AMERICAN JOURNAL OF PSYCHOLOGY
June 1980, Vol 93, No. 2, pp. 329-343

0002-9556/80/0600-0329$01.50/0

Judgments of knowing: The influence of retrieval practice

Joseph F. King
Loyola University of Chicago

Eugene B. Zechmeister
Loyola University of Chicago

John J. Shaughnessy
Hope College, Michigan

Four groups of college students first learned two paired-associate lists. Two experimental groups were asked to predict the likelihood that the response term of each pair would be recalled on the final test trial; learning either involved alternating study and test trials or only study trials before the prediction task. For the two control groups, the prediction task was omitted. All four groups then learned a third list without test trials and made judgments of knowing for items in this list. Subjects who had received test trials showed consistently higher prediction accuracy on the first two lists but their accuracy decreased on the third list when test trials were absent. A major source of facilitation produced by the opportunity for retrieval practice was traced to subjects' knowledge of their previous recall performance.

Contemporary views of learning and memory lead us to characterize the learner as a decision maker who selectively draws upon a repertoire of mnemonic devices depending upon the perceived task demands. For example, recent interpretations of phenomena such as the negative recency effect in free recall (e.g., Craik & Watkins, 1973) and the effect of input location on item strength (Conover & Brown, 1977) are based upon the premise that subjects emphasize a rehearsal strategy for recency items that is different from that for pre-recency items. Similarly, accounts of subjects' rehearsal patterns (e.g., Rundus, 1971) and distribution of study time in self-paced learning situations (e.g., Zimmerman, 1975) emphasize the strategic nature of human information-processing. These proposed control processes depend upon the learner being sensitive to what is "working" in a particular task; in other words, the learner must be able to judge both what is known and what is not known regarding the target material.

There have been few attempts to study the learner's ability to monitor the quality of a memory trace as it is being established, that is, the learner's ability to predict accurately whether material presently being studied has been encoded sufficiently to allow for subsequent recall of the material at the time of the memory test. We will refer to this kind of prediction as a "judgment of knowing." One notable study along these lines was done by Arbuckle and Cuddy (1969). They asked college students to predict the likelihood of subsequent recall in a paired-associate learning task, with the predictions being made during the initial, and only, study trial. Although the students showed an above-chance level of prediction accuracy, Arbuckle and Cuddy argued that this accuracy was probably not the result of assessments by the subjects of the effectiveness of their present mnemonic efforts. Instead, subjects appeared to base their predictions on perceived item difficulty (cf., Underwood, 1966). Thus, their results cannot be used to argue that the learner makes predictions of subsequent recall performance using knowledge gained from the encoding process.

One kind of information which could serve as a basis for judgments of knowing is the learner's memory for the outcome of previous recall attempts. A recent study by Gardiner and Klee (1976) demonstrates that this information is, in fact, available to the learner. In their experiment, college students were given a series of free-recall lists, each of which they studied and attempted to recall once. Following the recall test on the last list the subjects were given the words from all the lists and were asked to designate which words had been recalled on the initial tests. Although this type of test was unexpected, subjects were generally accurate in distinguishing previously recalled and non-recalled items, except for items taken from the recency portion of each list. Comparable performance on the "memory for remembered events" (MRE) test was obtained by Klee and Gardiner (1976) when a serial-recall test was used, but much lower MRE performance was obtained following recognition tests.

Gardiner, Passmore, Herriot, and Klee (1977) have suggested that memory for remembered events may play a critical memory-monitoring role in multiple-trial learning situations: "Here, the subject's knowledge of his previous performance can provide feedback information which may lead to decisions with respect to the regulation of a variety of control processes" (p. 53). Similar suggestions have been made by Thompson, Wenger, and Bartling (1978) and by Halff (1977) in their discussions of the role of recall trials in free recall and by LaPorte and Voss (1974) in their investigation of the role of test trials in paired-associate learning. In all these accounts, knowledge of previous test-trial performance is assumed to be the basis of subjects' decisions concerning encoding strategies on subsequent study trials.

