Psychology and Aging
1989, Vol. 4, No. 2, 131–135

0882-7974/89/$00.75

Age Differences in Predictions and Performance on a Cued Recall Task

Raymond J. Shaw and Fergus I. M. Craik
University of Toronto
Toronto, Ontario, Canada

The notion that older adults fail to use optimal encoding strategies in memory tasks because of a deficit in memory monitoring was examined in a cued recall task. Participants were given different types of descriptors (initial letters, rhyme, category) for each word during encoding, and these descriptors were later given as cues at recall. Participants also predicted the likelihood of recalling each item. Age differences were found in recall performance, but there were no age differences in average predictions. However, the prediction ratings given by younger adults showed a greater difference between words recalled and words not recalled in the subsequent test than did the ratings of the older adults. Differences in recall associated with different types of processing were predicted poorly by all of the participants, but the relative recallability of specific words was well assessed by both age groups. We concluded that a memory monitoring deficit is not likely to be responsible for age differences in memory.

The general characteristics of age-related deficits in memory performance are well documented. One problem that has received a great deal of attention is the poorer performance of older persons in recalling items from lists of words or of paired associates (Craik, 1977; Kausler, 1982). One proposed explanation of this deficit is the demonstrated failure of older adults to use organizational or mediational strategies spontaneously in studying words for later recall (e.g., Hultsch, 1969, 1974). It has also been demonstrated that older subjects benefit disproportionately from using such learning strategies when they do so under explicit instructions (Hulicka & Grossman, 1967; Hultsch, 1969; Treat & Reese, 1976). Why then do they not use strategies spontaneously?

One reasonable hypothesis as to why older people do not use mediational strategies is that they suffer a deficit in metamemory. As Lovelace and Marsh (1985) pointed out, the term *metamemory* has been used to refer to several different aspects of a person's knowledge of his or her own memory processes. In the present article, we are concerned with possible age differences in monitoring encoding processes; in particular, we are concerned with (a) differences in the ability to judge the relative memorability of particular items, (b) how these judgments change as encoding processes are varied, and (c) the relation between prediction and performance. If older adults are less aware of the consequences of different types of processing on subsequent memory performance than are their younger counterparts, then it might not occur to the older learner to switch to a more effective coding strategy, even though he or she is perfectly capable of making such a switch. Although this hypothesis is a plausible one, it is not supported by the bulk of the evidence; several studies have shown no, or very slight, age differences in the memory monitoring functions of metamemory (Lachman, Lachman, & Thronesbury, 1979; Lovelace & Marsh, 1985; Perlmutter, 1978; Rabinowitz, Ackerman, Craik, & Hinchley, 1982).

Some points remain to be clarified, however. For example, at first sight it seems reasonable to measure some aspects of memory monitoring by the difference between predicted and actual memory performance (Lachman et al., 1979). Using this logic, Bruce, Coyne, and Botwinick (1982) found no age differences in predicted word recall in a list-learning experiment, but they did find the typical age-related decrement in performance. The difference between prediction and performance was therefore greater in the older group, and the authors concluded that metamemory accuracy declines with age. However, in the experiment by Rabinowitz et al. (1982), if one assumes that predictions (divided by 10) and performance may be compared directly, visual inspection of their Figures 1 and 2 shows that the discrepancy between predicted and actual performance was smaller for younger than for older participants under standard learning instructions, but with interactive imagery instructions the difference was smaller for the older group. This pattern obviously yields no clear conclusion about age differences in predictive accuracy. A preferable way to describe both the Bruce et al. and the Rabinowitz et al. results is to state that there were no age differences in the *predicted* memory scores, whereas memory performance varied as a function of age and experimental condition. The changing pattern of prediction–performance differences can be attributed to the insensitivity of adults of any age to processing variables, such as interactive imagery, that have marked effects on recall.

On the basis of the Rabinowitz et al. (1982) findings, we argue that prediction–performance differences do not remain stable across experimental conditions and that the difference is thus a poor index of age changes in memory monitoring. A preferable

The research reported in this article formed part of the master's thesis (University of Toronto, 1984) of the first author. The study was supported by a grant from the National Sciences and Engineering Research Council of Canada to the second author.

We are grateful to Lily Moysiuk and Janine Jennings for technical help and to two anonymous reviewers for their helpful comments on an earlier version of this article.

