Psychology and Aging
2003, Vol. 18, No. 2, 340–345

0882-7974/03/$12.00 DOI: 10.1037/0882-7974.18.2.340

BRIEF REPORT

Training Monitoring Skills Improves Older Adults' Self-Paced Associative Learning

John Dunlosky and Alycia K. Kubat-Silman
University of North Carolina at Greensboro

Christopher Hertzog
Georgia Institute of Technology

We investigated a memory-enhancement program that involved teaching older adults to regulate study through self-testing. A regulation group was taught standard strategies along with self-testing techniques for identifying less well-learned items that could benefit from extra study. This group was compared with a strategy-control group, which was taught only strategies, and with a waiting-list control group. Greater training gains were shown for the regulation group (effect size, $d = 0.72$) than for the strategy-control ($d = 0.28$) and waiting-list control ($d = 0.03$) groups, indicating that training a monitoring skill—self-testing—can improve older adults' learning.

Age-related learning deficits are among the most widely studied phenomena in cognitive aging. Such deficits have been central to theorizing about how aging influences cognition and have provoked many attempts to improve older adults' learning. In a meta-analysis of the memory-improvement literature, Verhaeghen, Marcoen, and Goossens (1992) concluded that training older adults to use mnemonic strategies (e.g., imagery or the method of loci) can improve adults' learning of new materials. Even so, training improvements are not always found and occasionally are no larger for control groups who have received no training. Training also often shows minimal transfer (Stigsdotter Neely, 2000).

In this article, we introduce an intervention approach for improving older adults' learning that is based on theoretical frameworks of self-regulated learning. Self-regulated learning involves the interplay between metacognitive and cognitive processes as a means of adaptively guiding the learning process. The importance of metacognitive monitoring in particular is reflected by its presence in nearly every general theoretical approach to self-regulated learning, including information processing, social cognitive, Vygotskian, and constructivist approaches (Zimmerman, 2001). As Winne (2001) stated, "metacognitive monitoring is the pivot on which self-regulated learning turns because it creates opportunities to change tactics, to control how a task might be better dealt with" (p. 164). Thus, we trained older adults to accurately monitor learning and to use the output from it to achieve effective control of learning.

The use of monitoring may be easy for older adults to learn because monitoring accuracy appears to be relatively spared by aging (for a review, see Hertzog & Hultsch, 2000). In a typical study, participants have been asked to study paired-associate items, and for each one, to judge the likelihood of recalling the response when shown the stimulus on a subsequent test (called a *judgment of learning* [JOL]). Accuracy of monitoring is operationalized as the correlation between JOLs and later recall. Age has a negligible effect on these correlations, indicating that older and younger adults are equally accurate at discriminating between items that they will versus will not correctly recall (Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002).

For monitoring to assist in self-regulated learning, it must be sufficiently accurate to enable identification of items that have not already been learned so that they can receive additional study (Thiede, 1999). The premise here is based on metacognitive theories of self-regulated study, which assume that individuals use monitoring to control study time. In particular, a learner presumably uses his or her assessments of learning to decide whether (and how long) to restudy an item, and both younger and older adults often choose to focus restudy on those items that they perceive as not already being well-learned (Dunlosky & Connor, 1997). Inaccurate monitoring will result in ineffective regulation because individuals may choose to restudy material that is well-known and, even worse, they may not choose to restudy material that is poorly learned. In contrast, highly accurate monitoring enhances regulation of learning simply because it "shows you what you already know and what still demands close attention" (Woodworth, 1940, p. 334). Fortunately, the accuracy of monitoring associative learning can be substantial, with correlations between JOLs and crite-

John Dunlosky and Alycia K. Kubat-Silman, Psychology Department, University of North Carolina at Greensboro; Christopher Hertzog, Department of Psychology, Georgia Institute of Technology.