The major purpose of the present research was to examine subjects' memory-monitoring performance directly to determine whether previous test-trial experience facilitiates a learner's judgment of knowing. After several exposures of the to-be-recalled items, all subjects were asked to predict for each pair the likelihood that the response term would be recalled upon presentation of the stimulus term on the subsequent test trial. For half of the subjects, the exposures prior to the prediction task involved study trials only; for the other half of the subjects, study and test trials alternated. To the extent that test trials provide information regarding the distinction between known and unknown material, it was expected that those subjects having test trials would show a higher level of prediction accuracy than would subjects having only study trials.

An additional concern of the present research was the extent to which subjects' abilities to predict their own performance improve with practice. LaPorte, Voss, and Bisanz (1974) have previously tested the hypothesis that experience with test trials is a major component contributing to nonspecific transfer. Unfortunately, their findings did not support the hypothesis. Using a trials-to-criterion measure, they found significant and equal learning-to-learn for groups having and not having test trials during the acquisition for the first list. It is still possible that the discrimination between known and unknown information may depend on test-trial experience and may improve with practice even though test-trial experience is not a determining factor of learning-to-learn as indexed by overall learning rate. To examine this issue further, subjects in the present experiment learned two unrelated lists of paired-associates with performance on the prediction task serving as an index of learning-to-learn. If test-trial feedback is essential to learning to discriminate known and unknown information, then we should find superior learning to learn for those subjects having test trials as compared to those having only study trials.

Finally, all subjects were required to learn a third paired-associate list and make judgments of knowing without the benefit of test trials. If test trials are as critical as we presume them to be, then the accuracy of the predictions made by those subjects who have had test trials on the first two lists should decrease on this third list. Further, any difference in prediction accuracy on this third list between groups having and not having had test trials on this third list between groups having had test trials on the first two lists would indicate residual effects of the test-trial experience. To aid in this evaluation the design included two groups of subjects who did not make judgments of knowing before the third list.

METHOD

Design

College students were asked to learn three lists of 24 paired-associates under one of four experimental conditions. The conditions differed in terms of the procedures followed for learning of the first two lists. In two conditions the students received five study trials followed by either a prediction trial or another study trial and then a final test trial (i.e., SSSSS-P-T or SSSSS-S-T). The former group will be designated the Study-Only-Prediction (SOP) group and the latter group will be designated the Study-Only-Control (SOC) group. On the prediction trial both the stimulus and response member of each pair were presented and subjects were asked to predict the likelihood of their recalling the response member on the following test trial. In the other two conditions, paired-associate learning of the first two lists involved alternating study and test trials with three study-test alternations followed by a prediction trial or study trial and a final test trial (i.e., STSTST-P-T or STSTST-S-T). These groups will be designated the Study-Test-Prediction (STP) group and the Study-Test-Control (STC) group. Pilot work had indicated that three study-test cycles produced approximately the same level of learning as five study trials with no test trials. On list 3, in all conditions, three study trials were followed by a prediction trial and a final test trial (i.e., SSS-P-T).

Materials

Three different paired-associate lists were constructed in the following manner. Two-syllable concrete nouns with Thorndike-Lorge frequencies between 1 and 50 per million were selected from the Paivio, Yuille, and Madigan (1968) norms. These were then ranked according to rated imagery, and the 72 nouns with the lowest imagery ratings were chosen to serve as response terms. Stimulus terms were 72 CVC (consonant-vowel-consonant) trigrams with association values ranging between 70 and 80% (Noble, 1961). Stimulus and response terms were randomly paired; pairs were divided into three lists of 24 with the restriction that no two stimuli in a list had the same first two letters. Pairs were typed on index cards for presentation to the subjects. Each of the three lists was used equally often as the study list for the three stages of the experiment.

The 24 pairs were ordered differently on each study, test, or prediction trial. Further, in order to control the time between an item's presentation on a study trial and occurrence on the prediction trial, as well as between occurrences on prediction and test trials, pairs in a study trial were arranged in four blocks of six items each. On each study trial, the order of blocks and the appearance of items within a block were determined randomly. The prediction trials were divided into six blocks of four items each, and every block contained one item from each study block. Using this method, two different orders were used for the prediction trial in each list. Finally, the order of items on the test trials was determined by selecting one item from each of the prediction-trial blocks to appear in each of six test-trial blocks. Several orders of test items were used in each list.

