

Retention of Spanish Vocabulary Over 8 Years

Harry P. Bahrick
Ohio Wesleyan University

Elizabeth Phelps
Princeton University

Thirty-five individuals who had learned and relearned 50 English-Spanish word pairs were tested for recall and recognition after an interval of 8 years. Two variables, the spacing between successive relearning sessions and the number of presentations required to encode individual word pairs, are excellent predictors of the likelihood of achieving permastore retention. Optimum recall occurs for words encoded in 1-2 presentations and accessed at intervals of 30 days. Both variables yield monotonic retention functions that account for a range of variation from 0% to 23% recall. These variables also have very significant effects on the recognition of unrecalled words.

A recent investigation of the retention of Spanish language learned in school (Bahrick, 1984) shows that a portion of the acquired knowledge has a life span of more than 25 years even if the knowledge is not rehearsed or accessed during that long interval. Another part of the originally acquired knowledge is lost within 5 years after training terminates, and virtually no knowledge is lost during the interval between 5 and 25 years following acquisition. This finding suggests the challenging possibilities of identifying conditions of learning and/or characteristics of material associated with a prospective life span of more than 25 years and of differentiating these conditions and characteristics from those associated with material destined to have a relatively short life span (less than 5 years). Such information would augment the very limited knowledge of life span memory currently available and make memory research more relevant to the needs of educators who have an obvious interest in prolonging the life span of transmitted knowledge. However, the research is arduous because it requires longitudinal investigations to extend substantially beyond the 5-year period during which material with a short life span is likely to be forgotten.

The foregoing considerations led us to search for relevant longitudinal data, and we realized that such data might be available to us if we conducted a follow-up to an earlier investigation (Bahrick, 1979). The aim of the earlier investigation was to establish the effect of various rehearsal schedules on maintaining access to learned material over periods of 1-9 months.

One part of that investigation required college students to learn and relearn 50 English-Spanish word pairs in successive training sessions, spaced at intervals varying from a few seconds

to 30 days. We found that cumulative learning is faster with shorter intersession intervals, but retention between training sessions was nearly perfect after seven training sessions, even with a 30-day intersession interval. Furthermore, shorter intersession intervals resulted in substantial forgetting if a 30-day interval was introduced for the first time, whereas that same 30-day interval yielded improved performance for individuals who were trained with that interval all along.

It occurred to us that we might test retention of the 50 word pairs after an interval of approximately 8 years and thus establish whether the intersession interval used as the independent variable in the original study had any effect after such a long time. The prospect of using these data for our new purpose was very attractive, because the independent variable was equally relevant to the purposes of the present and the original investigation, and many years could be saved. In addition, detailed records of acquisition for each word pair were available, and these records would permit us to examine the effects of a number of other performance variables on very long term retention.

To be sure, the independent variables to be examined are among the most researched in the psychological literature. The topic of massed versus distributed practice was expertly reviewed by Underwood (1961), who concluded that increasing the distribution of practice involves a trade-off between the benefits of spacing associated with dissipation of response interference and the detriments attributable to losses that occur between successive practice sessions spaced far apart (Melton, 1970). More recent analyses of the spacing effect have stressed encoding variability (Glenberg, 1979) and the processes involved in successive attempts at retrieval (Cuddy & Jacoby, 1982). Delayed retention tests generally have been found to enhance the benefits of distributed practice (Keppel, 1964). Atkinson's (1972) research dealt most directly with the learning of foreign vocabulary. He showed that retention of German-English word pairs is far superior when a computer programs the spacing of individual item rehearsals on the basis of each subject's performance for each word pair than when the rehearsals are randomly programmed or programmed in accordance with the judgment of optimal rehearsal by the subject. The retention interval used by Atkinson was 7 days.

Thus, the effects of distribution of rehearsals have been well

This investigation was supported by National Science Foundation Grant BNS-8417788.

The authors are indebted to Shirley Bodi, Jillian Daly, Irene Jenkins, and Cara Wellman for help in the analysis of data; to Lynda Hall for many contributions to the analysis and to the editing of the manuscript; and to Ulric Neisser, Eugene Winograd, Arthur Glenberg, Charles Thompson, and Henry L. Roediger III for valuable reviews of the manuscript.