This research was supported by Grant R37 AG13148 from the National Institute on Aging, one of the National Institutes of Health. We greatly appreciate Robin West for sharing her expertise on how to successfully conduct a memory-training program. We are also indebted to Nancy Abrams for her help in training our participants and to Debbie Owen and Michelle Parker for orchestrating collection and entry of data. We also thank Greg Matvey for writing data collection and analyses programs and Katherine Rawson for helpful feedback on a draft of the article. More extensive description and analysis of this research is available at http://psychology.gatech.edu/CHertzog.

Correspondence concerning this article should be addressed to John Dunlosky, Psychology Department, P.O. Box 26164, University of North Carolina at Greensboro, Greensboro, North Carolina 27402-6164. E-mail: dunlosky@uncg.edu

rion performance typically exceeding +.80 when the JOLs are delayed briefly after study and are prompted by only the stimulus of an item (i.e., if "scream–canyon" were studied, "scream–?" would be used to prompt the JOL; e.g., Nelson & Dunlosky, 1991). Delayed JOLs produce high levels of accuracy because they are based on covert self-tests for responses from long-term memory, which are highly predictive of later success of retrieval.

Accordingly, our approach was to train older adults to accurately monitor their learning by relying on delayed self-tests (for further rationale behind this approach, see Dunlosky & Hertzog, 1998; and for evidence that links self-testing to student achievement, see Leal, 1987). More specifically, after initially studying a list of paired associates, older adults were trained to return to each one and conduct a self-test by covering the response. They were then instructed to regulate learning by restudying those items that were not retrieved during the self-tests. Albeit simple, this use of delayed self-testing may benefit older adults' learning of many real-world materials, and its simplicity may even contribute to its success and foster transfer to other kinds of learning.

Two potential constraints to the benefits of self-testing techniques merit discussion. First, one boundary condition was implied by the mechanism of self-testing described above. Namely, for self-testing to maximally benefit learning, one must have the opportunity to restudy after self-testing is used to identify those items that have not been well-learned. Thus, training adults to use self-testing is expected to have minimal benefits when the criterion tasks are experimenter paced, especially when they do not allow participants to restudy items. A second potential limitation is that older adults may already rely on delayed self-tests during study. If so, training this skill would be unnecessary. This possibility was investigated by Murphy, Schmitt, Caruso, and Sanders (1987), who observed older adults as they studied a list of words for a free-recall test. Although some older adults appeared to self-test, many did not. Moreover, when older adults were told to self-test prior to terminating study, not only were they more likely to do so but their learning also improved. Hence, Murphy et al. (1987) provided promising evidence that training older adults to self-test will benefit their associative learning.

To explore the efficacy of training older adults to regulate associative learning by self-testing, we conducted a training-based experiment. Older adults were assigned to one of three groups. A *waiting-list control* group received no training between the pretraining and posttraining tests. A *strategy-control* group was trained to use interactive imagery and sentence generation to learn paired associates.[1] The experimental *regulation* group was also trained to use these two mnemonics, but more critically, their training featured the use of delayed self-tests to regulate learning. By training both strategies and self-testing, we were able to assess whether self-testing skills boosted learning beyond any benefits obtained by training strategies alone. Training focused exclusively on learning paired associates, and hence the focal criterion task was associative learning. Nevertheless, training efficacy was also evaluated with list learning of individual words. This task was chosen because it involved the transfer of training to different materials (single words vs. word pairs) yet would still afford the benefits of self-testing (Murphy et al., 1987).

Experimenter-paced and self-paced versions of the tasks were also administered both prior to and after training. For the experimenter-paced tasks, participants were given a chance to study each item individually at a fixed presentation rate for one study trial. Such tasks are often used to assess learning in the aging literature, and more important, they were specifically selected to minimize the likelihood that participants could take advantage of self-testing. In contrast, the self-paced tasks allowed participants to study items up to 20 min and in any order, which afforded the use of regulation training. In particular, if they desired, the participants had the opportunity to use self-testing so as to focus restudy on "what still demands close attention" (Woodworth, 1940, p 334).