Correspondence concerning this manuscript should be addressed to Raymond J. Shaw, Department of Psychology, University of Toronto, Toronto, M5S 1A1, Canada.

approach is to study age differences in memory predictions and to examine age differences in the sensitivity of such predictions to variables that affect performance. In this regard, it is noteworthy that Rabinowitz et al. found both younger and older participants' predictions to be sensitive to changes in the associative strength between the stimulus and response members of word pairs, whereas neither age group's predictions were sensitive to the difference between standard learning and interactive imagery instructions. Speculatively, it seems possible that learners know that differences among *materials* will affect their later memory performance, but that they are unaware of the effects associated with different types of *mental operations.*

The present experiment was carried out to explore further the relations among aging, types of mental processing, memory predictions, and memory performance. It is known that differences in levels of processing are associated with large differences in subsequent recall and recognition (Craik & Tulving, 1975). In the present study, therefore, the type of processing at input was varied, and the effects of these variations on both predicted and actual recall were examined in younger and older participants. The use of orienting tasks during learning has the important consequence that encoding processes are largely controlled. In this way we can assess the participants' ability to monitor these different mental processes and to predict their effects on later memory performance. We do not of course expect lay persons to be familiar a priori with the effects of various orienting tasks; rather, we wish to observe their ability to assess the efficacy of the different tasks as they are being performed. In addition, the paradigm allowed us to assess participants' ability to predict recall on an item-by-item basis. As each word was presented for study, the participant rated the likelihood that he or she would recall it later; it was therefore possible to examine participants' sensitivity to processing variables by comparing predicted with actual recall across different types of processing and across age groups.

Method

Subjects

A total of 36 adults (18 in each of two age groups) participated in the experiment. The younger participants (M age = 19.4 years, SD = 1.25) were University of Toronto undergraduates, who received course credit for their participation; the older group (M age = 68.7 years, SD = 4.8) consisted of volunteers from the community. This second group scored higher (M = 16.0, SD = 2.3), $t(34) = 4.35, p < .01$, on a 20-item subset of the Mill Hill vocabulary test than did the younger group (M = 12.7, SD = 2.2). The older group also had a slightly greater number of years of formal education (M = 14.6 years, SD = 2.5), $t(34) = 2.25, p < .05$, than did the younger group (M = 13.2 years, SD = 0.79). Furthermore, the older subjects reported no unusual health problems. The higher vocabulary scores for the older sample, otherwise comparable with the younger group in terms of social and educational background, is typical in our experience.

Materials

A total of 60 common, concrete nouns were chosen so as to be unique within the list on three dimensions. No two started with the same first two letters, no two rhymed, and all were from separate semantic categories. Furthermore, all 60 words rhymed with at least four other common nouns in the English language and started with the same first two letters as at least three other nouns. Each word in the list was presented with one of three types of cue: a letter cue (e.g., "starts with ic: ice"), a rhyme cue (e.g., "rhymes with dice: ice"), or a category/descriptive cue (e.g., "something slippery: ice"). These same cues were used in the cued recall test. On the basis of previous work (e.g., Craik & Tulving, 1975), we expected that cued recall would be best for the category cues and poorest for the letter cues. There were 20 words of each type. Three lists were constructed so that each word appeared equally often across subjects in all three conditions. The items were presented in random order. Age was the only between-subjects factor.

Procedure

Participants were tested individually or in pairs. The list of words and the cued recall test were described with examples from each condition. The experimenter then told participants that he was interested in their ability to "rate their own memory," or to predict the likelihood of recalling a word they were studying. They were told not to rate the words on "some inherent memorizability" but to try to assess their own memory for the words and the likelihood of their recalling the words, given the cue. They were asked to give this assessment a rating from 0 to 10, with the lower end indicating certainty of not recalling the word and the upper end indicating certainty of recalling the word. Participants were encouraged to spread their predictions over the entire range of the scale.

A practice set of six items was presented orally by the experimenter. In the actual experiment, participants read the words with their cues from a typed sheet, one pair at a time. For each item, at the end of 10 s the experimenter indicated vocally that the participant should rate the item and move on to the next one. A mask was used so that participants could study only one item at a time. Items were presented in random order with respect to cue type.

After studying the list, participants spent 5 min working on a vocabulary test, after which they were given the cued recall test. The cues were presented on a typed sheet in a different random order than in the study phase, although participants could go through the list in any order, with no time limit.

