APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **19**: 361–374 (2005)
Published online 17 December 2004 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1083

The Effect of Overlearning on Long-Term Retention

DOUG ROHRER[1], KELLI TAYLOR[1], HAROLD PASHLER[2], JOHN T. WIXTED[2] and NICHOLAS J. CEPEDA[2]

[1]*University of South Florida, USA*
[2]*University of California, San Diego, USA*

SUMMARY

Once material has been learned to a criterion of one perfect trial, further study within the same session constitutes overlearning. Although overlearning is a popular learning strategy, its effect on long-term retention is unclear. In two experiments presented here, 218 college students learned geography facts (Experiment 1) or word definitions (Experiment 2). The degree of learning was manipulated and measured via multiple test-with-feedback trials, and participants returned for a final cued recall test between 1 and 9 weeks later. The overlearners recalled far more than the low learners at the 1-week test, but this difference decreased dramatically thereafter. These data suggest that overlearning (and its concomitant demand for additional study time) is an inefficient strategy for learning material for meaningfully long periods of time. Copyright © 2004 John Wiley & Sons, Ltd.

Study duration may be the most frequently manipulated variable in the field of memory. The ubiquity of this manipulation probably reflects its theoretical importance, but its practical implications are arguably even greater. Of particular concern is the relationship between study session duration and long-term retention. For example, after a student has learned the definitions for each of 10 vocabulary words, what are the benefits of continued study of the same material?

The immediate continuation of practice beyond the criterion of one perfect instance is defined as *overlearning*. Thus, if criterion is reached but further study is delayed until a subsequent session, the post-criterion practice is *not* an instance of overlearning. Another point to be clarified is the distinction between overlearning and the ultimate degree of mastery. As defined above, material is overlearned if it is learned with an overlearning strategy, regardless of how well the material was learned. Likewise, material can be extremely well learned without the use of an overlearning strategy. For example, virtually everyone has mastered the names of the 12 calendar months, but few have accomplished this by overlearning (i.e. a single study session with post-criterion learning). It is the *strategy* of overlearning that is assessed in the present studies rather than the utility of mastery *per se* (which, incidentally, does appear to boost long-term retention, e.g. Conway, Cohen, & Stanhope, 1992).

*Correspondence to: Dr D. Rohrer, Department of Psychology, PCD 4118G, University of South Florida, Tampa, FL 33620, USA. E-mail: drohrer@cas.usf.edu

Contract/grant sponsor: Institute of Education Sciences, US Department of Education; contract/grant number: R305H020061.

Overlearning occurs frequently in education and training, perhaps because it is commonly cited as a useful strategy in review chapters and textbooks regarding education and training (e.g. Aamodt, 1999; Fitts, 1965; Foriska, 1993; Hagman & Rose, 1983; Spector, 2000). Overlearning is generally described as a means of ensuring long-term retention. As Foriska (1993) explained, 'The information or skill to be learned is finally moved from short-term memory to long-term memory by overlearning the information . . . ' (p. 40). Similarly, Hall (1989) wrote, 'The overlearning effect would appear to have considerable practical value since continued practice on material already learned to a point of mastery can take place with a minimum of effort, and yet will prevent significant losses in retention' (p. 328). Fitts (1965) concluded, 'The importance of continuing practice beyond the point in time where some (often arbitrary) criterion is reached cannot be overemphasized' (p. 195). In summary, overlearning is widely advocated, and this advice is entirely consistent with empirical studies of overlearning.

THE EMPIRICAL LITERATURE

In the majority of previous overlearning studies, the data show that overlearning leads to greater recall than lesser degrees of learning. One such finding was presented by Krueger (1929), which may be the most frequently cited paper in this area. In that study, participants completed multiple learning trials with the same list of words, and the overlearning condition included twice as many learning trials as the control condition. This margin of overlearning is commonly referred to as 100% overlearning. When participants were tested between 1 and 28 days later, the overlearned material was better recalled than the non-overlearned material. This finding has since been replicated many times with varying procedures and different kinds of materials (e.g. Bromage & Mayer, 1986; Craig, Sternthal, & Olshan, 1972; Earhard, Fried, & Carlson, 1972; Gilbert, 1957; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982, Postman, 1962; Richardson, 1973; Rose, 1992). By contrast, we know of one study that revealed virtually no benefit of overlearning (Kratochwill, Demuth, & Conzemius, 1977).

A similar account of the literature is given by a meta-analysis by Driskell, Willis, and Cooper (1992). These authors examined 11 overlearning studies with cognitive tasks, and they found that the effect of overlearning on a subsequent test was moderate in size ($d = 0.753$). In brief, then, the results of previous overlearning studies have shown that overlearning can boost subsequent retention.