Prediction task

The prediction task was introduced as a game wherein subjects could earn points by correctly predicting recall or nonrecall of the paired-associate response terms. The game format was originally used by Pasko (Note 1) and its purpose was to obtain estimates of memory-monitoring accuracy when subjects were both motivated to attempt accurate predictions and uninclined to withhold responses in order to guarantee correct prediction of nonrecall. An added benefit was that subjects found the game interesting.

For each pair during a prediction trial, subjects were asked to use a six-point scale to indicate how confident they were that the response term in the pair would be recalled on the following test trial. The scale, however, was clearly described as reflecting two judgments: "No, I will not recall the item," and "Yes, I will recall the item." The first three scale points (1, 2, 3) were used to predict nonrecall with lower numbers indicating greater confidence in nonrecall. Scale points 4, 5, and 6 were to be used when recall was predicted, and increasing scale values indicated increasing confidence in predicting recall. In other words, the higher the scale value selected, the greater the predicted confidence in recall.

All subjects were told that recalling an item would earn them 5 points and that nonrecall would be penalized by subtracting 5 points from their point total. The contingency was in effect no matter what was predicted. However, subjects were informed that additional points could be earned if their predictions matched their test trial performance, and conversely, that penalty points would be given when performance was not accurately predicted. The bonus points and penalty points were earned according to the degree of confidence that was indicated. Specifically, 1, 2, or 3 bonus or penalty points were given according to the confidence indicated. For example, if an item was actually recalled and recall had been predicted, then 1, 2, or 3 bonus points were earned for indicating a confidence value of 4, 5, or 6, respectively. If recall was predicted and the item was not recalled, 1, 2, or 3 points were substracted depending again on the confidence value. This time higher confidence (6) yielded the most penalty points (3).

The system worked in the opposite manner for predicting nonrecall. For example, subjects earned 3 points for predicting nonrecall with a confidence of 1, when, in fact, recall was not observed. Because nonrecall always resulted in subjects losing 5 points, predicting nonrecall allowed subjects to earn back some of the lost points. The game was explained to the subjects with the aid of a diagram illustrating the various point combinations and instructions pointed out that recall earned subjects the most points no matter what was predicted and that accurate prediction was the subjects' best strategy.

Procedure

Subjects were tested in pairs and each pair was assigned upon appearance at the laboratory to an experimental group using a block randomization procedure. Instructions informed subjects that they were to participate in a memory study and that their ability to predict would be important. All groups were given typical paired-associate learning instructions and were told that the list would be presented several times for study. Instructions for the prediction task were given to groups SOP and STP prior to the presentation of the first list. The other two groups, SOC and STC, received these instructions immediately prior to presentation of the third list. Subjects were not informed as to the number of study lists in the experiment or the number of pairs in a list.

Pairs were presented on the study and prediction trials at a constant rate which was indicated to the experimenter and subject by tape-recorded tones. The same was true for the presentation of stimuli on test trials. Pairs were shown on the study trials at a 3-sec rate; prediction and test trials were paced at a 5-sec rate. Subjects responded on both test trials and prediction trials by recording responses on sheets of paper which they pulled from large envelopes, exposing one new item each time a tone was heard. An exception to these presentation rates occurred when subjects in the prediction conditions received a prediction trial before the last test trial on the first two paired-associate lists. Subjects not asked to perform the prediction task were given an additional study trial at a 5-sec rate in order to equate experience with the list.

On the third list all subjects learned the pairs with no intervening test trials. Pairs were presented at a 3-sec rate for three trials, followed by a prediction trial and then a test trial, both at 5-sec rates. When instructions regarding the prediction task were given to those subjects in groups SOC and STC, subjects in the prediction groups were given a brief review of the instructions and task. Therefore, time between the second and third lists was held constant for all groups.

