	0.67 (.05)
5	0.53 (.06)	0.56 (.05)	0.55 (.05)	0.59 (.04)	0.59 (.04)	0.60 (.05)
6	0.54 (.06)	0.48 (.05)	0.52 (.05)	0.54 (.05)	0.56 (.05)	0.49 (.05)
	Experimenter-paced			*Self-paced*		
1	0.05 (.01)	–[a]	0.08 (.01)	0.14 (.02)	–[a]	0.13 (.02)
2	0.06 (.01)	–[a]	0.11 (.02)	0.15 (.02)	–[a]	0.20 (.03)

[a]For Experiment 4, values were collapsed across 1–10 points (low) and 90–100 points (high).

For Experiments 1–3, low, medium and high refer to the value awarded for correct recall, with values being collapsed across 1–4 points (low), 5–8 points (medium) and 9–12 points (high).

TABLE 2
Average self-paced study time (in seconds) for Experiments 1, 2, 3 and 4 as a function of group

Value	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
List	*Experimenter-paced*			*Self-paced*		
1	–	–	–	11.0 (14.6)	11.1 (13.7)	11.4 (14.4)
2	–	–	–	10.2 (14.8)	10.2 (13.0)	10.5 (13.4)
	Penalty			*No penalty*		
1	3.8 (3.5)	4.1 (4.6)	3.4 (2.8)	6.9 (9.7)	5.9 (6.3)	6.7 (7.3)
2	3.0 (2.7)	3.0 (3.2)	2.7 (2.5)	6.4 (7.8)	7.6 (7.9)	7.4 (8.7)
3	2.7 (3.2)	2.7 (3.5)	2.8 (3.0)	5.4 (6.4)	5.6 (5.8)	5.9 (5.8)
4	2.2 (2.2)	2.4 (2.5)	2.5 (2.6)	5.9 (7.1)	5.9 (6.8)	5.9 (8.0)
5	2.2 (2.4)	2.1 (2.1)	2.1 (2.1)	4.1 (4.3)	4.6 (4.0)	5.0 (5.0)
6	1.9 (2.1)	1.9 (1.7)	2.1 (2.5)	6.0 (9.0)	5.9 (10.0)	6.0 (6.2)
	Standard			*Value-emphasised*		
1	8.9 (9.9)	9.3 (10.5)	10.7 (14.6)	10.9 (12.7)	10.9 (12.4)	11.6 (14.1)
2	7.9 (10.9)	8.8 (10.5)	8.7 (11.2)	9.3 (11.7)	8.3 (9.0)	11.5 (14.3)
3	8.5 (9.3)	8.9 (10.0)	7.9 (9.1)	7.7 (9.5)	9.4 (12.1)	9.8 (12.1)
4	7.3 (8.3)	8.5 (10.3)	8.6 (9.5)	6.8 (8.4)	8.1 (9.5)	8.2 (8.8)
5	7.5 (12.0)	7.7 (10.3)	7.5 (8.0)	7.9 (9.1)	8.2 (9.4)	7.5 (7.4)
6	6.8 (8.9)	7.4 (8.6)	7.7 (9.8)	6.5 (7.8)	8.0 (14.9)	6.0 (6.9)
	Experimenter-paced			*Self-paced*		
1	–	–	–	5.8 (6.4)	–[a]	6.2 (5.8)
2	–	–	–	3.9 (4.3)	–[a]	4.9 (5.1)

[a]For Experiment 4, values were collapsed across 1–10 points (low) and 90–100 points (high).
For Experiments 1–3, low, medium and high refer to the value awarded for correct recall, with values being collapsed across 1–4 points (low), 5–8 points (medium) and 9–12 points (high).

p = .84. Participants also did not focus their study time on the higher-value words differently on list 1 than on list 2, as indicated by the lack of a list × points interaction, $F(2, 76) < 1.0$.

Selectivity index. To supplement the analyses of recall, we also computed a selectivity index (SI; Watkins & Bloom, 1999) that reveals the degree to which participants selectively recalled the higher-value words over others. The SI involves computing the degree to which a participant's recall is sensitive to point value relative to chance and to ideal scores given a specific level of recall: SI = (participant's score–chance score)/(ideal score–chance score). For example, if a participant recalled three words, with values of 2, 3 and 4, that participant's SI would be low, resulting in a SI of (9–19.5)/(33–19.5) = −.78. The ideal SI for recalling three words would occur if the recalled words were valued at 12, 11, 10; in this case, SI would reach the maximum of 1.0. When words are recalled close to chance, with no regard to their respective point values, the SI would be close to 0 (for additional details, see Castel, Benjamin, Craik, & Watkins, 2002; Watkins & Bloom, 1999). For the experimenter-paced group, mean SIs for lists 1 and 2 (respectively) were .06 (SE = .09) and −.04 (.07), and for the self-paced group, they were −.03 (SE = .09) and .21 (.07). An ANOVA did not yield a significant main effect for study group, $F(1, 63) = 1.35$, $p = .25$, or for list, $F(1, 63) = 1.52$, $p = .22$, but the list × study group interaction was significant, $F(1, 63) = 4.95$, MMSE = 1.02, $p = .03$, $\eta^2 = .073$. The interaction revealed that participant's recall did become more selective across trials, but only for the self-paced group.