However, these studies leave open the possibility that the benefits of overlearning may not be long lasting. Of the 11 studies in the Driskell et al. meta-analysis, for instance, a majority relied solely on retention intervals of 1 week or less, and none included a retention interval greater than 4 weeks. Moreover, Driskell et al. observed that the difference between the overlearners and low learners was smaller at longer retention intervals (i.e. retention interval was a moderator). However, this finding is a correlation and therefore subject to alternative explanations. For instance, most of the 11 studies relied on a single retention interval, which means that this correlation may have partly reflected procedural differences between the studies with short retention intervals and the studies with longer retention intervals. Nevertheless, the results of the Driskell et al. meta-analysis are consistent with the possibility that the benefits of overlearning may dissipate at longer retention intervals.

In summary, it seems clear that overlearning can positively affect subsequent performance, but little is known about how this relationship is modulated by retention interval. An understanding of this relationship is essential to making rational decisions about the optimum degree of study when long-term retention is the goal. Toward this aim, the present studies included manipulations of learning level and retention interval (RI), and each study included a RI of at least 4 weeks.

CRITERION- VS. DURATION-BASED PROCEDURES

In overlearning studies, the degree of learning is manipulated by one of two procedures. In the criterion-based procedure, participants study or practice until they reach a criterion of one perfect trial before stopping or continuing. Thus, this method requires that participants' performance be measured during the learning session. For example, the learning of paired associates (e.g. Talara-Peru) is typically assessed by the use of multiple trials with test items (Talara - ?) followed immediately by feedback (Peru).

In the second type of procedure, the duration of study (or the number of learning trials) is predetermined for each degree of learning. For example, the overlearning condition might include 20 learning trials, whereas the control condition might include only 10 trials. This technique ideally includes a measure of performance during the learning session in order to determine, for example, whether the participants in the overlearning condition actually reached and surpassed criterion.

Each procedure has its advantages and disadvantages. The criterion-based procedure ensures the precise level of learning for all participants, but the analyses are complicated by the variability in the study duration *within* a condition (as some participants reach criterion more quickly than others). The duration-based procedure ensures an equal number of learning trials (or equal amount of study time) for all participants in the same condition, but it can be difficult for researchers to choose study durations that produce the desired degree of learning. Of the two procedures, the duration-based procedure has been used slightly more often than the criterion-based procedure.

OVERVIEW OF EXPERIMENTS

The two experiments presented here rely on the duration-based procedure, and we assessed performance during the learning session of each experiment. In addition, we conducted secondary analyses that incorporated the criterion-based definition of overlearning. These analyses essentially excluded participants who achieved too few or too many perfect trials.

In Experiment 1, participants studied paired associates (city-country pairs), and the number of learning trials was manipulated. In Experiment 2, participants studied paired associates (word-definition pairs), and the number of study words (but not total study duration) was varied, which effectively manipulated the amount of study time per word. In both experiments, RI was manipulated as well.

EXPERIMENT 1

The first study examined the effects of overlearning paired associates on a subsequent test. College students learned 10 city-country pairs (e.g. Talara-Peru) by test-with-feedback trials, with each trial including all 10 pairs. The number of learning trials equalled five or 20 (Lo vs. Hi Learning), and participants returned for a test after 1, 3, or 9 weeks.

Method

Participants

The sample included 130 undergraduates at the University of South Florida (USF). An additional eight students completed the learning session but failed to show for the test.

Design

Learning Level (Lo, Hi) and RI (1, 3, 9 weeks) were between-subjects variables, and each participant was assigned randomly to one of the six groups.

Materials

Participants studied the 10 city-country pairs (e.g. Talara-Peru) listed in the Appendix. We used real world knowledge rather than random word pairings, because we believed that participants would find the task more interesting and therefore exhibit greater motivation. In order to assess whether the participants may have known these city-country pairs prior to the experiment, we surveyed a different sample of 50 participants from the same population. They were asked to supply the country for each city (e.g. Talara-?) and were informed that each country name included five or fewer letters. Their accuracy averaged 1.4% (i.e. 7 of 500).

Procedure

During the learning session, each participant received a booklet with a different page devoted to each trial, with each trial including all 10 items. On the first trial, participants simply saw a list of 10 city-country pairs for 1 min. Immediately afterwards, the Lo and Hi Learners completed five or 20 test-with-feedback trials, respectively. For each of these 1-min trials, the booklet page included a column of the 10 cities. Participants were asked to write each city's corresponding country on a horizontal line immediately to the right of each city name. The order of the cities varied randomly across trials to ensure that participants were not merely learning serial position. Handwriting time was minimized by the use of countries with five or fewer letters and by encouraging participants to write in cursive. After 50 s, participants were prompted to unfold the right side of the page, which revealed each correct answer in a location directly to the right of the corresponding written response. Participants then studied for the remaining 10 s of the trial before turning the page and beginning the next trial. After the fifth test-with-feedback trial, all participants ceased studying and the Lo Learners departed the room. A few minutes later, the Hi Learners completed the remainder of their 20 learning trials.