Correspondence concerning this article should be addressed to Harry P. Bahrick, Department of Psychology, Ohio Wesleyan University, Delaware, Ohio 43015.

established for retention intervals of 2 weeks or less, but not for longer intervals of time. A major goal of the present study was to determine if such effects continue to play a role in establishing memory with a life span of many years.

The plan to use the available data also involved a number of serious problems: (a) the original participants had dispersed and were no longer readily available to us; (b) many of the participants in the original investigation were subjected to a change of the intersession interval after three to seven relearning sessions, and this change would make it more difficult to attribute differences in retention to particular intersession intervals; (c) individuals trained with a given intersession interval varied in the number of successive relearning sessions (ranging from six to nine); and (d) we doubted whether retention of the word pairs after 8 years would be sufficient to warrant the investigation. However, preliminary data, obtained from a few individuals still available to us for testing, convinced us that the investigation had merit in spite of the above concerns, and we decided to go ahead.

Method

A brief review of the method used for training subjects in the 1979 investigation is in order. The first training session began with a presentation trial in which 50 English-Spanish word pairs were presented visually, individually at a 5-s rate. The experimenter also pronounced each Spanish word during this first presentation trial. The presentation trial was followed immediately by a test trial in which only the English words were presented visually in random sequence, and the subject was allowed 10 s to pronounce the corresponding Spanish word. A dropout technique was used so that on the following presentation trial only those word pairs that the subject had failed were presented. Alternating test and presentation trials continued until the session ended with the first test trial on which all remaining items were passed.

Subsequent relearning sessions began with a test trial that included all of the 50 word pairs. This trial was again followed by alternating presentation and test trials according to the dropout technique, until all remaining items were passed on a test trial. Thus, the original sessions and subsequent relearning sessions differed in that the original sessions began with a presentation trial, whereas subsequent sessions began with a test trial. Because presentation of an item was contingent on failure to recall that item on the preceding test trial, it was possible for an item to be presented only once, on the initial presentation trial in the first

session, provided the item was passed on the initial test trial of all subsequent relearning sessions.

The interval between successive relearning sessions was a between-subjects independent variable; the three intervals of interest here are 30 days, 1 day, and 0 (no interval). Subjects trained with the 0 intersession interval began a new retraining session immediately after the previous session was completed. Intersession intervals were changed for most subjects after five to seven relearning sessions, and one to four additional sessions were administered following the change of the intersession interval. Table 1 summarizes the training conditions for subjects who participated in the present study.

The original training took place during a period of 14 months, the training of some individuals continued for 8 months, and others were trained in a single session (see Table 1). The follow-up retention tests were administered over a period of 6 months. The average interval from the termination of training to the retention test was approximately 8 years, with a range from 81 to 103 months.

A total of 64 individuals who had served in the original study were sent a letter reminding them of their participation as undergraduates and requesting them to agree to take two additional memory tests. Follow-up letters and phone calls yielded a total of 48 affirmative responses.

Those persons who agreed to participate were sent a questionnaire that inquired about various types of possible exposure to the Spanish language during the 8-year retention interval. Individuals were eliminated if they had resided in Spanish-speaking countries or bilingual areas, or if they had taken formal instruction in the Spanish language. Some of the original participants could not be located; a number failed to reply for unknown reasons, and one was deceased. The yield of 75% is probably a reasonably unbiased sample of the original participants. The yield is quite high because most of the participants were former students of one of the investigators, and they were supportive of his research. Of those willing to be tested, 13 were subsequently eliminated because they had been exposed to the Spanish language during the retention interval or because they failed to return a test. Most of the participants had received hourly pay in the original investigation, but they agreed to serve in the current investigation without pay.