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 that introduced negative consequences in addition to incentives, which could have increased the likelihood that participants would develop an agenda to allocate more time to higher-value words. In our study, little pressure was put upon participants to differentially allocate time to higher- (vs. lower-) value words. That is, with unlimited study time, participants could use whatever strategies they thought best to recall *all* the words, and hence may not have believed they needed to differentially allocate more time to the higher-value words.

To evaluate this possibility in Experiment 2, we contrasted a no-penalty group (as per Experiment 1) with a penalty group. For the penalty group, participants could study as long as they chose, but they lost 1 point for each second they spent studying. Thus, if they spend more than a few seconds studying lower-value words, they would actually lose points even if they correctly recalled them. With this penalty in place, participants may view selective encoding (i.e., relatively more time to higher-value words) as essential for maximising points, and hence selectivity may contribute to gains in earned points across trials. Besides adding a penalty group, we also used six study-test trials, so as to evaluate whether participants 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. Seventy-two word–value pairs (from the same pool used in Experiment 1) were used. Procedures in Experiment 2 were nearly identical to those of Experiment 1, with the following exceptions. First, the penalty group was additionally instructed that they would lose one point for each second of study. This reduced score was then shown to the penalty group after each study-test trial. As in Experiment 1, the no-penalty group was informed of their total score without point deductions for time spent studying. After reading the instructions, participants were required to answer a brief test on the instructions before beginning to study the words. Test questions included "Will you be permitted to study the material as long as you want?" If participants answered a question incorrectly, they were redirected back to the instruction screen, where they could find the correct answer before continuing. Second, participants had six study-test trials, with each trial involving 12 word–value pairs randomly selected (without replacement) from the list of 72 words.

Results and discussion

As in Experiment 1, for each variable, we present two analyses—the first analysis is collapsed across all point values, and the second analysis presents the dependent variable as a function of point values.

Points earned. We analyzed the number of points earned in two ways, with no penalty for each second used studying (as in Experiment 1) and with one point being deducted from the score for each second that elapsed during study. The former analysis is most appropriate for the no-penalty group, and the latter analysis is most appropriate for the penalty group. Nonetheless, for completeness, we present both measures for each group.

Total points earned without penalty. For the no-penalty group, total point scores for lists 1–6 (respectively) were 31.2 (SEM = 4.1), 40.9 (4.1), 38.0 (4.3), 39.7 (4.0), 40.7 (3.4) and 44.4 (3.8). For the penalty group, total point scores for lists 1–6 (respectively) were 20.8 (SEM = 3.1), 23.1 (2.8), 28.3 (3.1), 26.9 (2.9), 27.2 (3.1) and 30.4 (2.7). 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 (study group) × 6 (list) ANOVA revealed a main effect for study group, $F(1, 59) = 13.92$, MSE = 15,548.13, $p < .001$, $\eta^2 = .19$, indicating that when no points were deducted for study time in either group, more points were earned by the no-penalty group than by the penalty group. A main effect was also found for list, $F(5, 295) = 3.85$, MSE = 845.43, $p = .002$, $\eta^2 = .061$. To follow-up on the main effect of list, we conducted a priori comparisons on points earned on adjacent lists, collapsed across groups, because the list × study group interaction was not significant, $F(5, 295) = .57$, $p = .72$. Points earned increased from list 1 to list 2, $t(60) = 1.95$, $p = .03$, $d = .28$ (one-tailed test, given replication of gains in Experiment 1) but did not significantly increase on subsequent lists.

Total points earned with penalty. As shown in Figure 1, when the total points were adjusted for the penalty (subtracting one point per second during study), participants tended to earn more