One, 3, or 9 weeks later, participants returned for the test. Each received a page listing the 10 cities and was asked to recall the corresponding countries in 3 min.

Results and discussion

Learning session

Virtually all Hi Learners but only a minority of Lo Learners produced more than one perfect learning trial. Specifically, 58 of the 63 Hi Learners produced at least three perfect trials, and this subgroup was dubbed the True Hi Learners. By contrast, 44 of the 67 Lo

Learners produced no more than one perfect trial, and this subgroup was dubbed the True Lo Learners. All subsequent analyses were performed twice: once for the Hi vs. Lo Learners and once for the True Hi vs. True Lo Learners.

Accuracy during the learning session improved with successive trials, of course, as illustrated in Figure 1A. The Hi Learners' and Lo Learners' curves coincide during the first five trials, because these two groups undergo identical procedures during this period. By contrast, there is no such overlap between the True Hi and True Lo Learners despite identical procedures during the first five trials, because these two groups were not selected randomly. That is, the True Hi Learners represent the best performing Hi Learners, and the True Lo Learners comprise the worst performing Lo Learners. Incidentally, the dip in the curve after trial five for both the Hi and True Hi Learners reflects a brief rest period, as described in the procedure section.

Figure 1A also illustrates that both the Hi and True Hi Learners neared the ceiling well before the end of the learning session, whereas their counterparts failed to reach asymptote. On the last learning trial, the Hi Learners averaged 97%, and the Lo Learners averaged 85%, F (1, 128) = 15.53, $p < 0.001$. Likewise, the True Hi Learners averaged

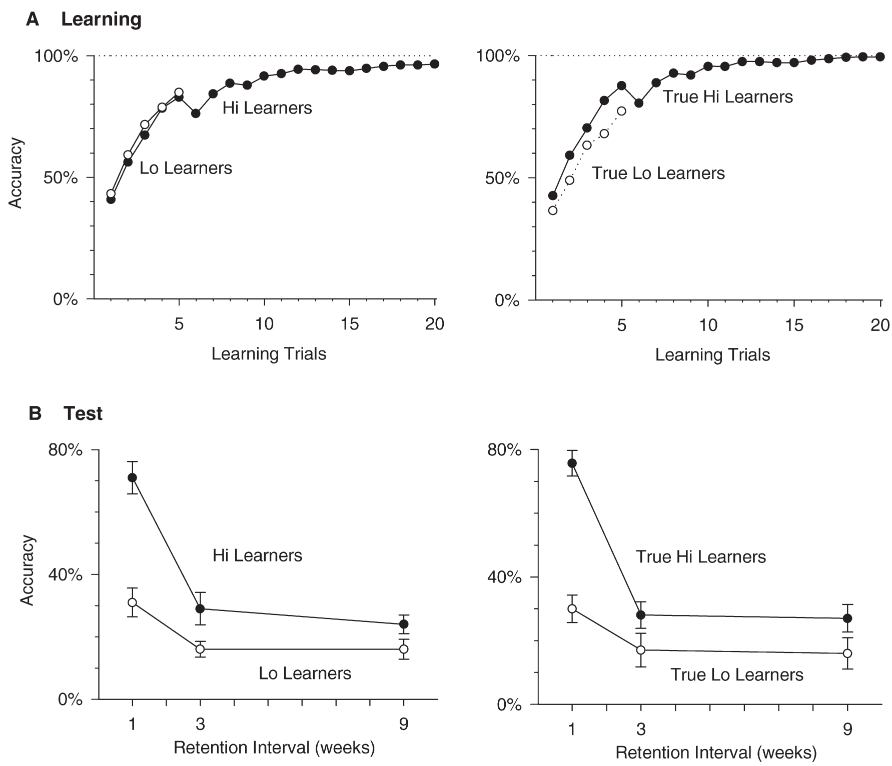

Figure 1. Experiment 1. The Hi Learners ($n = 63$) who achieved at least three perfect learning trials were dubbed the True Hi Learners ($n = 58$). The Lo Learners ($n = 67$) who achieved no more than one perfect trial were the True Lo Learners ($n = 44$). (A) The Hi Learners and Lo Learners underwent the same procedure during the first five trials. The dip in the Hi Learners data at Trial 6 reflects a brief rest after Trial 5. (B) For both graphs, both main effects and the interaction are statistically significant. Error bars indicate ± 1 standard error

99% on their last learning trial, while the True Lo Learners averaged only 77%, F (1, 100) = 68.729, $p < 0.001$.

Test session

Test performance is shown in Figure 1B. The Hi Learners recalled more than the Lo Learners, as indicated by a significant main effect of learning level given by a two-factor analysis of variance, F (1, 124) = 33.29, $p < 0.001$. Likewise, the main effect of RI was also reliable, F (2, 124) = 34.78, $p < 0.001$. More importantly, the difference between the Hi and Lo Learners declined with RI, as evidenced by a significant interaction between learning level and RI, F (2, 124) = 8.65, $p < 0.001$. This declining difference between the Hi and Lo Learners was further evidenced by Holm-Sidak multiple comparison tests (with alpha = 0.05) showing that the Hi and Lo Learners differed significantly at week 1 ($t = 6.85$) and week 3 ($t = 2.03$) but not week 9 ($t = 1.26$).