Several a priori predictions were evaluated by using planned comparisons. First, and of most interest, greater gains from the pretraining to posttraining tests were expected from the regulation group than either for the strategy-control group or for the waiting-list control group. A more specific prediction was available from theory of self-regulated learning described above: The benefits of regulation training were expected to be most evident on the self-paced associative-learning task and less so on the experimenter-paced task. Second, on the basis of the existing literature on memory training, lesser gains in performance were expected for the waiting-list control group than for either the strategy-control group or the regulation group, because both of the latter groups included strategy training. Finally, given the lack of transfer of training in previous studies, performance on list learning was not expected to differ between the three groups. Even with this pessimistic expectation in mind, one possible outcome was that in contrast to training mnemonic strategies, regulation training would show some transfer because self-testing is readily applicable to learning lists of single words.

Method

Participants

A total of 95 older adults between the ages of 65 and 85 years participated in this study. The regulation group consisted of 33 participants with a mean age of 68.4 years ($SD = 6.7$). The strategy-control group consisted of 31 participants with a mean age of 69.9 years ($SD = 6.1$). The waiting-list control group consisted of 31 participants with a mean age of 64.9 years ($SD = 7.5$). All participants were recruited through a newspaper advertisement describing a "Memory Improvement Training Program." The Short-Portable Mini-Screening Questionnaire (SPMSQ) was used to screen participants for dementia. Only 1 individual, who had more than two incorrect answers (which indicates at least mild intellectual impairment) was excluded from the study. A standard 4-min vocabulary test (36 items) from the Educational Testing Service Reference Kit was administered (Ekstrom, French, Harman, & Dermen, 1976). The mean vocabulary score of the participants was 19.96 ($SD = 7.88$), and there were no reliable differences between the groups on this measure ($F < 1.0$). Each adult was paid $30 for participating. Participants were randomly assigned to one of the three training groups (regulation, strategy control, and waiting-list control) and were not informed of that assignment.

Testing Materials and Test Procedures

Four memory tasks were administered during both the pretraining and posttraining tests. The tests involved studying pairs of either unrelated

[1] We used the term *control* in the label for the group that received strategy training to emphasize that it served as an important control for our theoretically motivated comparisons involving the training of self-testing skills. Nevertheless, to connect with previous intervention research, we also planned comparisons between the strategy-control and waiting-list control groups.

words or single words. The study trial was either experimenter paced or self-paced so that two versions of each kind of test were given. All items used in associative learning and list learning consisted of nouns taken from Paivio, Yuille, and Madigan's (1968) word norms and were not repeated in any of the tasks or the two testing sessions. Across tasks, we attempted to equate the single words (for list learning) and stimuli and responses (for associative learning) as closely as possible on key dimensions, such as imagery (range across tasks = 4.9–5.3, M = 5.20), frequency (range = 28–70, M = 51.00), and concreteness (range = 5.2–5.9, M = 5.40). Items and their characteristics (from Paivio et al., 1968) are available on request.

Associative learning. Participants were given 40 word pairs to study. In the experimenter-paced version of this task, pairs were individually presented at a fixed rate (5 s per pair) on a Macintosh computer using Hypercard programming software. After all pairs were presented, each stimulus was individually presented for the test, and participants were asked to type the response term that had been paired with each stimulus during study. Participants were given as much time as needed to complete the recall test. In the self-paced version of the task, participants studied a given word for as long as they wanted. Each word pair was printed in the center of a 5 × 7 index card. The 40 cards were handed to participants, who were instructed to study the words any way they chose. Twenty minutes were provided for study, and participants were told to notify the experimenter if they finished before the 20-min study period was done. We chose a 20-min limit (instead of unlimited study time) because results from a pilot study indicated that it provided enough time to restudy the items multiple times yet still ensured performance was off the ceiling. The recall test was the same as that used for the experimenter-paced version of the task.