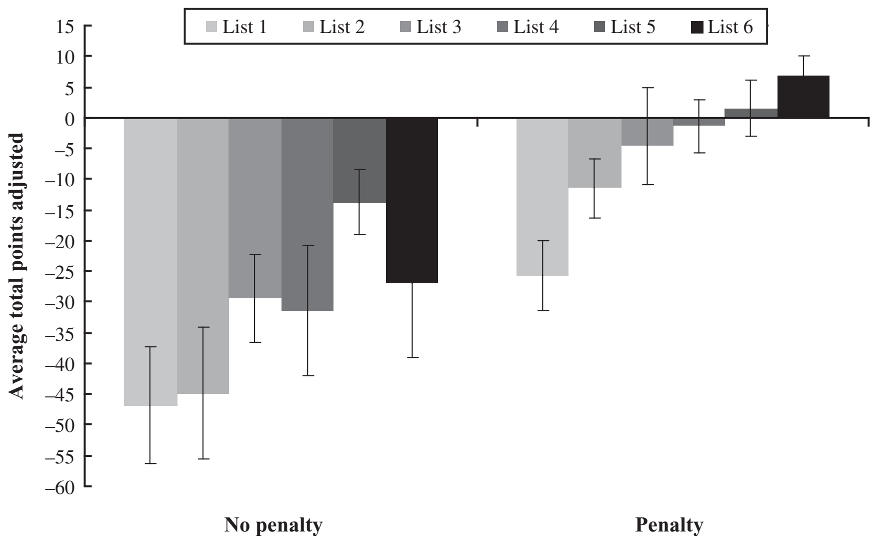

Figure 1. Mean total points after penalties for time spent studying (calculated for both groups). Error bars are standard errors of the mean.

points across lists. Consistent with these observations, a 2 (study group) × 6 (list) ANOVA for adjusted totals revealed a main effect of study group, $F(1, 59) = 9.86$, $MSE = 63{,}712.05$, $p = .003$, $\eta^2 = .14$, indicating that when participants in both groups were penalised for each second spent studying, more points were earned by the no-penalty group. A main effect was also found for list, $F(5, 295) = 10.08$, $MSE = 7798.01$, $p < .001$, $\eta^2 = .15$. To follow-up on the main effect of list, we conducted a priori comparisons on points earned on adjacent lists, collapsed across groups, because the list × study group interaction was not significant, $F(5, 295) = 1.06$, $p = .38$. Points earned increased from list 1 to list 2, $t(60) = 1.66$, $p = .05$, $d = .18$ (one-tailed test), from list 2 to list 3, $t(60) = 2.95$, $p = .01$, $d = .26$ and from list 4 to list 5, $t(60) = 2.17$, $p = .03$, $d = .26$ (two-tailed tests). It is noteworthy that totals adjusted for penalty typically improved across all lists.

Recall performance. One reason why scores may have improved is that participants recalled more words on later than earlier lists. To evaluate this possibility, mean recall performance was computed for these lists. For the no-penalty group, recall proportions for lists 1–6 (respectively) were .39 ($SEM = .05$), .51 (.05), .47 (.05), .50 (.05), .48 (.04) and .52 (.05). For the penalty group, recall proportions for lists 1–6 (respectively) were .27 ($SEM = .04$), .30 (.03), .36 (.04), .33 (.03), .33 (.03) and .37 (.03). A 2 (study group) × 6 (list) ANOVA revealed that recall was higher for the no-penalty than the penalty group, $F(1, 59) = 12.06$, $MSE = 2.17$, $p = .001$, $\eta^2 = .170$, and that recall differed across lists, $F(5, 295) = 2.74$, $MSE = .08$, $p = .019$, $\eta^2 = .044$. The list × study group interaction was not significant, $F(5, 295) = .63$, $p = .680$, so a priori comparisons on adjacent lists were collapsed across groups. Recall increased from list 1 to list 2, $t(60) = 2.05$, $p = .045$, $d = .28$ (two-tailed *t*-test). All other comparisons were not significant, *t*s < 1.44. Thus, it appears that one reason why learners earned more points was because they recalled more words across the first two lists.

Of course, another reason why learners earned more points could be that as compared to earlier lists, they also recalled more of the words with the higher point values on later lists. Because most of these differences occurred between lists 1 and 2, all further analyses focused on only these two lists. For completeness, values for all six lists as a function of points are presented in Table 1. A 2 (study group) × 2 (list) × 3 (points: low, medium, high) ANOVA was conducted. As expected from prior analyses, a main effect was found for group, $F(1, 59) = 10.45$, $MSE = 2.54$, $p = .002$, $\eta^2 = .15$, and for list, $F(1, 59) = 4.31$, $MSE = .49$, $p = .04$, $\eta^2 = .07$. No main effect was found for points, $F(2, 118) = 1.14$, $p = .32$, and no interactions were significant, all *F*s < 1.55.

Study time. As evident from inspection of Table 2, participants did not spend relatively more time on the most valuable words as they progressed across lists. Participants spent more time studying in the no-penalty group than in the penalty group, $F(1, 59) = 12.27$, $MSE = 1084.01$, $p = .001$, $\eta^2 = .17$. No main effect was found either for list, $F(1, 59) = .18$, $p = .67$, or for points, $F(2, 118) = .13$, $p = .88$, and no interactions were significant, all Fs < 3.82. Thus, in contrast to expectations, participants in the penalty group did not selectively encode any more than did participants who were not being penalised for each second spent studying.

Selectivity index. For the no-penalty group, SIs for lists 1–2 (respectively) were: −.08 ($SEM = .09$) and .16 (.10). For the penalty group, SIs for lists 1–2 (respectively) were: .14 ($SEM = .14$) and .04 (.09). No main effect was found for group, $F(1, 47) = .25$, $p = .62$, or for lists, $F(1, 47) = .95$, $p = .33$, and the list × study group interaction was not significant, $F(1, 47) = 2.03$, $p = .16$.