Subjects

There were 30 Loyola University undergraduates assigned to each group. Students participated in order to earn extra credit for an introductory psychology course. However, two subjects in group STP obtained perfect recall and one subject in group SOP obtained zero recall. Since the measure to be used as an index of prediction accuracy is undefined if recall is either perfect or zero, the data from these subjects could not be used and an attempt was made to reestablish the equivalence of the four groups by dropping the two subjects from each group with the highest total recall and the one subject with the lowest total recall across the first two lists. Thus, the analyses to follow were based on sample sizes of 27 per group.

RESULTS

Throughout this paper, effects characterized as statistically significant or reliable exceed the .05 level of significance.

Paired-associate recall

Although the accuracy of subjects' predictions is of primary interest, the analyses of recall performance will be presented first. The mean number of correct responses on the test trials of each of the three lists for the four different conditions is presented in Table 1. The columns labeled "Criterion" represented recall on the test trials following the prediction trial or following the study trial which was substituted for the prediction trial. Overall, performance on the criterion trials of list 1 and 2 was higher for groups having test trials, but all groups showed learning-to-learn as evidenced by the consistent increases from list 1 to list 2. Finally, differences among the groups were substantially smaller on the criterion trial of list 3.

Table 1. Mean number of correct responses across lists on test trials preceding (T-1, T-2, T-3) and following (Criterion[a]) the prediction task

	List 1				List 2				List 3
Condition	T-1	T-2	T-3	Criterion	T-1	T-2	T-3	Criterion	Criterion
Study-Test-Prediction (STP)	1.7	5.2	9.3	12.11	3.0	8.7	12.8	15.81	10.89
Study-Test-Control (STC)	1.1	5.2	8.5	11.85	3.1	8.1	12.6	15.41	10.56
Study-Only-Prediction (SOP)	—	—	—	9.15	—	—	—	11.00	8.67
Study-Only-Control (SOC)	—	—	—	8.59	—	—	—	11.59	10.11

[a]Criterion represents recall on the test trials following the prediction trial or following the study trial which was substituted for the prediction trial.

The analysis of the recall on the criterion trials of the first two lists included two between-subjects variables along with the within-subject variable of lists. The first of these represented the presence or absence of the prediction task and the second the presence or absence of test trials. Thus, the analysis was a 2 × 2 × 2. There were no reliable differences attributable to the main effect of having had the prediction task nor were there any interactions involving this variable, $F\text{s} < 1.12$. This indicates that performing the prediction task was comparable to having an additional study trial. The effect of the test trial variable, however, was statistically significant, $F(1, 104) = 17.57, MS_E = 42.38$. The mean recall for the two conditions with test trials (summing over lists) was 13.80 while that for the conditions having only study trials was 10.09. The increase in mean recall across lists was also reliable, $F(1, 104) = 97.83, MS_E = 5.06$. The mean number of correct responses summing over the four conditions was 10.43 for list 1 and 13.46 for list 2.

There was also some suggestion of greater learning-to-learn in those conditions having test trials. Those groups having test trials showed a slightly greater increase across lists (11.98 vs. 15.61) than the groups having only study trials (8.87 vs. 11.30), and this interaction was statistically significant, $F(1, 104) = 3.87, MS_E = 5.06$. This effect should be interpreted with caution, however, because the first-list recall indicates that the adjustment in the number of presentations for the study-only groups based on the results of the pilot study (i.e., 5 study trials as compared to 3 study-test cycles) was not successful in equating initial recall levels on the criterion trial. Therefore, the greater increase across lists for the test-trial groups may be attributable to the higher degree of learning of the first list rather than to the effect of the test trials per se.

The second major analysis of the recall data involved performance on list 3. The results of this analysis showed no effect of having had test trials or the prediction task on the first two lists nor was there any interaction, $F\text{s}\,(1, 104) < 2.43, MS_E = 19.97$. As can be seen in Table 1, there was a slight advantage in mean recall for those groups having had test trials (10.73) as compared to the groups having had only study trials (9.39); however, this difference can be attributed to the unexpectedly low recall level in group SOP.