List learning. Participants were given 40 single words to study for test. The experimenter-paced and self-paced versions of this task were identical to those used in the associative-learning task with two exceptions. First, only single words were presented for study. Second, participants were asked to recall the words by typing them into a blank screen.

Procedure

For the regulation and strategy-control groups, four sessions were conducted in the lab: two 2.5-hr testing sessions and two 2-hr training sessions. Each participant in the regulation and strategy-control groups received the same order of testing and training: pretraining tests, two training sessions separated by approximately a week, and posttraining tests. The waiting-list control group participated in the pretraining and posttraining sessions and then received regulation training after the posttraining tests. The mean lag between the two test sessions was 22.5 days for the regulation group, 25 days for the strategy-control group, and 17.7 days for the waiting-list control group. The mean amount of time between Day 2 of the training session and the posttesting session was 5.1 days for the regulation group and 7.5 days for the strategy-control group, $t(62) = 1.31$. Differences in lag were not related to final test performance on any of the memory tasks, $-.13 < r\text{s} < .03$.

After completing the SPMSQ phone interview, participants were mailed packets that included a consent form and a general demographics questionnaire. These measures were collected or completed immediately before the pretraining test session. During the pretraining test session, participants sat alone in a quiet room that had a computer (for experimenter-paced tasks and for recall), an extra table (to lay cards on during self-paced study), and a one-way mirror in front of the table. Participants were informed that a video camera was on the other side of the mirror and that they would be videotaped throughout their session.

For the self-paced versions of each task, the instructions were printed on individual sheets of paper and given to the participants to read. The participants were then given the index cards (with either paired associates or single words) for study. For the experimenter-paced version of each task, the instructions were presented to the participants on a computer screen. When the participants finished reading them, the experimenter started the computer program and left the participants alone until the participants were finished studying. For the recall tests for both the self-paced and experimenter-paced tasks (either paired-associate recall for the associative-learning task or free recall for the list-learning task), the experimenter instructed the participants to type as many responses as they could recall.

The order of the memory tests was identical for both the pretraining and posttraining sessions: self-paced associative learning, self-paced list learning, experimenter-paced associative learning, experimenter-paced list learning. This order was selected so that if participants performed more poorly than expected on the experimenter-paced tasks, such performance would not undermine participants' willingness to engage in self-paced learning.

Training Sessions

The 2-hr training sessions were conducted in a small classroom with no more than six adults in a single session. The training sessions were spaced 7 days apart. All training was done by one of two female trainers, with the same trainer being used for both training sessions for a single group. The training included a Microsoft PowerPoint presentation of how to use mnemonics alone (strategy-control group) or mnemonics and self-testing (regulation group). Two memory mnemonics were taught: sentence generation and interactive imagery. For sentence generation, participants were instructed to make a semantic link between the words in a word pair by making up a sentence by using the pair (e.g., for *scream–canyon*, a participant may generate the sentence, "The woman screamed loudly over the deep canyon"). For interactive imagery, participants were instructed to imagine an active picture of the two words of the pair together (e.g., a mental image of a woman screaming into a deep canyon). For the regulation group, training of self-testing was straightforward and intermixed with the training of the two mnemonics. In particular, participants were instructed to first generate a mnemonic for each pair. Afterward, they were trained to cover the response of each pair to elicit a self-test and then to set aside for restudy those items that they believed that they had not correctly recalled.

Both training groups received the same amount of practice with the paired associates during the two training sessions. During the training sessions, word pairs were presented on index cards, and participants practiced the mnemonics (and self-testing procedure for the regulation group) on these pairs. Initially, practice lists were short and gradually increased across the two training sessions. In particular, during the first session, they were given practice lists composed of 3, 10, 15, and 25 word pairs. During the second session, practice lists were composed of 5, 10, 25, and 40 word pairs. Immediately after studying each list, participants were given a paired-associate recall test for the pairs just studied. Participants were then invited to discuss their answers and the specific mnemonic (or self-testing) used on the items.