EXPERIMENT 3

In both Experiments 1 and 2, participants did earn more points across lists (and in particular, from list 1 to list 2), but what causes such increases in value-based learning? One factor is that overall recall improved across the first lists. Somewhat surprisingly, however, changes in the allocation of study time to higher-value words across lists did not appear to contribute to the increase in points earned. Perhaps participants will regulate more selectively if they are instructed to improve their scores by focusing on higher-value words across lists. Castel and his colleagues typically instruct participants to selectively focus on higher-value words during study; for instance, Castel, Humphreys, et al. (2011) instructed participants that "the goal of the task was to earn as many points as possible by remembering as many of the high point value words as they could" (p. 1556).

With these instructions, perhaps participants will selectively focus on higher-value words as they progress across lists, especially in the present context where they can pace their study. Alternatively, self-paced study of single words in this context may be largely driven by item difficulty (Koriat et al., 2014), and if so, study times may not be longer for higher-value words than lower-value words even when participants are instructed to focus on the higher-value words. We evaluated these predictions in Experiment 3 by having participants self-pace their study (without a penalty) either with instructions that we used in the prior experiments (i.e., "Your object is to maximise your score") or with value-emphasised instructions (i.e., "Your object is to maximise your score, and you can do this by remembering as many of the high point value items as you can").

Method

Participants. Sixty-one undergraduates from Kent State University participated for course credit in introductory psychology; they were randomly assigned to the standard group ($n = 39$) or the value-emphasised group ($n = 40$).

Materials and procedure. The materials and procedure for both groups were nearly identical to those used for the no-penalty group in Experiment 2, with one exception. Whereas the standard instruction group received the same standard set of instructions used in the prior experiments, the instructions for the value-emphasised group emphasised that participants should focus on higher-value words. In particular:

> In this experiment, you will be asked to study a series of words and recall them; each word has a different point value assigned to it, and your goal is to gain as many points as possible. After clicking "Continue Experiment," you will be presented with the words and their values. After 12 word-value pairs have been presented, you will work on a crossword puzzle for a predetermined length of time, and you will then be asked to recall as many words as you can. For every word you recall correctly, you will receive the point value originally paired with that word. After you study and recall a list, the computer will tell you how many points you earned, and then, you will move on to the next list. Additional points will be given for each completed word or number in the puzzle. This entire procedure will continue for 6 lists, so do your best to maximize your score.
>
> Importantly, you may study each word-value pair for as long as you wish before continuing to the next pair. You will be shown the word-value pair with a button underneath it called "Next Item." When you're finished studying a particular word-value pair, simply click on the "Next Item" button to move to the next item.

Your object is to maximize your score, and you can do this by remembering as many of the high point value items as you can. You will not receive any points or any penalties for incorrectly answering or for failing to recall a word.

Results and discussion

Points earned. For the standard instruction group, total point scores for lists 1–6 (respectively) were 30.4 (SEM = 3.7), 36.0 (3.1), 41.7 (3.3), 45.8 (3.4), 43.0 (3.7) and 39.4 (3.5). For the value-emphasised group, total point scores for lists 1–6 (respectively) were 34.1 (SEM = 4.2), 40.5 (4.1), 44.2 (3.6), 48.9 (3.1), 46.1 (3.0) and 40.4 (3.6). A 2 (study group) × 6 (list) ANOVA revealed no main effect of group, $F(1, 77) = .65$, $p = .421$, whereas a main effect was found for list, $F(5, 385) = 8.27$, $MSE = 2241.21$, $p < .001$, $\eta^2 = .097$. A priori pairwise comparisons were used to compare points earned on adjacent lists, collapsed across group, because the list × study group interaction was not significant, $F(5, 385) = .10$. Participants achieved significantly higher scores on list 2 than on list 1, $t(78) = 2.19$, $p = .015$, $d = .25$ (one tail, per replication of prior experiments), higher scores on list 3 than on list 2, $t(78) = 2.15$, $p = .03$, $d = .21$, and higher scores on list 4 than on list 3, $t(78) = 2.11$, $p = .04$, $d = .21$ (both two-tailed). After list 4, however, the trend tended towards the opposite direction; the difference between scores on list 5 and list 4 was not significant, $t < 1.25$, but scores were significantly lower on list 6 than on list 5, $t(78) = 2.17$, $p = .03$, $d = .27$.

Recall performance. For the standard instruction group, recall proportions for lists 1–6 (respectively) were .42 (SEM = .05), .53 (.05), .45 (.04), .59 (.04), .55 (.05) and .51 (.04). For the value-emphasised group, recall proportions for lists 1–6 (respectively) were .44 (SEM = .05), .53 (.05), .56 (.05), .60 (.04), .59 (.04) and .53 (.04). A 2 (study group) × 6 (list) repeated measures ANOVA did not reveal a main effect for group, $F(1, 77) = .54$, $p = .47$. A main effect was found for list, $F(5, 385) = 7.10$, $MSE = .29$, $p < .001$, $\eta^2 = .084$. The list × study group interaction was not significant, $F = .29$, $p = .92$. A priori pairwise comparisons were conducted to explore recall on adjacent lists, indicating that participants recalled more words on list 2 than on list 1, $t(78) = 1.78$, $p = .04$, $d = .20$ (one-tailed test, as per the replication of prior experiments). Participants also recalled more words on list 3 than on list 2, $t(78) = 2.28$, $p = .03$, $d = .24$.