Prediction accuracy

Two measures of prediction accuracy were used. The first measure was comparable to the one developed by Zimmerman, Broder, Shaughnessy, and Underwood (1977). The measure is a ratio, the numerator of which is the difference between the mean confidence judgment assigned to items which were recalled on the criterion trial and the mean confidence assigned to items which were not recalled. The denominator of the ratio is the square root of the pooled variance of the subject's confidence judgments for recalled and nonrecalled items. This measure will be referred to as the confidence-judgment accuracy quotient (CAQ), and it can be considered roughly analogous to d' in a signal-detection analysis. Complete failure to distinguish known from unknown information would result in a CAQ score of zero. The denominator of the CAQ score adjusts the difference between mean confidence judgments for known and unknown information for how "tightly" a given subject used the confidence scale. The second measure of prediction accuracy was the Type II d' score based upon the hit rate and false alarm rate for each subject's predictions. For this measure, the hit rate and false alarm rate were determined by dichotomizing the confidence judgments into categories including ratings of 1, 2, or 3 (predictions of nonrecall) and ratings of 4, 5, or 6 (predictions of recall).

A separate CAQ and d' score was calculated for each subject for each of the three lists in conditions STP and SOP, and for the third list in conditions STC and SOC. The mean CAQ and d' scores across subjects are presented in Table 2.[1] As was true for recall performance, the mean prediction accuracy scores on the first two lists were higher for those subjects who had test trials. The mean CAQ score in group STP (2.10) was statistically significantly greater than that in group SOP (1.27), F (1, 52) = 26.46, MS_E = .70. Contrary to the recall scores, however, the mean accuracy scores did not show much of an increase across lists. For the analysis of lists 1 and 2, neither the main effect of lists nor the interaction of lists with conditions was reliable. As expected, however, the drop in CAQ scores from list 2 to list 3 in the STP group was reliable, F (1, 26) = 12.20, MS_E = .74. Finally, the analysis of mean CAQ scores on list 3 for the four conditions produced no reliable effects, Fs (1, 104) < 1, MS_E = .19.

As was mentioned previously, the degree of learning of the first two paired-associate lists was consistently higher in the groups which had test trials. It could be suggested that the greater degree of prediction accuracy in condition STP simply reflects this higher degree of learning. To examine this possibility, subsets of subjects from groups SOP and STP were matched on the basis of total recall for the criterion trials of lists 1 and 2. Appropriate matches were made for 18 pairs of subjects. The mean recall, CAQ, and Type II d' scores for these subjects are presented in Table 3. If anything, this matching procedure

Table 2. Mean CAQ and Type II d' scores across lists

	List 1	List 2	List 3
Study-Test-Prediction (STP)			
CAQ	1.91	2.29	1.47
d'	2.06	2.34	1.72
Study-Test-Control (STC)			
CAQ	—	—	1.45
d'	—	—	1.74
Study-Only-Prediction (SOP)			
CAQ	1.18	1.35	1.66
d'	1.60	1.56	1.86
Study-Only-Control (SOC)			
CAQ	—	—	1.50
d'	—	—	1.76

Table 3. Mean recall, CAQ, and Type II d' scores for subjects matched on total recall for lists 1 and 2

	List 1	List 2
Study-Test-Prediction (STP)		
Mean recall	9.83	13.22
CAQ	2.04	2.08
d'	2.11	2.19
Study-Only-Prediction (SOP)		
Mean recall	10.39	12.72
CAQ	1.21	1.29
d'	1.69	1.56

is biased against showing greater prediction accuracy in condition STP because matches required pairing an STP subject with relatively poor recall within that group with an SOP subject with relatively good recall in that group. Clearly, however, greater prediction accuracy is present in group STP than in group SOP even when differences in recall level have been eliminated. The mean CAQ score across lists in group STP (2.06) was significantly greater than that in group SOP (1.25), F (1, 34) = 17.82, MS_E = .65. In fact, the mean differences in prediction accuracy across conditions for both the CAQ and Type II d' scores for the matched subjects are quite comparable to those obtained for the groups as a whole. These results are obviously not consistent with the notion that the greater degree of prediction accuracy in group STP is simply a reflection of a higher degree of learning in that group.