Participants in the training groups had 3 days of homework assignments between the two training sessions. The homework assignments were provided in a brief workbook that included reviews of mnemonics (and self-testing for the regulation group) that had been taught in the training session as well as directed practice with 5 sets of word pairs. Homework assignments were checked, and every participant had completed assignments as instructed.

Results

Criterion Test Performance

For each individual, we computed the proportion of correct responses for each of the criterion tests (Table 1). Because we derived a priori predictions from theory of self-regulated learning, planned comparisons were conducted to evaluate the training effects. Three planned comparisons were conducted for each of the four criterion tests (i.e., self-paced vs. experimenter-paced study × associative vs. list learning). Because the focal theory-based com-

Table 1
Performance on Criterion Tests for the Experimenter-Paced Tasks and for the Self-Paced Tasks

	Associative learning				List learning			
	Self-paced		Exp paced		Self-paced		Exp paced	
Group	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)
Waiting-list control								
Pretraining test	.57	(.27)	.18	(.18)	.61	(.29)	.17	(.12)
Posttraining test	.58	(.27)	.17	(.15)	.63	(.30)	.18	(.12)
Strategy control								
Pretraining test	.48	(.27)	.13	(.13)	.47	(.31)	.13	(.07)
Posttraining test	.56	(.31)	.14	(.17)	.52	(.31)	.14	(.08)
Regulation training								
Pretraining test	.44	(.30)	.11	(.11)	.51	(.32)	.11	(.07)
Posttraining test	.65	(.27)	.17	(.15)	.58	(.29)	.15	(.08)

Note. Participants in the waiting-list control group received no training between the pretraining and posttraining tests. See text for corresponding pre-to-post training effect sizes for each condition. Exp = experimenter.

parisons involved the regulation group, we first present the two comparisons involving this group. Finally, we evaluate whether strategy training alone was successful by comparing gains for the strategy-control and waiting-list control groups.

For these planned comparisons, gain scores were computed by subtracting the score on the pretraining test from the corresponding score on the posttraining test. Each of the three planned contrasts was then conducted on these gain scores, with the mean square error being based on scores from all participants. Thus, the best estimate of the population variance in training gains would be obtained from the sample and would be identical for all three contrasts. Each contrast yielded an F value and one-tailed p value to evaluate whether differential gains occurred in performance on the criterion task for the comparison groups. Effect sizes (d) for each group were also computed by dividing the difference between pretraining and posttraining scores for a given task by the pooled standard deviation of the pretraining scores for that task.[2]

Associative learning. As predicted from theory of self-regulated learning, when individuals self-paced associative learning, training regulation skills produced gains that were substantially greater than no training and strategy training alone. For the self-paced task, gains in performance were reliably greater for the regulation group ($d = 0.72$) than for the waiting-list control group, $F(1, 91) = 13.0$. Gains were also reliably greater for the regulation group than for the strategy-control group ($d = 0.28$), $F(1, 91) = 5.68$. The planned comparison between the strategy-control group and the waiting-list control group ($d = 0.03$) did not reveal reliably different gains, $F(1, 91) = 1.49$ ($MSE = 0.05$).

As evident from inspection of Table 1, gains in performance for the experimenter-paced associative-learning task were relatively minimal. Although the gains were reliably greater for the regulation group ($d = 0.55$) than for the waiting-list control group ($d = -0.09$), $F(1, 86) = 4.71$, they were not reliably greater for the regulation than the strategy-control group ($d = 0.09$), $F(1, 86) = 1.85$. The planned comparison for the strategy-control and waiting-list control groups did not reveal differential gains, $F(1, 86) = 0.59$ ($MSE = 0.01$).