To evaluate whether learners earned more points across lists because they recalled more higher-value words across lists (Table 1), we conducted a 2 (study group) × 6 (list) × 3 (points: low, medium, high) ANOVA. No main effect was found for group, $F(1, 77) = 0.54$, $p = .47$. A main effect was found for list, $F(5, 385) = 7.10$, $MSE = .86$, $p < .001$, $\eta^2 = .08$. No main effect was found for points, $F(2, 154) = .08$, and no interactions were significant, all Fs < .30.

Study time. While the improvement in points earned is not due to increased recall of higher-value (vs. lower-value) words, perhaps participants in the value-emphasised group attempted to selectively encode. A 2 (study group) × 6 (list) × 3 (points) ANOVA was conducted (Table 2). No main effect was found for group, $F(1, 77) = 0.10$, $p = .76$. A main effect was found for list, $F(5, 385) = .659$, $MSE = 330.33$, $p < .001$, $\eta^2 = .08$, More important, a main effect was found for points, $F(2, 154) = 4.0$, $MSE = 83.99$, $p = .020$, $\eta^2 = .05$. No interactions were significant, all Fs < 1.62. Concerning the main effect of list (collapsed across group), participants spent significantly less time studying words on list 2 ($M = 9.1$, $SE = .91$) than list 1 ($M = 10.4$, $SE = .99$), $t(78) = 2.56$, $p = .012$, $d = .16$ (two-tailed), but no other comparisons were significant. Concerning the main effect of points (collapsed across list), participants spent significantly less time studying lower-value words ($M = 8.00$, $SEM = .65$) than either medium-value words ($M = 8.61$, $SEM = .78$), $t(78) = 2.23$, $p = .03$, $d = .08$, or higher-value words ($M = 8.81$, $SEM = .71$), $t(78) = 2.39$, $p = .019$, $d = .13$.

Selectivity index. Mean SI values for lists 1–6 (respectively) for the standard instruction group were: $M = .09$ (SEM = .07), .09 (.08), .05 (.07), .22 (.07), −.02 (.07) and −.05 (.08). Mean SI values for lists 1–6 (respectively) for the value-emphasised group were: .15 (SEM = .10), .07 (.09), .04 (.08), .21 (.07), .00 (.07) and −.05 (.09). An ANOVA did not reveal a main effect for group, $F(1, 58) = .14$, $p = .71$, or for list, $F(5, 290) = 1.47$, $p = .20$, and the list × study group interaction was not significant, $F(5, 290) = 1.23$, $p = .29$.

EXPERIMENT 4

In Experiments 1–3, participants increased their total point scores across lists, primarily due to increases in recall performance rather than to selective encoding of higher-value words. In Experiment 4, we explored two other conditions that may promote selective encoding. First, participants may selectively encode if the point difference between higher-value and lower-value words are made more obvious by increasing the range of points. In support of this possibility, Dunlosky and Thiede (1998, Experiment 2) found that learners spent more time studying higher-value (64 pints) than lower-value word pairs (8 points), whereas participants did not selectively encode when the range was smaller (10 vs. 1; Experiment 1). To explore this possibility, we increased the range of point values by including 10 words with lower values (10 and below) and 10 words with much higher values (90 and above).

Second, learners may be more likely to selectively encode when they feel pressured to do so, such as when it is evident that not all words can be learned during a single-study trial. To evaluate this possibility, we included one group in which the experimenter-paced presentation rate for each word was speeded at 1 second/word. In this case, participants may elect to focus rehearsal on higher-value words and tend to ignore lower-value words as they are presented. If so, recall performance is expected to be higher for the higher-value than lower-value words and this selectivity in recall may contribute to gains in points across lists.

Method

Participants. Sixty undergraduates from Kent State University participated for course credit in introductory psychology; they were randomly assigned to either the experimenter-paced group (n = 30) or the self-paced group (n = 30).

Materials and procedure. The materials and procedure were nearly identical to those used in Experiment 1, with the following exceptions. First, both groups used the value-emphasised instructions from Experiment 3. Second, the presentation rate for the experimenter-paced group was 1 second per word–value pair. Finally, 20 word–value pairs were used in each list, and these pairs were randomly assigned a different point value within these two ranges: 1–10 and 90–100.