Those who agreed to be tested and survived the rejection criteria were sent a recall test listing the 50 English words. They were asked to respond by writing down the equivalent Spanish words, to do so at their own pace and in any order, to complete the test in one session, and to return the test immediately without consulting any other sources. Five participants who resided in the central Ohio area were tested individually by the investigators. As soon as participants returned their completed recall tests, they were sent recognition tests. The recognition tests listed all of the words (in English) the subject had failed to recall on the

Table 1
Original Training Conditions of the Participants

Primary intersession interval (in days)	No. of sessions prior to change	Changed intersession interval (in days)	No. of sessions after change	<i>N</i>
0	6	none	0	1
0	5	30	1	1
0	6	1	1	3
0	6	30-1	1, 2	5
1	6	none	0	1
1	6	30	1	2
1	6	30	2	6
1	3	30-1	2, 2	5
30	7	1	1	7
30	7	none	0	4

recall test, together with five alternative words (in Spanish). The foils were matched with the correct response on part of speech. The instructions required the participants to respond to each word, guessing if necessary.

A control group of individuals who had not participated in the original investigation was recruited to provide a baseline for differentiating memory content established during the original training sessions from content derived incidentally from other sources. The control group was recruited by writing letters to Ohio Wesleyan University alumni who were members of the same graduating classes as the participants in the original investigation. Invitations were extended to individuals who met the criteria applied in the original study. Thus, only persons who had no formal training in the Spanish language and who had not lived in Spanish speaking or bilingual areas were recruited. Volunteers were sent the questionnaires inquiring about exposure to the Spanish language during the years following graduation from college, and the same criteria for rejection were applied as for the participants in the original investigation. Ten individuals met the criteria and agreed to serve in the control group. They were mailed the tests for recall and recognition of the 50 word pairs. Instructions and procedures for this group were the same as those previously described.

Results and Discussion

Table 1 lists the number of individuals in each original training condition who completed the 8-year retention test. In spite of the variation in the number of retraining sessions and of the changes in the intersession interval, the data permit analysis of the effects of two types of variables on retention: the manipulated variables (e.g., the intervals between successive training sessions) and variables of performance during acquisition (e.g., the record of successes and failures on test trials for individual words).

The Effect of the Intersession Interval

Recall tests were scored and minor spelling errors were disregarded because participants had originally seen the correct Spanish spelling but had been trained to pronounce rather than spell responses. For this reason, responses were scored as correct if they could be identified phonetically without ambiguity. Recognition tests were limited to questions failed on the recall test. As a result, the respective scores are mutually exclusive; the recognition scores do not include recalled words.

Table 2 gives the mean percentage of responses recalled, the percentage of nonrecalled words that were recognized, and the percentage of words failed on both tests as a function of the primary intersession interval. The data were analyzed on the basis of the primary interval, without regard to the changes designated in Table 1. Thus, participants who received from six to nine training sessions are grouped together in Table 2. The average number of training sessions varies only slightly across the primary intervals (7.3–7.8); the zero interval is associated with the largest number of sessions.

The data show that the intersession interval has a very important effect on retention after eight years. The recall probability associated with the 30-day interval is 2.5 times the probability associated with the zero interval. The effect on recognition exclusive of recall is much less pronounced.

Although the effects shown in Table 2 are quite large, they almost certainly underestimate the true retention differences at-

Table 2

Percentage of Words Recalled, Recognized but Not Recalled, and Failed as a Function of the Primary Intersession Interval

Primary intersession interval	% Correct		% Failed on both tests
	Recall test	Recognition test	
30 days	15	83	14
1 day	8	80	18
0	6	71	27
control	1	62	37

tributable to the three intersession intervals. The underestimation stems from the fact that the primary intersession intervals used in this investigation were modified during training, as shown in Table 1, and these changes are certain to dilute the effects of the primary interval.

Incidental learning as reflected by the performance of the control group makes a trivial contribution to recall performance but a substantial contribution to recognition. Individuals in the control group were able to identify 62% of the Spanish words on the recognition test, probably because most subjects had some knowledge of French or Latin. This knowledge may provide a common root sufficient to identify the correct alternative on a multiple choice test, but it does not permit independent retrieval of the correct response. This interpretation is particularly plausible because no obvious Spanish-English cognates were included among the 50 word pairs.