List learning. For self-paced list learning, the benefits of training people to use strategies in learning paired associates were also negligible, with effect sizes being 0.06 for the waiting-list control group, 0.16 for the strategy-control group, and 0.23 for the regulation group. For all three planned comparisons, the groups did not show reliable differences in gains, $F\text{s} < 1.0$ ($MSE = 0.09$).

Training had a minimal influence on experimenter-paced list learning, with effect sizes being 0.10 for the waiting-list control group, 0.10 for the strategy-control group, and 0.40 for the regulation group. The three planned comparisons revealed no reliable differences in gains, all $F\text{s} < 1.20$ ($MSE = 0.01$). Thus, as in previous research, negligible transfer occurred.[3,4]

[2] We pooled the pretraining standard deviations only because they would not be influenced by any changes in variability due to testing or interventions. However, other methods to estimate effect sizes yielded the same conclusions.

[3] Although the comparisons involving the regulation group did not yield statistically reliable effects, inspection of the mean values for list learning revealed trends in interactions that are in the expected direction. Because these interactions would have significant implications for theory, we conducted power analyses to determine the number of participants required to obtain reliable outcomes. Assuming that the effect sizes would not be diminished with further data collection, to obtain reliable outcomes for the contrasts between the regulation and waiting-list control groups, we would require 107 (self-paced task) and 139 (experimenter-paced task) participants per group. The contrasts between the regulation and the strategy-control groups would require 3,054 (self-paced task) and 36 (experimenter-paced task) participants per group. Only the latter contrast may reach statistical significance with the addition of a reasonable number of participants. Note, however, that even if this contrast were statistically significant, the interaction itself may not hold much applied significance because the training gain in free recall for the regulation group was only 4%.

[4] In some conditions, the waiting-list control group obtained a higher level of pretraining performance than did the other two groups. These initial differences in pretraining performance (which were not expected given random assignment to groups) may have influenced the outcome of the planned comparisons involving the waiting-list control group. To explore this possibility, we conducted all the planned comparisons involving the waiting-list control group (for associative and list learning) again, using the performance on the corresponding pretraining tests as a covariate. In all cases, these new analyses yielded the same statistical and substantive conclusions as in the original planned comparisons. Another technique to control for initial differences in performance involves analyzing a subset of

Study Time

For the self-paced tasks, participants chose when to terminate study of each set of items. Given that regulation training may have improved associative learning by encouraging participants to merely study longer (Murphy et al., 1987), we compared the groups on overall study time, which was operationalized as the amount of time between when a participant first began to inspect items for study to when study was terminated either by the participant or by the experimenter (i.e., when the time limit had elapsed). We then computed the mean study time across participants. Because we had no theoretically motivated expectations about the effects of the various groups on overall study time, we did not conduct planned comparisons. For each task, we analyzed study time with a 3 (group) $\times$ 2 (time of test) analysis of variance (ANOVA).

For associative learning, mean study time for the regulation group was 19.5 min ($SEM = 0.5$) for the pretraining test and was 20.8 min ($SEM = 0.3$) for the posttraining test. For the strategy-control group, mean study time was 20.0 min ($SEM = 0.6$) for the pretraining test and was 20.7 min ($SEM = 0.3$) for the posttraining test. And for the waiting-list control group, mean study time was 19.5 min ($SEM = 0.7$) for the pretraining test and was 19.5 min ($SEM = 0.4$) for the posttraining test. The ANOVA revealed no reliable effects, Fs < 2.90 (MSEs < 8.5).