This convergence of retention curves was also observed for the True Hi and True Lo Learners, as shown in Figure 1B. Specifically, a two-factor analysis of variance revealed statistical significance for learning level, F (1, 96) = 37.19, $p < 0.001$, RI, F (2, 96) = 34.26, $p < 0.001$, and the interaction, F (2, 96) = 10.95, $p < 0.001$. The Holm-Sidak multiple comparison tests revealed significance differences between Hi and Lo Learners at week 1 ($t = 7.82$) but not week 3 ($t = 1.67$) or week 9 ($t = 1.58$).

Finally, we assessed whether the interaction of retention curves shown in Figure 1B may partly reflect a floor effect on the Lo Learners at the longer RIs. To examine this possibility, we tabulated the number of participants who recalled zero items at test. Although zero scores were rare and equally frequent among Hi and Lo Learners at weeks 1 and 3, there were more zero scores among the Lo Learners (six of 23) than Hi Learners (two of 20) at week 9. To examine whether these zero scores contributed to the findings described above, we analysed the test data shown in Figure 1B without the data from the 9-week test. Specifically, a two-factor analysis of variance revealed statistical significance for learning level, $F(1, 83) = 31.56$, $p < 0.001$, RI, $F(1, 83) = 38.81$, $p < 0.001$, and the interaction between learning level and RI, $F(1, 83) = 8.58$, $p < 0.01$. Likewise, when the same analysis was restricted to the True Hi and True Lo Learners, statistical significance was again observed for learning level, $F(1, 66) = 32.78$, $p < 0.001$, RI, $F(1, 66) = 37.81$, $p < 0.001$, and their interaction, $F(1, 66) = 12.02$, $p < 0.001$.

Summary

The Hi Learners recalled more than the Lo Learners at each RI, but the Hi Learners' retention declined at a greater rate than the Lo Learners' retention, as indicated by the interaction in Figure 1B. Moreover, the Hi Learners' retention declined by a greater *proportion* as well. Specifically, the Hi Learners' retention declined by about two-thirds (from 70% at week 1 to 24% at week 9), while the Lo Learners' retention declined by less than one half during the same period (from 31% at week 1 to 17% at week 9).

A similar decline in overlearning benefits emerged in the analyses restricted to the higher achieving Hi Learners (or True Hi Learners) and the lesser achieving Lo Learners (or True Lo Learners). Although these two subgroups probably differed in their ability and motivation, this confound presumably worked against the difference between the two groups (by magnifying the observed difference between the two groups). Also, a majority of both the Lo and True Lo Learners failed to achieve even one perfect learning trial, which is typically the criterion for the control condition in overlearning studies. Thus, the inclusion of these 'less-than-criterion learners' presumably led to an observed difference that further overestimated the true benefit of studying after reaching criterion.

EXPERIMENT 2

Whereas the first experiment included a manipulation of total study time, the second experiment induced overlearning by varying the amount of learning material (and not the total study time). Specifically, the Lo Learners studied 20 word-definition pairs, and the Hi Learners were given the same amount of time to learn just 10 word-definition pairs. Thus, the Hi Learners received twice as much study time *per item*, and this was enough to produce overlearning. The learning session was again comprised of test-with-feedback trials, and participants were tested either 1 or 4 weeks after learning.

Method

Participants
A total of 88 USF undergraduates completed the experiment. An additional 13 students completed the learning session but did not return for the test.

Design
Learning Level (Lo or Hi) and RI (1 or 4 weeks) were between-subjects variables, and each participant was randomly assigned to one of the four groups.

Materials
Participants studied the word-definition pairs (e.g. cicatrix-scar) listed in the Appendix. Each definition was a single word comprised of four or fewer letters in order to minimize participants' writing time. The Lo Learners studied all 20 pairs, and the Hi Learners studied a random subset of 10 pairs. The base rate knowledge of these vocabulary words was assessed through a survey of the 50 different USF students drawn from the same sample. None was able to provide a correct definition or synonym for any of the words.

Procedure
The procedure was identical to that of Experiment 1 with the following exceptions. The initial study-only trial was 2 min. All participants completed 20 test-with-feedback 'sub-trials.' Each sub-trial lasted 30 s and included five words. For the Hi Learners (who studied the 10-word list), the 20 sub-trials were grouped into 10 trials, with each trial spanning two consecutive five-word sub-trials so that each word appeared once in every trial. Likewise, the Lo Learners (who studied the 20-word list) completed five trials, with each trial spanning four sub-trials so that each word appeared just once in every trial. The order of the words varied across trials.