Results

Figure 1 shows the probability of cued recall and mean recall predictions as a function of age and cue type. With regard to predictions, the figure shows that younger and older participants did not differ and that there was a slight increase in mean predictions from letter to rhyme to category cues. An analysis of variance (ANOVA) on predictions confirmed the lack of age difference, $F < 1.0$. There was a significant main effect of cue type, $F(2, 68) = 19.4, p < .01, \omega^2 = .12$. Newman-Keuls comparisons showed that participants made lower predictions for letters (5.6) than for either rhymes (6.4) or categories (6.7), but did not differentiate reliably between rhymes and categories. The interaction of age and cue type was not reliable, $F < 1.0$.

Cued recall performance is also shown in Figure 1. The figure demonstrates that, as expected, there was a dramatic increase in performance with deeper levels of processing. Consistent with notions of reduced age differences with increasing semantic guidance (Craik, 1986; Craik & Rabinowitz, 1984), the performance of older adults equaled that of their younger counterparts in the category/descriptive condition. These results were confirmed by an ANOVA. There was a significant main effect of age, $F(1, 34) = 5.92, p < .025, \omega^2 = .04$, and a significant main effect of cue type, $F(2, 68) = 130.65, p < .01, \omega^2 = .53$. The apparent interaction of age and cue type was also reliable, $F(2, 68) = 4.75, p < .025, \omega^2 = .02$. This interaction reflects the fact

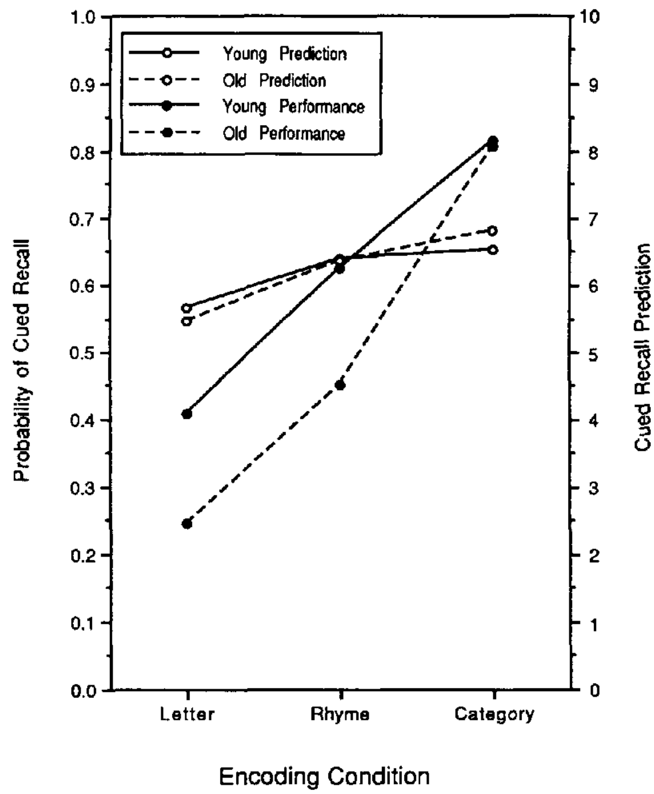

Figure 1. Probability of cued recall and cued recall predictions as a function of age and encoding condition.

that whereas, overall, young participants recalled significantly more words (.61) than did older participants (.50), the age difference in performance disappeared in the category condition.

A comparison of the results from predictions and recall performance shows that encoding conditions had a much larger effect on the latter measure (Figure 1). Although the measures are inherently rather different, it was considered of interest to assess the effect statistically after making several simplifying assumptions. A single ANOVA was therefore conducted on all the data points shown in Figure 1. To accomplish this, predictions were divided by 10 and treated as though they were estimates of proportions of words that subjects would recall. For the purposes of this secondary analysis, then, predictions and performance were considered as two levels of a single factor and as being on the same scale. Such a comparison assumes that subjects used the prediction scale as a way to estimate probability of recall; this was implied in the instructions, but not explicitly stated.