For list learning, mean study time for the regulation group was 19.0 min ($SEM = 0.7$) for the pretraining test and was 19.5 min ($SEM = 0.5$) for the posttraining test. For the strategy-control group, mean study time was 19.9 min ($SEM = 0.5$) for the pretraining test and was 19.8 min ($SEM = 0.6$) for the posttraining test. And for the waiting-list control group, mean study time was 17.4 min ($SEM = 0.9$) for the pretraining test and was 18.4 min ($SEM = 0.8$) for the posttraining test.[5] The ANOVA revealed a main effect for group, $F(2, 87) = 3.13$ ($MSE = 19.0$), which resulted from the slightly shorter time used by those in the waiting-list control group than by the other two groups. The main effect for time of test and the interaction were not reliable, Fs < 1.0 ($MSE = 11.24$).

Moreover, the Pearson correlation (across groups) between study time and performance on the posttraining tests was .02 for associative learning and was .20 for list learning, suggesting that any individual differences in the amount of time used for study contributed minimally to differences in test performance. These outcomes indicate that any differential gains between groups cannot be due to differences in the amount of time used to study items. Instead, how the regulation group used study time to control learning was responsible for their gains.

Discussion

Training gains in performance were substantially enhanced by supplementing standard mnemonics with self-testing, as long as older individuals were able to self-pace their study. This outcome is quite encouraging for applications, given that training older adults to self-test can be done easily and may be useful for studying many kinds of material. At least two factors may have contributed to the positive effects of self-testing. First, self-testing may have supported more effective use of study time. This possibility is based on the theory of self-regulated theory and does not necessarily entail that individuals who self-test will study longer than will those who do not guide learning by self-testing. In fact, the groups did not differ reliably with respect to the amount of time that was allocated to self-paced study. Instead, those who self-tested were trained to use the time more effectively by isolating less well-learned items for restudy. Second, self-testing itself may have enhanced memory performance (Bjork, 1988). When an item is correctly recalled during a practice test, its influence on a subsequent criterion test is greater than if the item is merely restudied (Modigliani, 1976). Given that both factors likely contribute to the effects of regulation training, a challenge for future research will be to estimate their relative contributions to training gains.

The benefits of regulation training were primarily evident on the self-paced tasks. Of course, training individuals on self-testing techniques was not expected to have a major impact on experimenter-paced tasks because self-regulatory behavior is largely undermined by standard experimenter-paced procedures. Nevertheless, given previous training research (Verhaeghen et al., 1992), we expected reliable gains on the experimenter-paced task of associative learning because both the regulation and strategy-control groups were trained to use effective strategies for learning associates. In contrast to this expectation, the mean effect size for the experimenter-paced task of associative learning was only about 0.32, which is smaller than the mean effect size (0.73) reported in a meta-analysis of the training literature (Verhaeghen et al., 1992).

Multiple explanations are available concerning why our strategy training did not substantively boost performance for the experimenter-paced tasks. First, performance on the experimenter-paced tasks was uniformly low, which may have artifactually constrained gains. Although possible, all mean values for pretest performance on these tasks were above the floor and still afforded much room for gains. A second explanation may be traced to a "file drawer" problem. Given that null training effects typically will not have theoretical significance, unsuccessful training studies will often not be reported in the literature. Censoring such null results would lead to inflated meta-analytic estimates of training effect sizes (Bradley & Gupta, 1997). Third, perhaps the amount of strategy training in the present intervention was not sufficient to

participants from each group who show similar levels of pretraining performance. For instance, for the critical comparison between the waiting-list control and regulation groups on self-paced associative learning (which yielded a reliable training effect), we matched groups by excluding participants with extreme scores on pretraining performance. After matching, pretraining performance was nearly identical for the regulation group ($M = 0.51$, $n = 18$) and for the waiting-list control group ($M = 0.52$, $n = 20$); by contrast, posttraining performance was still reliably greater for the regulation group ($M = 0.70$) than for waiting-list control group ($M = 0.52$), $F(1, 35) = 5.45$, $MSE = 0.03$, for the Group $\times$ Time of Test interaction. As with the covariate analyses described above, comparisons based on this matching technique did not change conclusions based on the original planned comparisons.