One or 4 weeks later, participants returned for the test. They were given 4 min to provide the one-word definition for each study word (e.g. cicatrix - ?).

Results and discussion

Learning session
Most Hi Learners and virtually no Lo Learners produced more than one perfect trial. Specifically, 40 of the 46 Hi Learners achieved at least three perfect trials, and this

subgroup was dubbed the True Hi Learners. By contrast, 41 of the 42 Lo Learners produced no more than one perfect trial, and this subgroup was dubbed the True Lo Learners. Consequently, the True Hi and True Lo Learners included all but seven of 88 participants, thereby ensuring that the differences between the Hi and Lo Learners were virtually identical to the differences between the True Hi and True Lo Learners.

Across learning trials, the Hi Learners performed much better than the Lo Learners, as shown in Figure 2A. On the last trial, the Hi Learners averaged 97%, and the Lo Learners averaged 70%, $F(1, 86) = 50.12$, $p < 0.001$. Likewise, the True Hi Learners averaged 98% on their last learning trial, compared to 70% for the True Lo Learners, $F(1, 79) = 54.59$, $p < 0.001$.

Test session

The Hi Learners, who received twice as much study time per word as the Lo Learners, recalled a greater proportion of study words on the 1-week test. But this difference disappeared by week 4, as shown in Figure 2B. A two-factor analysis of variance revealed a main effect of Learning Level, $F(1, 84) = 12.21$, $p < 0.001$, a main effect of RI, F (1,

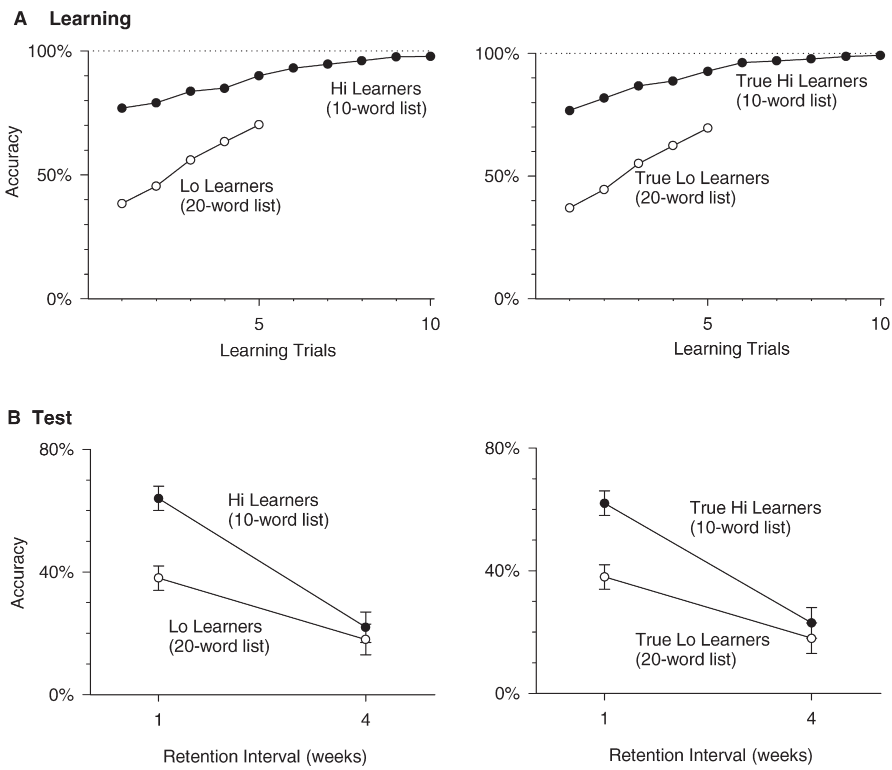

Figure 2. Experiment 2. The Hi Learners ($n = 46$) who achieved at least three perfect learning trials were dubbed the True Hi Learners ($n = 40$). The Lo Learners ($n = 42$) who achieved no more than one perfect trial were the True Lo Learners ($n = 41$). (A) The Hi Learners (and True Hi Learners) completed 10 learning trials with a 10-word list, whereas the Lo Learners (and True Lo Learners) completed five learning trials with a 20-word list. Total study time did not differ. (B) For both graphs, both main effects and the interaction are statistically significant. Error bars indicate ± 1 standard error

84) = 49.54, $p < 0.001$, and a reliable interaction between learning level and RI, F (1, 84) = 6.15, $p < 0.02$. This convergence of retention curves was again illustrated by Holm-Sidak multiple comparison tests (with alpha = 0.05) showing that the Hi and Lo Learners differed significantly at week 1 ($t = 4.42$) but not at week 4 ($t = 0.69$).