Results and discussion

Points earned. For the experimenter-paced group, mean total points earned were 159.47 (SE = 26.42) for list 1 and 217.87 (SE = 30.18) for list 2. For the self-paced group, mean total points earned were 261.50 (SE = 40.29) for list 1 and 392.83 (SE = 49.05) for list 2. A 2 (study group) × 2 (list) ANOVA revealed that more points were earned by the self-paced than experimenter-paced group, $F(1, 58) = 11.8$, $p = .001$, $\eta^2 = .17$, and that more points were earned on list 2 than on list 1, $F(1, 58) = 7.59$, $p = .008$, $\eta^2 = .12$. The list × study group interaction was not significant, $F(1, 58) = 1.12$, $p = .29$.

Recall performance. For the experimenter-paced group, the mean proportion of words recalled was .13 (SE = .02) for list 1 and .17 (SE = .03) for list 2. For the self-paced group, the mean proportion of words recalled was .27 (SE = .03) for list 1 and .35 (SE = .04) for list 2. Recall was greater for the self-paced group, $F(1, 58) = 28.55$, $p < .001$, $\eta^2 = .330$, and improved from list 1 to list 2, $F(1, 58) = 4.31$, $p = .042$, $\eta^2 = .069$. The list × study group interaction was not significant, $F(1, 58) = .43$, $p = .517$.

To evaluate whether differential recall of higher-value words contributed to the increase in point score between lists 1 and 2 (Table 1), a 2 (study group) × 2 (list) × 2 (points: low, high) ANOVA was conducted. A main effect was found for group, $F(1, 58) = 28.55$, $p < .001$, $\eta^2 = .33$, and for list, $F(1, 58) = 4.31$, $p = .04$, $\eta^2 = .07$. Most important, a main effect was found for points, $F(1, 59) = 10.07$, $p = .002$, $\eta^2 = .15$, with recall being higher for higher-value words than for lower-value words. The list × points interaction was significant, $F(1, 58) = 4.47$, $p = .04$, $\eta^2 = .07$, but no other interactions were significant, all Fs < 1.7. A priori pairwise comparisons (two-way) revealed that recall for higher-value words was greater than for lower-value words on list 2, $t(59) = 3.38$, $p = .001$, $d = .46$, but recall did not differ significantly between higher-value and lower-value words on list 1, $t(59) = 1.08$, $p = .29$. In summary, recall tended to increase across lists, and participants demonstrated greater recall for higher-value (vs. low value) words across the two lists.

Study time. Mean time (in seconds) spent studying words was 6.02 (SE = .70) for list 1 and 4.42 (SE = .46) for list 2, and mean time as a function of point value is presented in Table 2. A 2 (list) × 2 (points: low, high) ANOVA revealed a main effect for list, $F(1, 29) = 12.46$, $p = .001$, $\eta^2 = .30$, indicating that participants spent less time studying words on list 2 than on list 1. A main effect was found for points, $F(1, 29) = 4.94$, $p = .034$, $\eta^2 = .146$, indicating that participants spent more time studying higher-value words than lower-value words. The list × points interaction was not significant, $F(1, 29) = 1.45$, $p = .24$, which indicates that although participants were selectively encoding, they did not do so more strongly across lists.[1]

GENERAL DISCUSSION

The main goal of the present research was to evaluate whether students improve their performance (in terms of points earned) as they gain experience by selectively encoding higher-value words relative to lower-value ones. Across all experiments, students' performance (in terms of points earned) did improve with task experience (especially across the first two lists). Why did students earn more points across lists? Although the self-paced group in Experiment 1 did show increases in the SI from the first to second list, such increased selectivity did not replicate in either Experiments 2 or 3. Moreover, study time did not consistently favour higher-value words as compared to lower-value words, and even when study time did favour higher-value words (when the range of values was increased in Experiment 4), such selective encoding as indicated by self-paced study did not increase across lists. Taken together, this evidence is inconsistent with the encoding-selectivity hypothesis as an explanation for why learners earn more points across lists.

Although these outcomes may be viewed as surprising, they are consistent with theory that indicates self-paced study of individual words is largely driven by item difficulty (Koriat et al., 2014). One possibility is that participants study a particular item until they believe no more progress is being made with respect to improving memory (Metcalfe & Kornell, 2005; Thiede & Dunlosky, 1999), and hence the rate of progress itself is largely determined by word difficulty. Sometimes value does affect self-paced study, with people spending more time studying items of higher than lower value (e.g., Thiede & Dunlosky, 1999), but even here, the differences in study time are typically small and do not necessarily translate into differential recall. We found this same result in Experiment 4, where in spite of the large difference between higher-value and lower-value words, participants only studied higher-value words about 1 second longer than lower-value words. Moreover, this (minor) prioritisation of higher-value words did not increase across lists, again suggesting that selective encoding did not contribute to increases in participants' total point scores.