A final analysis was done to determine whether the source of the superior prediction accuracy in condition STP could be attributed to the subjects' memory for their recall performance on the test trial preceding the prediction trial. For each subject, items were classified into those recalled and not recalled on the last trial prior to the prediction trial. Then, a separate classification was made of the items recalled and not recalled on the criterion trial. The mean confidence judgment assigned to each of these four item types was then determined by summing across subjects and these means are presented in Table 4. The data for the criterion trial are simply the means of the components of the numerator of the CAQ score. The data for the trial preceding judgments of knowing can be considered a way of representing the subject's memory for remembered events if it is assumed that subjects are likely to have given a relatively high confidence judgment to an item on the prediction task if they remembered having recalled the item on the preceding test trial. When averaged across lists, there was a substantial difference between the mean confidence judgment assigned to recalled and nonrecalled items both when the classification was based on the preceding trial performance (5.48 vs. 2.47) and when it was based on the criterion trial performance (4.91 vs. 2.32). In fact, the difference was greater for the former classification. These findings suggest that a substantial source for the greater prediction accuracy in condition STP was the subjects' memory for their recall performance on the last trial prior to the prediction trial. The exact extent of the influence of this factor is difficult to specify because performance on the test trials preceding and following the prediction trial would be expected to be highly correlated.

DISCUSSION

The intent of the present research was to determine the influence of test-trial experience on the accuracy of subjects' judgments of knowing. In line with our expectations, prediction accuracy was consistently higher for those subjects who were given test trials prior to the prediction task. The critical role of retrieval opportunities for judgments of knowing was further evidenced in the decrease in prediction accuracy on the third list for those subjects who had test trials on the first two lists but not on the third list. A major source of the facilitation produced by test-trial experience was traced to subjects' knowledge of their previous recall performance. However, there was little or no improvement in prediction accuracy across lists regardless of whether or not test trials were given.

The results of this study clearly demonstrate that subjects can accu-

Table 4. Mean confidence judgments for recalled (R) and nonrecalled (NR) items on the trial preceding and following judgments of knowing in condition STP

	List 1		List 2	
	R	NR	R	NR
Preceding trial	5.41	2.48	5.55	2.46
Criterion trial	4.78	2.43	5.04	2.22

rately discriminate between known and unknown information. In fact, even though this discrimination was facilitated by test-trial experience, subjects who had only study trials showed a relatively high level of prediction accuracy. There are at least three possible bases for making accurate judgments of knowing. The first possibility was suggested by Arbuckle and Cuddy (1969) in their early studies on this problem, and it can perhaps best be described as a "direct-access" hyopthesis. They argued that if the memory trace is like other types of input signals, then subjects should be able to make accurate decisions simply by reading the strength of the appropriate trace. Although this hypothesis is intuitively appealing, we know of no evidence providing direct support for the hypothesis.

A second possible basis for judgment-of-knowing accuracy is memory for how a given item has been processed at the time of study and an appropriate mapping of these process factors onto the recall likelihood scale. For example, if a subject remembered forming an image for a given item, this item would be given a higher rating and would be likely to be recalled. The most extreme form of this process knowledge is provided when test trials are given. There is probably no better cue that an item is hard to remember than to have the learner fail to recall that item. If this is the fundamental basis of memory-monitoring accuracy, there is considerable slippage in subjects' mapping of the process factors onto recall likelihood. Zechmeister and Shaughnessy (1980) have found that subjects give equally high judgment-of-knowing ratings to items repeated according to massed and distributed schedules even though these two types of items are not equally likely to be recalled.