The same pattern emerged when the analyses were restricted to the True Hi and True Lo Learners, as shown in Figure 2B. This is not surprising, because these two groups included virtually the same participants as the Hi and Lo Learners. The statistical analyses produced the same findings: a main effect of learning level, F (1, 77) = 10.20, $p < 0.002$, a main effect of RI, F (1, 77) = 39.76, $p < 0.001$, and an interaction, F (1, 77) = 4.07, $p < 0.05$. Likewise, the Holm-Sidak tests revealed significance differences between Hi and Lo Learners at week 1 ($t = 3.86$) but not at week 4 ($t = 0.80$).

The convergence of the retention curves shown in both panels of Figure 2B raises the possibility that this interaction is due to a floor effect that disproportionately affects the Lo Learners at the long RI. Although it is difficult to strictly rule out this possibility, this rival hypothesis is at odds with an analysis of the individual data. Specifically, at the long RI, the number of Lo Learners who recalled zero items was relatively small (two of 20) and actually less than the number of Hi Learners who recalled nothing (five of 20). Incidentally, there were no test scores of zero at the short RI.

Although the Hi and Lo Learners recalled the same *proportion* of words on the 4-week test, the Lo Learners recalled a greater *absolute* number of words at both RIs. The conversion from proportions to absolute totals is straightforward. For the Hi Learners, who studied 10 words, their 1- and 4-week proportions of 64% and 22% convert to 6.4 and 2.2 words, respectively. For the Lo Learners, who studied 20 words, their 1- and 4-week proportions of 38% and 18% convert to 7.5 and 3.5 words (after rounding). Thus, the Lo Learners recalled *more* words, on average, than the Hi Learners at both the 1-week RI (7.5 vs. 6.4) and the 4-week RI (3.5 vs. 2.2). Hence, the margin of difference equalled 1.3 words at each RI (coincidentally), meaning that there was no interaction like that observed with the proportional measures (Figure 2B). However, significance was obtained for Learning Level, F (1, 84) = 4.14, $p < 0.05$, and RI, F (1, 84) = 48.19, $p < 0.001$. The same pattern was observed for the absolute test scores of the True Hi and True Lo Learners: a main effect of learning level, F (1, 77) = 4.27, $p < 0.05$, a main effect of RI, F (1, 77) = 38.26, $p < 0.001$, and no interaction, F (1, 77) < 1.

Summary

Participants either overlearned a 10-word list or underlearned a 20-word list. Overlearning boosted the chances of each item being recalled 1 week after learning, but no such benefit remained at week 4. In addition to this convergence of retention curves, the Hi Learners forgot at a greater proportional rate. Specifically, the Hi Learners' retention dropped by about two thirds (from 64% at week 1 to 22% at week 4), while the Lo Learners' retention dropped by about one half during the same time period (from 38% to 18%). Finally, with respect to the total number of words recalled, the underlearning of 20 words proved superior to the overlearning of 10 words at both RIs.

GENERAL DISCUSSION

We presented two experiments that assessed the effect of overlearning on the long-term retention of paired associate learning. In Experiment 1, college students learned 10 paired associates, with one group of participants studying about four times as much as the other. This additional study effort produced overlearning, as evidenced by multiple perfect learning trials, which in turn boosted their subsequent test performance beyond that of their more moderately learning cohorts. Yet the size of the benefit declined sharply within several weeks of the learning session, as shown in Figure 1B. In Experiment 2, college students studied 10 or 20 paired associates, with the same study time allotted for either list. The 10-item group was able to achieve overlearning, as evidenced by multiple perfect learning trials, while the 20-item group was unable to reach a criterion of one perfect trial. The overlearned 10 items were more often recalled than the poorly learned 20 items when participants were tested 1 week after learning, but this benefit disappeared by 4 weeks, as shown in Figure 2B. Hence, in both experiments, overlearning boosted subsequent cued recall performance in comparison to a lesser degree of learning, but this benefit dissipated within several weeks.

The drawbacks of overlearning

These findings have implications for anyone who wants to retain information for at least several weeks, because the data cast doubt on the efficiency of overlearning as a strategy for achieving long-term retention of paired associates. Thus, these data are at odds with the pedagogical and empirical literature cited in the introduction.

The question that arises, then, is why most previous overlearning studies have found benefits of overlearning that exceed those observed in the present studies. The discrepancy appears to be at least partly due to three procedural differences. First, most previous overlearning studies relied on RIs that were far shorter than those in the present studies. For instance, of the papers included in the Driskell et al. (1992) meta-analysis described in the introduction, nine of 11 relied solely on RIs of 1 week or less. That these studies would find strong support for overlearning is neither surprising nor inconsistent with the results presented here. Second, the majority of previous overlearning studies included a single RI, and this prevents the possibility of observing the declining benefits of overlearning that is indicated by interactions like those shown in Figures 1B and 2B. In Driskell et al., seven of the 11 papers included only one RI. Third, and less significantly, a small number of purported overlearning studies required participants to delay their post-criterion study to a second session on a later day, whereas overlearning is usually defined as immediate post-criterion study. By delaying their post-criterion practice, these learners were able to exploit the benefit of the spacing effect, which is described in greater detail further below. This was the case in two of the studies in the Driskell et al. meta-analysis (Ausubel & Youssef, 1965; Ausubel, Stager, & Gaite, 1968), and each of these studied boosted the overall effect size given by the meta-analysis.