In contrast to self-paced study, however, learners do consistently use value (in terms of points rewarded for correct recall) when they *select* items for study (e.g., Ariel & Dunlosky, 2013; Ariel et al., 2009; Castel et al., 2013). In these cases, the decision concerning how to allocate study time is made separately from allocating time; that is, a participant first selects an item for study and then studies it sometime later. Item-selection tasks may allow learners to more fully consider possible ways to strategically allocate study, without the risk of attention being captured by the difficulty of individual to-be-learned items (as in self-paced study). Based on this rationale, directly comparing people's value-based learning within a selection task versus a self-paced study task should be informative with respect to understanding when value will have its largest effects on study decisions and overall performance.

Even though selective encoding did not contribute to performance gains, another possibility is that selective retrieval did; that is, perhaps across the first two lists, participants increased their score by changing the order in which they outputted higher- vs. lower-value items during each test. For instance, participants could output the higher-value words first, by virtue of their importance. Alternatively, assuming participants believed higher-value words were more memorable than lower-value words (e.g., Soderstrom & McCabe, 2011), they may have output lower-value words first (before losing access to them), leaving time to later recall the higher-value words. The latter outcome would be consistent with triaging theory, which predicts that with unlimited time for recall, materials are output

[1] Because Experiment 4 used values from discrete ranges of values (higher and lower), some estimates of the selectivity index would fall outside the interpretable range of −1 to +1 (Watkins & Bloom, 1999).

in a pattern of weak–strong–weak (Brainerd, Reyna, Howe, & Kevershan, 1990). Accordingly, we examined the average output order as a function of value for all four experiments. Unfortunately, most participants failed to output words from all three point value categories (low, medium and high), so omnibus inferential analyses did not produce interpretable outcomes. However, examination of mean output order (for these descriptive statistics, see the Appendix) suggests that the average output order is about the same for all item values, and no systematic changes in relative output order for higher- and lower-value items appears across lists. Therefore, selective retrieval also does not appear to contribute to improved performance across lists.

Selective encoding and selective retrieval appear to play a minimal role, and it appears that the gains in performance (in terms of points earned) largely arose from general improvements in recall performance. That is, regardless of point value, students consistently showed higher levels of recall as they gained task experience. This increase in recall has also been reported in other experiments focusing on value-based learning (Castel et al., 2009; Castel, Lee, Humphreys, & Moore, 2011). Note that in this prior work, increases in recall occurred with increases in the SI, so it is not evident whether increases in recall per se contributed to overall improvements in points earned. Fortuitously, several groups in the present research did increase their recall across lists, yet showed no increase in the SI. Evidence from these groups demonstrates that score and selectivity are not always related (and one can change without the other), providing relatively definitive evidence that one factor contributing to performance improvements is students' ability to improve their recall across lists. The improvement in recall across lists represents a practice effect whose constituent processes are beyond the scope of the current project. At least one factor, however, can be ruled out. Specifically, consider the overall study times presented in Table 2. In almost every condition, the magnitude of study times decreases (even if not significantly so) from the first to the second trial. Thus, merely spending more time studying words on the second (as compared to the first) list cannot explain the boost in recall performance. Perhaps students are developing better strategies to encode the words across lists or the improved recall reflects a task warm-up effect; these and other explanations for the current practice effects are worth exploring in future research.

Before concluding, we consider an outcome that could be viewed as inconsistent with conclusions from prior research. In the present experiments, value was not related to recall performance when study was experimenter-paced (see first two rows and last two rows of Table 1). By contrast, Castel, Humphreys, et al. (2011) have consistently found higher recall performance for higher-value than lower-value words. The present outcomes may represent a boundary condition for value-based effects that occurs when overall recall levels are low. For instance, average recall in Castel's research ranged from .25 to .59 (Castel et al., 2002, 2007, 2009; Castel, Humphreys, et al., 2011; McGillivray & Castel, 2011), whereas overall recall for the experimenter-paced groups in the present study were relatively low (range: .05 to .30). Even in the present context, however, a trend towards a value-based effect was evident, with three of the four conditions demonstrating a higher magnitude of recall for the higher-value words than for lower-value words. Thus, although the value-based effects on recall in the present groups were not significant, the trends are suggestive and may not necessarily be inconsistent with prior value-based research.

In conclusion, across four experiments, learners did increase the number of points earned across lists in multiple contexts, including experimenter-paced study, self-paced study with and without penalties, and regardless of whether instructions did (or did not) emphasise learning higher-value words. Such improvements in value-based learning across lists were largely isolated to learners' ability to learn how to improve recall and not to allocating extra time to higher-value words across lists.

REFERENCES

Ariel, R., & Dunlosky, J. (2013). When do learners shift from habitual to agenda-based processes when selecting items for study? *Memory & Cognition*, *41*, 416–428. doi:10.3758/s13421-012-0267-4

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override material-based monitoring. *Journal of Experimental Psychology: General*, *138*, 432–447. doi:10.1037/a0015928

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kevershan, J. (1990). The last shall be first: How memory strength affects children's retrieval.

Psychological Science, *1*, 247–252. doi:10.1111/j.1467-9280.1990.tb00208.x

Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225–270). London: Academic Press.