The third possible basis of memory-monitoring accuracy is the subjects' knowledge of the characteristics of the to-be-remembered material itself. The subjects in condition SOP of the present study may have been no more accurate in their judgments of knowing than a group of subjects might have been who had made their predictions before being given an opportunity to study the material, i.e., if the subjects made ease-of-learning judgments (Underwood, 1966). Nonetheless, the judgment-of-knowing accuracy in group STP of the present experiment can be added to the growing list of evidence that learners have considerable knowledge of the contents of their episodic memories. The list includes research on topics such as the feeling of knowing (e.g., Blake, 1973) and memory for remembered events (Gardiner & Klee, 1976).

In addition to its theoretical import, the study of memory-monitoring processes has clear implications for applied learning situations. The learner must not only master the material but must also make decisions about when study should be terminated or redirected. In general, we not only need to teach people various techniques for acquiring information, but we must also teach them ways in which they can evaluate the effectiveness of these techniques. The present research indicates that one component of such a memory-monitoring training program would be retrieval practice. For example, the recent work of Bjork and Geiselman (1978) suggests that the effectiveness of retrieval opportunities is dependent upon factors such as the delay between study and retrieval attempt. It may be that particular types of retrieval practice are ideally suited for developing memory-monitoring abilities.

Notes

Portions of this research were reported by King and Zechmeister at the Forty-seventh Annual Meeting of the Midwestern Psychological Association, 1975 and by all three authors at the Nineteenth Annual Meeting of the Psychonomic Society, 1978.

Requests for reprints should be sent to Eugene B. Zechmeister, Department of Psychology, Loyola University, 6525 North Sheridan Road, Chicago, IL 60626. Received for publication September 19, 1979; revision received November 27, 1979.

1. All reported analyses were also done using the Type II d' measure. The pattern of results described using the CAQ measure was obtained with the d' measure. The correlation of the d' and CAQ score were calculated separately for each list in each condition; and these ranged from $r\,(25) = .64$ to $r\,(25) = .90$, with a mean correlation across conditions and lists of $r\,(25) = .78$. Analyses of the CAQ measure are reported in the body of this report because this measure is more closely tied to the analytical analyses which were done to try to specify the sources of differences in prediction accuracy.

Reference note

1. Pasko, S. J. The judgment of knowing as related to practice, ease of learning judgments, and individual differences. Unpublished doctoral dissertation, Loyola University of Chicago, 1977.

References

Arbuckle, T. Y., & Cuddy, L. L. Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, 1969, *81*, 126–131.

Blake, M. Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 311–319.

Bjork, R. A., & Geiselman, R. E. Constituent processes in the differentiation of items in memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 347–361.

Conover, J. N., & Brown, S. C. Item strength and input location in free-recall learning. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 109–118.

Craik, F. I. M., & Watkins, M. J. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 599–607.

Gardiner, J. M., & Klee, H. Memory for remembered events: An assessment of output monitoring in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 227–233.

Gardiner, J. M., Passmore, C., Herriot, P., & Klee, H. Memory for remembered events: Effects of response mode and response-produced feedback. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 45–54.

Halff, H. M. The role of opportunities for recall in learning to retrieve. *American Journal of Psychology*, 1977, *90*, 383–406.

Klee, H., & Gardiner, J. M. Memory for remembered events: Contrasting recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 471–478.

LaPorte, R., & Voss, J. F. Nonspecific transfer as a function of number of lists acquired and criterion of learning. *American Journal of Psychology*, 1974, *87*, 463–474.

LaPorte, R., Voss, J. F., & Bisanz, G. Nonspecific transfer with and without test trials on the first list. *American Journal of Psychology*, 1974, *87*, 475–479.

Noble, C. E. Measurements of assoication value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, 1961, *8*, 487–521.

Paivio, A., Yuille, J.C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, 1968, *76*, (1, Pt. 2).

Rundus, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, *89*, 63–77.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 210–221.

Underwood, B. J. Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, 1966, *71*, 673–679.

Zechmeister, E. B. & Shaughnessy, J. J. When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, 1980, *15*, 41–44.

Zimmerman, J. Free recall after self-paced study: A test of the attentional explanation of the spacing effect. *American Journal of Psychology*, 1975, *88*, 277–291

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. *Intelligence*, 1977, *1*, 5–31.