The benefits of overlearning

Although an overlearning strategy may have limitations, we should emphasize that this learning strategy may be advisable in certain instances. For example, because overlearning proved extremely beneficial at the shortest RIs in the present studies, it might be ideal for learners who seek only short-term retention. For instance, a student with an exam later in the day might benefit from overlearning, as would anyone trying to amass foreign language vocabulary just before a brief trip. Furthermore, overlearning would be appropriate when there are dire consequences of forgetting. For example, if an employee must know certain safety procedures, overlearning might be advised. Similarly, Schendel and Hagman (1982) advocate an overlearning strategy for soldiers learning disassembly and assembly of machine guns, and the present results would not challenge that advice.

A final set of caveats concern the potentially limited generality of the present results. In Experiments 1 and 2, for instance, participants learned paired associates, and it is thus possible that overlearning may provide better long-term retention for different kinds of tasks. For instance, it remains unknown whether overlearning is advisable for the long-term retention of more abstract skills such as mathematical procedures. Likewise, overlearning may provide longer lasting benefits when the material lends itself to different kinds of encoding strategies. For example, strategies such as visual imagery or deep processing are likely precluded by the use of city-country pairs and word-definition pairs. Finally, it is possible that overlearning would have proven more useful at longer RIs if retention had been assessed by a free recall test or a recognition test.

Maximizing recall total

Although the studies presented here focused on the proportion of items recalled, the measure of recall total has more practical significance in some situations. When enriching one's vocabulary, for instance, the total number of words learned is arguably more important than the proportion of studied items that were forgotten. For example, rather than study 25 vocabulary words and later recall 15, a student who devoted the same total study time to 50 words might later recall 20. Here, then, the longer study list provides a greater boost in vocabulary.

This hypothetical example is analogous to the results of Experiment 2. The 20-word learners were able to recall more words than the 10-word learners even though total study duration did not vary. Hence, although students are often required to master relatively short lists of vocabulary words, the use of much longer vocabulary lists (and a concomitant reduction in expected accuracy) might increase students' vocabulary by a greater number of words. Such a strategy would also be appropriate during the preparation for a standardized test that includes vocabulary words, because these words are usually drawn from a list of thousands.

This raises the question of how many items one should study during a session of a given duration in order to maximize recall total, which is equivalent to determining the optimal study time per item. In Experiment 2, recall total was greater for the 20-word learners (who averaged 36 s per word) than for the 10-word learners (who averaged 66 s per word). Here, then, it was better to study more words (and devote less time to each word). However, we surmise that the optimal study duration per item is not necessarily as brief as possible. In a compilation of data from Experiments 1 and 2 and a number of additional unpublished studies in our laboratory, recall total was an inverted U-function of study time per item, meaning that recall total was optimized by an intermediate value of study time per item (when total study duration remains constant). For a 1-week RI, for example, this optimum was about 30 s per item. Thus, by this preliminary analysis, if students wish to increase their vocabulary in preparation for a standardized test that is to be taken in 1 week, the number of study words should be that which produces an average of about 30 s of study time per word (although this duration is probably best distributed across multiple presentations). Of course, this value would most likely vary with the kind of material, the amount of material, and the type of test.

The benefits of distributed practice

While the data presented here suggest that the benefits of overlearning diminish with time, we are certainly not advocating that people should avoid large amounts of practice to achieve long-term retention. Instead, we suggest that post-criterion practice be delayed until a later day in order to reap the benefits of the so-called spacing effect. Specifically, by distributing (or spacing) practice across multiple sessions rather than concentrating (or massing) the same amount of practice into one session, long-term retention is often boosted dramatically (Austin, 1921; Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bjork, 1979; Bloom & Schuell, 1981; Cull, 2000; Dempster, 1989; Earhard & Landry, 1976; Glenberg & Lehman, 1980; Greene, 1989; Reynolds & Glaser, 1964; Schmidt & Bjork, 1992; Shaughnessy, 1977).

Unfortunately, the strategy of distributed practice is underutilized. For example, the majority of vocabulary words presented in foreign language textbooks appear in one chapter rather than multiple chapters. Likewise, for each daily lesson within many mathematics textbooks, the majority of the following exercises concern that day's lesson. The spacing effect also appears to be underappreciated by cognitive psychologists, as it is mentioned in only three of the 12 introductory cognitive psychology textbooks belonging to the first author.