[5] Study times occasionally exceeded 20 min because in some cases the experimenter did not terminate study exactly at the 20-min mark (e.g., she was with another participant, etc.). In these cases, the extra time rarely exceeded a minute and was inconsequential to test performance.

produce large training gains. Although some mnemonics do require substantial time to train (e.g., method of Loci), the amount of time used in our program is arguably sufficient because the mnemonics trained were straightforward. Moreover, the amount of training was sufficient to demonstrate reliable gains for self-paced associative learning. Our final and preferred explanation is that because differential strategy production cannot adequately account for age-related deficits in associative learning (for references and exceptions, see Kausler, 1994), the potential training gains associated with producing a trained strategy are limited. For instance, Dunlosky and Hertzog (2001) reported that when older adults studied paired associates, they spontaneously used imagery and sentence generation, although not quite as often as did younger adults. Even these age deficits in the production of effective strategies accounted for a small proportion of the age effects on memory. Thus, training these strategies without self-testing may have had modest benefits because many older adults would have used them even prior to training.

Finally, consider performance on the transfer task. Even though list learning would afford the use of self-testing (Murphy et al., 1987), regulation training showed minimal transfer. Why? Self-testing may produce relatively minimal gains when participants do not also have effective mnemonic strategies to restudy items that are not initially well-learned. That is, even if individuals use self-testing to identify items that are currently unknown, such information will have limited value if they do not know how to learn those items. Free recall of unrelated words is enhanced when people organize words into meaningfully related chunks. This strategy was not included in our training. In the present program, the strategies (interactive imagery and sentence generation) were selected to improve participants' skill at associating word pairs, and they would apparently be less beneficial to tasks that are not obviously associative in nature. This hypothesis—that the benefits of self-testing partly rely on having strategies to learn difficult materials—yields testable predictions. For instance, the benefits of regulation training should be largest when it is combined with strategy training that is aimed specifically at improving performance on the criterion tasks. In the case of list learning, combining self-testing with an organizational strategy would be expected to produce considerable gains. Such predictions can be used to further explore the efficacy of regulation training.

In summary, standard strategy training appeared to yield modest training gains. By contrast, an intervention that involved training older adults to self-test yielded substantial performance gains (a 21% boost in performance, $d = 0.72$) when participants were given the opportunity to regulate their associative learning through self-paced study.

References

Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 396–401). New York: Wiley.

Bradley, M. T., & Gupta, R. D. (1997). Estimating the effect of the file drawer problem in meta-analysis. *Perceptual and Motor Skills, 85,* 719–722.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25,* 691–700.

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

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

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 417–466). Hillsdale, NJ: Erlbaum.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17,* 209–225.

Kausler, D. H. (1994). *Learning and memory in normal aging.* New York: Academic Press.

Leal, L. (1987). Investigation of the relation between metamemory and university students' examination performance. *Journal of Educational Psychology, 79,* 35–40.

Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning & Memory, 2,* 609–622.

Murphy, M. D., Schmitt, F. A., Caruso, M. J., & Sanders, R. E. (1987). Metamemory in older adults: The role of monitoring in serial recall. *Psychology and Aging, 2,* 331–339.

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

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76,* 1–25.

Stigsdotter Neely, A. (2000). Multifactorial memory training in normal aging: In search of memory improvement beyond the ordinary. In R. D. Hill, L. Bäckman, & A. Stigsdotter Neely (Eds.), *Cognitive rehabilitation in old age* (pp. 63–80). New York: Oxford University Press.

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multi-trial learning. *Psychonomic Bulletin & Review, 6,* 662–667.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging, 7,* 242–251.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 153–190). Hillsdale, NJ: Erlbaum.

Woodworth, R. S. (1940). *Psychology* (4th ed.). New York: Henry Holt.

Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 1–38). Hillsdale, NJ: Erlbaum.

Received December 18, 2001
Revision received August 22, 2002
Accepted August 27, 2002 ■