Castel, A. D., Balota, D. A., & McCabe, D. P. (2009). Memory efficiency and the strategic control of attention at encoding: Impairments of value-directed remembering in Alzheimer's disease. *Neuropsychology*, *23*, 297–306. doi:10.1037/a0014888

Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078–1085. doi:10.3758/BF03194325

Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition*, *35*, 689–700. doi:10.3758/BF03193307

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the life span: A cross-sectional study of memory and selectivity. *Developmental Psychology*, *47*, 1553–1564. doi:10.1037/a0025623

Castel, A. D., Lee, S. S., Humphreys, K. L., & Moore, A. N. (2011). Memory capacity, selective control, and value-directed remembering in children with and without attention-deficit/hyperactivity disorder (ADHD). *Neuropsychology*, *25*(1), 15–24. doi:10.1037/a0020298

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging*, *28*, 232–242. doi:10.1037/a0030678

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, *3*, 425–453. doi:10.1111/j.1745-6924.2008.00089.x

Dunlosky, J., & Ariel, R. (2011). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 103–140). San Diego, CA: Academic Press.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, *98*(1), 37–56. doi:10.1016/S0001-6918(97)00051-6

Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory & Cognition*, *32*, 779–788. doi:10.3758/BF03195868

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General*, *143*, 386–403. doi:10.1037/a0031768

Le Ny, J.-F., Denhière, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica*, *36*, 280–289. doi:10.1016/0001-6918(72)90012-1

McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement by younger and older adults. *Psychology and Aging*, *26*(1), 137–142. doi:10.1037/a0022681

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477. doi:10.1016/j.jml.2004.12.001

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 676–686. doi:10.1037/0278-7393.14.4.676

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*, 1236–1242. doi:10.1037/a0023548

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

Watkins, M. J., & Bloom, L. C. (1999). *Selectivity in memory: An exploration of willful control over the remembering process*. Unpublished manuscript.

APPENDIX

TABLE A1
Average output order as a function of value (low, medium, high) in Experiments 1, 2, 3 and 4 as a function of group

Value	*Low*	*Medium*	*High*	*Low*	*Medium*	*High*
Experiment 1						
	Experimenter-paced			Self-paced		
List						
1	2.3 (.29)	2.6 (.28)	2.3 (1.9)	3.4 (.37)	4.3 (.46)	4.1 (.50)
2	2.7 (.22)	2.6 (.27)	2.7 (.22)	3.1 (.24)	4.4 (.29)	4.8 (.46)
Experiment 2						
	No penalty			Penalty		
List						
1	3.3 (.34)	4.1 (.40)	3.5 (.36)	2.3 (.25)	3.1 (.52)	2.9 (2.6)
2	3.6 (.36)	4.0 (.33)	4.4 (.42)	2.9 (.32)	2.8 (.29)	2.6 (.29)
3	3.2 (.37)	3.9 (.46)	4.0 (.38)	2.8 (.38)	3.7 (.31)	2.7 (.29)
4	3.6 (.35)	4.3 (.32)	3.9 (.46)	2.7 (.34)	2.8 (.34)	2.6 (.22)
5	3.8 (.34)	3.3 (.31)	3.6 (.41)	3.2 (.40)	2.2 (.22)	3.2 (.28)
6	4.0 (.43)	3.5 (.41)	4.2 (.37)	3.4 (.35)	2.9 (.31)	2.8 (.27)
Experiment 3						
	Standard			Value-emphasised		
List						
1	3.6 (.33)	3.9 (.36)	3.9 (.36)	3.6 (.35)	4.2 (.41)	4.6 (.40)
2	3.3 (.30)	4.5 (.35)	4.8 (.49)	3.1 (.25)	3.4 (.30)	4.0 (.37)
3	4.3 (.31)	4.0 (.39)	4.3 (.32)	4.4 (.35)	3.8 (.33)	3.9 (.35)
4	3.8 (.27)	4.4 (.32)	4.7 (.33)	4.2 (.32)	4.4 (.32)	4.4 (.39)
5	4.0 (.27)	4.0 (.29)	4.8 (.29)	4.4 (.27)	4.3 (.36)	4.2 (.38)
6	3.9 (.34)	3.9 (.31)	4.1 (.32)	3.9 (.32)	4.0 (.34)	4.2 (.36)
Experiment 4						
	Experimenter-paced			Self-paced		
List						
1	2.0 (.21)	–[a]	2.4 (.18)	3.8 (.34)	–[a]	3.6 (.39)
2	3.0 (.34)	–[a]	2.5 (.21)	3.8 (.34)	–[a]	4.3 (.43)

[a]For Experiment 4, values were collapsed across 1–10 points (low) and 90–100 points (high).
Standard errors of the mean are in parentheses.
For Experiments 1–3, low, medium and high refer to the value awarded for correct recall, with values being collapsed across 1–4 points (low), 5–8 points (medium) and 9–12 points (high).