In fact, the paucity of spacing in educational curriculum may partly reflect the popularity of overlearning, as the two strategies often conflict in applied settings. That is, for a study session of a given duration, the overlearning of any particular skill leaves less time for the review of previously learned skills. Thus, every additional math problem done for the sake of overlearning is one less problem devoted to the principle of spaced practice. Hopefully, an appreciation of the severe limitations of overlearning will encourage a greater reliance on the productive principle of spaced learning.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Institute of Education Sciences, US Department of Education (R305H020061). We thank Erica Porch for her assistance with data collection, and Glen Bodner, Robert Greene, and J. Don Read for their helpful comments regarding an earlier draft of this paper.

REFERENCES

Aamodt, M. G. (1999). *Applied industrial/organizational psychology*, 3rd ed. Pacific Grove, CA: Wadsworth.

Austin, S. D. M. (1921). A study in logical memory. *American Journal of Psychology, 32*, 370–403.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign-language vocabulary and the spacing effect. *Psychological Science, 4*, 316–321.

Ausubel, D. P., & Youssef, M. (1965). The effect of spaced repetition on meaningful retention. *Journal of General Psychology, 73*, 147–150.

Ausubel, D. P., Stager, M., & Gaite, J. H. (1968). Retroactive facilitation in meaningful verbal learning. *Journal of Educational Psychology, 59*, 250–255.

Bjork, R. A. (1979). Information-processing analysis of college teaching. *Educational Psychologist, 14*, 15–23.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*, 245–248.

Bromage, B. K., & Mayer, R. E. (1986). Quantitative and qualitative effects of repetition on learning from technical texts. *Journal of Educational Psychology, 78*, 271–278.

Conway, M. A., Cohen, G., & Stanhope, N. (1992). Very long-term memory for knowledge acquired at school and university. *Applied Cognitive Psychology, 6*, 467–482.

Craig, C. S., Sternthal, B., & Olshan, K. (1972). The effect of overlearning on retention. *Journal of General Psychology, 87*, 85–94.

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*, 309–330.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology, 77*, 615–622.

Earhard, B., & Landry, D. (1976). Dual vs. single encoding in recognition memory as a function of lag, delay, and availability of semantic information. *Memory & Cognition, 4*, 609–615.

Earhard, B., Fried, C., & Carlson, G. (1972). Interference, overlearning, and anticipation time. *Journal of Experimental Psychology, 94*, 345–347.

Fitts, P. M. (1965). Factors in complex skill training. In R. Glaser (Ed.), *Training research and education* (pp. 177–197). New York: Wiley.

Foriska, T. J. (1993). What every educator should know about learning. *Schools in the Middle, 3*, 39–44.

Gilbert, T. F. (1957). Overlearning and the retention of meaningful prose. *Journal of General Psychology, 56*, 281–289.

Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8*, 528–538.

Greene, R. L. (1989). Spacing effects in memory: evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 371–377.

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: a review. *Human Factors, 23*, 199–213.

Hall, F. H. (1989). *Learning and memory.* Boston: Allyn & Bacon.

Kratochwill, T., Demuth, D. M., & Conzemius, W. C. (1977). The effects of overlearning on preschool children's retention of sight vocabulary words. *Reading Improvement, 14*, 223–228.

Krueger, W. C. F. (1929). The effect of overlearning on retention. *Journal of Experimental Psychology, 12*, 71–78.

Nelson, T. O., Leonesio, R. J., Shimamura, A. P., Landwehr, R. F., & Narens, L. (1982). Overlearning and the feeling of knowing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 8*, 279–288.

Postman, L. (1962). Retention as a function of degree of overlearning. *Science, 135*, 666–667.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55*, 297–308.

Richardson, J. (1973). Effect of speed of learning and degree of learning on cue selection. *Journal of Experimental Psychology, 98*, 396–403.

Rose, R. J. (1992). Degree of learning, interpolated tests, and rate of forgetting. *Memory & Cognition, 20*, 621–632.

Schendel, J. D., & Hagman, J. D. (1982). On sustaining procedural skills over a prolonged retention interval. *Journal of Applied Psychology, 67*, 605–610.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Shaughnessy, J. J. (1977). Long-term retention and the spacing effect in free-recall and frequency judgments. *American Journal of Psychology, 90*, 587–598.

Spector, P. E. (2000). *Industrial and Organizational Psychology: Research and Practice*, 2nd ed. Wiley: New York.

APPENDIX

Experiment 1

city	*country*
Axim	Ghana
Bari	Italy
Chiba	Japan
Doba	Chad
Lugo	Spain
Mago	Tonga
Pune	India
Rabat	Malta
Talara	Peru
Yumen	China

Experiment 2

word	*definition*
acrogen	fern
anta	pier
beldam	hag
cerate	wax
cess	tax
cicatrix	scar
elver	eel
emmet	ant
excrescence	wart
fosse	moat
mentum	chin
mome	fool
peruke	wig
ruga	fold
salver	tray
stannum	tin
talar	robe
tippet	cape
vizard	mask
weir	dam