Long-Term Retention as a Function of Word Difficulty During Acquisition

The acquisition data available from the Bahrack (1979) investigation made it possible to determine each subject's relative ease or difficulty of learning each of the 50 word pairs. We considered several indicants of word difficulty, and all of them yielded similar findings. The most informative indicant is based on the total number of presentation trials an individual subject required to learn and relearn each word. As previously stated, presentation of each word pair was contingent on failure to recall the Spanish response word on the preceding test trial. Thus, the most readily learned words were presented only once during the entire retraining process, provided they were recalled on the initial test trial in all subsequent retraining sessions. Most words required repeated presentations in several retraining sessions before they were mastered by the subject.

The cumulative presentation count is flawed as an indicant of word difficulty because it may be confounded by the variable number of retraining sessions. However, an alternative indicant based upon the average number of presentations per training session proved to be less useful because it yielded distributions of word difficulty that were strongly affected by arbitrary decisions regarding the selection of an appropriate scaling interval. Results obtained from several indicants of word difficulty were quite similar.

The cumulative number of presentations required during training was established for each word learned by each subject,

Table 3
Probability of Retention Over 8 Years as a Function of Number of Presentations During Training

Measure	Number of presentations during training					
	1-2	3-4	5-6	7-8	9-10	11+
Recall	.14	.08	.07	.06	.04	.02
Recognition	.83	.79	.74	.78	.56	.63
Failure	.14	.19	.24	.21	.42	.35
<i>N</i>	634	562	254	122	47	81

Note. One subject's acquisition data were not available for the analysis of number of presentations but were included in that of the effect of retraining interval.

and it was further determined whether the word was recalled, recognized, or failed on the 8-year test. Table 3 presents the recall, recognition, and failure probabilities as a function of the number of presentations during acquisition. These data show that retention 8 years after learning is strongly influenced by the total number of presentations required to reach the successive relearning criteria during acquisition. Furthermore, the table considerably understates the strength of this relation because it combines data for all three intersession intervals used in training. The longer intersession intervals not only yield better retention after 8 years, but also slowed down acquisition, that is, they are associated with a larger number of presentations per word during training. More specifically, the average number of presentations per word pair is 3.0 for the zero intersession interval, 4.0 for the 1-day interval, and 5.5 for the 30-day interval. Combining the data from the three intersession intervals therefore obscures the strength of the effect of relative word difficulty on long term retention. Table 4 shows the full strength of the effect; it gives retention probability as a function of the number of presentations separately for each of the three intersession intervals.

The recall and recognition data in Table 4 were subjected to separate analyses of variance (ANOVAs) that confirmed the statistical significance ($p < .001$) of the obvious main effects of intersession interval, $F(2, 1688) = 22.96$, $MS_e = .083$ for recall, and $F(2, 1526) = 20.56$, $MS_e = .164$ for recognition; number

Table 4
Probability of Retention as a Function of the Intersession Interval and the Number of Presentations During Acquisition

Intersession interval (in days)	Acquisition presentations			
	1-2	3-4	5-6	7+
Recall				
0	.09	.03	.02	.00
1	.15	.06	.03	.01
30	.23	.15	.14	.07
Recognition				
0	.78	.66	.62	.54
1	.84	.80	.76	.72
30	.92	.92	.78	.71

Table 5
Probability of Recall for the 10 Most Frequently Recalled Words

Intersession interval (in days)	Acquisition presentations		
	1-2	3-4	5+
0	.24	.05	.17
1	.36	.28	.20
30	.48	.47	.33

of presentations during acquisition, $F(3, 1688) = 15.37$, $MS_e = .083$ for recall, and $F(3, 1526) = 13.57$, $MS_e = 0.164$ for recognition; and the interaction between the two, $F < 1.0$ for both recall and recognition.

The recall data in Table 4 indicate monotonic functions for both rows and columns. The magnitude of the effects is impressive, especially when considering that an interval of 8 years is involved, that the manipulation of the intersession interval is imperfect, and that the degree of control over other variables (e.g., retention testing conditions) does not meet the standards typically observed in laboratory investigations.

The recall data in Table 4 indicate trade-offs between the manipulated intersession interval and the word difficulty variable. Words acquired in one or two presentations with a zero intersession interval are recalled with about the same probability as words acquired in seven or more presentations with the 30-day interval. Because the manipulation of the intersession interval was flawed, the effects attributable to that variable are probably underrepresented in the table, and this underestimate also affects estimates of trade-off between the two variables.

The effects on recognition of nonrecalled words are also large, but chance success and incidental learning account for a base performance of approximately 62% correct. The experimental effects must be interpreted in relation to this control.

We wanted to establish to what extent results of the study were affected by the difficulty level of individual word pairs, defined on the basis of the number of subjects who recalled each word after 8 years. Figure 1 shows the distribution of word difficulty based upon this indicant. It can be seen that the best retained words were recalled by about one third of the subjects; 22 words were so difficult that they were recalled by no subject or by only one subject. The effect of the independent variables cannot be established for the most difficult words because their recall frequency is too low, but Table 5 shows the effects for the 10 most frequently recalled words. Clearly, recall probabilities are higher in Table 5 than in Table 4, but the effects of the intersession interval and the number of presentations during training are comparable in the two tables.

The present results show that approximately 10% of vocabulary learned and relearned in six to nine successive sessions can still be recalled 8 years later. A large portion of the nonrecalled words can be identified on recognition tests. More important, the three intersession intervals used in this study affect recall after 8 years by a factor of at least 2.5 to 1, with the 30-day interval yielding the best performance.

The effect of speed of learning of individual words is even more impressive. Fourteen percent of words acquired in one or

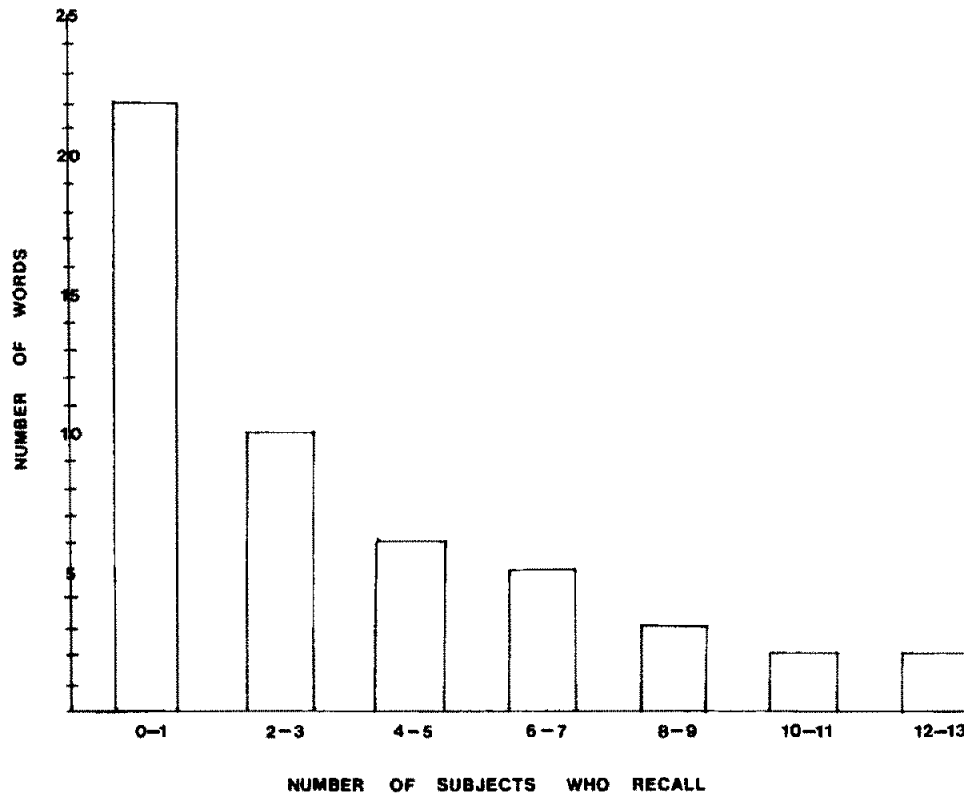


Figure 1. Frequency distribution of 8-year recall for 50 word pairs.

two presentations are still recalled after 8 years, but only 2% of words acquired in more than ten presentations are recalled. This is true in spite of the fact that training for all words was terminated on the basis of the same criterion within each session.

It has been known for a long time, of course, that spaced practice yields superior retention (Cain & Willey, 1939) and that terminating practice at a common criterion yields better retention for target items that are learned more quickly (Underwood, 1966, p. 548). However, the available literature offers no information about the persistence or the magnitude of such effects after intervals of many years. The present data show that the combined effects of the intersession interval and the presentation trials account for a range of variation from 0% to 23% recall for the least favorable versus the most favorable condition.

The Implications for Permastore Retention

Results from Bahrick (1984) show that Spanish vocabulary acquired in high school or college courses will be recallable for more than 25 years if the vocabulary is not lost during the first 5 years following training. This finding holds even for material not rehearsed during the 25-year retention interval. The term *permastore* is used to designate unrehearsed memory content with a life span in excess of 25 years. There is no guarantee, of course, that the words recalled after an interval of 8 years in the present investigation will still be recalled after an interval of 25 years. However, such a result is probable, because the 1984 data

are based on the same type of material and on the performance of more than 800 individuals whose training occurred under a considerable variety of conditions.

Other conclusions from the Bahrick (1984) investigation can be examined more directly. The 1984 data indicate that permastore longevity of individual target items is determined by conditions of acquisition and is relatively independent of conditions encountered during the retention interval. This conclusion was based on the fact that the proportion of original knowledge retained for more than 25 years varies greatly among subjects, and this variation is predictable on the basis of the level of original training (number of Spanish courses taken and grades received), but is largely independent of the conditions during the retention interval. The 1984 investigation is based upon cross-sectional, rather than longitudinal, data, and conclusions regarding variables that affect the fate of individual target items may be questioned on that basis (Neisser, 1984). The present data are longitudinal; that is, they relate retention of individual words to the original training record for these same words. The longitudinal data are definitive; that is, they leave no doubt concerning the decisive role of conditions of acquisition (including training variables and Stimulus \times Subject interaction effects) in the determination of the life span of the individual target items. Specifically, it is now clear that 8-year retention probability is greatly enhanced for words that are well encoded in one or two presentations (a Stimulus \times Subject interaction effect) and are subsequently accessed several times at intervals of 30 days (a training variable). Equally frequent access at shorter intervals is

much less effective for words that are originally equally quickly learned.

That 8-year recall is optimal for word pairs presented only once or twice suggests that encoding variability during the presentation of items is not critical for permastore memory. Rather, as suggested by Thompson, Wenger, and Bartling (1978), encoding may also occur during retrieval, without presentation of the item, and the spacing of successful retrievals is a critical determinant of life-span retention.

The results of this study have implications for educators. Permastore retention may not be facilitated by repeated presentations of the material to be learned, particularly if the intervals between successive presentations is only 1 day or less. Instead, educators need to identify effective encoding conditions and to make certain that students independently retrieve target information at intervals that are as long as 1 month, over a period of several years, instead of the more typical intervals of 1 to 2 days over periods of from 10 to 15 weeks. Extended training schedules would slow down acquisition somewhat, but they would yield disproportional benefits for long-term retention. Optimal rehearsal schedules must be based on empirically established optimum intervals between sessions. The optimum interval is likely to be the longest interval that avoids retrieval failures, and techniques for establishing such intervals have already been described by Landauer and Bjork (1978), but they have not been applied to the problems of education. Programmatic research can establish optimum intervals as a function of the type of material, the stage of acquisition, and the type of student. Such findings will yield major contributions of memory research to education.

References

- Atkinson, R. (1972). Optimizing the learning of a second language vocabulary. *Journal of Experimental Psychology*, 96, 124-129.
- Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, 108, 296-308.
- Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113, 1-29.
- Cain, L. F., & Willey, R. (1939). The effect of spaced learning on the curve of retention. *Journal of Experimental Psychology*, 25, 209-214.
- Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, 21, 451-467.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition* 7, 95-112.
- Keppel, G. (1964). Facilitation in short- and long-term retention of paired associates following distributed practice in learning. *Journal of Verbal Learning and Verbal Behavior*, 3, 91-111.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). New York: Academic Press.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 596-606.
- Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General*, 113, 32-35.
- Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 210-221.
- Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review*, 68, 229-247.
- Underwood, B. J. (1966) *Experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

Received April 14, 1986
Revision received July 3, 1986 ■