In the resultant Age × Cue Type × Prediction/Performance mixed-design ANOVA, the main effects of interest concern the prediction/performance factor and its interactions. The main effect of prediction/performance was reliable, $F(1, 34) = 5.14$, $p < .05$, $\omega^2 = .014$, reflecting the fact that predictions were higher than performance levels (.62 and .56, respectively). The interaction of age and prediction/performance was also reliable, $F(1, 34) = 4.60$, $p < .05$, $\omega^2 = .013$, reflecting the finding of no age difference in average predictions (.62 for both age groups), but a difference in performance scores (.62 for younger and .52 for older participants). It seems from this result that younger adults predicted their performance accurately, whereas older adults overestimated their performance by 10%. This straightforward interpretation is questioned, however, by the presence of a strong crossover interaction between cue type and prediction/performance, $F(2, 68) = 68.6$, $p < .001$, $\omega^2 = .113$. It is clear from Figure 1 that performance was much more affected by cue type than were predictions, and that both younger and older participants overestimated performance when it was low and underestimated performance when it was high. The three-way interaction of age, cue type, and prediction/performance only approached significance, $F(2, 68) = 2.42$, $p < .10$, $\omega^2 = .002$.

These results make it clear that the simple difference between prediction and performance is not particularly useful as a way of assessing memory monitoring because it can be either positive or negative depending on processing circumstances. This is an important point for future researchers to bear in mind because it may resolve apparent contradictions between different experimental results. It is also important to note that in the present results, older and younger adults were equally insensitive to the magnitude of changes in performance caused by processing manipulations. Apparently, this particular aspect of memory monitoring does not change with age. The finding that performance changes more than do predictions across types of processing confirms previous findings using the levels-of-processing paradigm (Cutting, 1975) and also findings contrasting interactive imagery instructions with standard learning instructions (Rabinowitz et al., 1982). However, it should be noted that in the present results predictions and performance showed the same general pattern: Participants were somewhat aware of the higher likelihood of recalling a word in the rhyme and category condition than in the letter condition.

The sensitivity of predictions for individual items was assessed by comparing predictions for items that were subsequently recalled to predictions for nonrecalled items. If participants were able to assess their encoding processes accurately, then predictions for items that were actually recalled should be higher than for items that were not recalled. It is also of interest to assess whether any such predictive ability varies as a function of either age or type of processing.

Table 1 shows the relevant data. Predictions for subsequently recalled items are higher than for items that were not recalled for both age groups; Table 1 also shows a tendency for predictions to rise from letter to category processing, as would be expected from Figure 1. These effects were assessed by a three-way ANOVA comparing age group (young/old), type of processing (letter/rhyme/category), and recall status (recalled/not recalled). The analysis showed that age was not a significant factor, $F < 1.0$, but that both type of processing, $F(2, 68) = 6.66$, $p < .01$, $\omega^2 = .032$, and recall status, $F(1, 34) = 56.8$, $p < .001$, $\omega^2 = .070$, were statistically reliable. The interaction between age and recall status was also significant, $F(1, 34) = 5.12$, $p < .05$, $\omega^2 = .005$, but no other interaction approached significance.

These more detailed results on individual item predictions thus show that young participants were better able than their elders to predict which items they would subsequently recall. Note, however, that this age difference is quite small; the mean predictions for recalled and unrecalled items were 6.6 and 5.5, respectively, for the younger group, compared with correspond-

Table 1
Mean Predictions and Standard Deviations for Recalled and Nonrecalled Items as a Function of Age and Type of Processing

	Young				Old			
	Recalled		Nonrecalled		Recalled		Nonrecalled	
Type of processing	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Letter cues	6.22	1.34	5.34	1.15	5.85	1.64	5.31	1.51
Rhyme cues	6.72	1.02	5.79	1.14	6.33	1.25	6.13	1.51
Category cues	6.80	1.36	5.35	1.95	6.94	1.40	6.23	1.67

ing values of 6.4 and 5.9 for the older group. The more striking finding is that participants of all ages were largely insensitive to the recall differences associated with different types of processing (Cutting, 1975; Rabinowitz et al., 1982). It is also of interest that the ability to discriminate potentially recallable items from nonrecallable items did not vary as a function of type of processing and that the age difference in discrimination did not change with processing type.

A final analysis was carried out to assess the nature of the information used by participants to make their predictions. Rabinowitz et al. (1982) suggested that in this situation people may use culturally shared information about the properties of words. Also, Rubin (1980) has shown that the relative recallability of words is rather stable across a wide range of experimental situations. If participants are sensitive to the properties of words that determine (in part at least) later memory performance, then there should be a positive correlation between the average prediction for a given word and its later probability of recall. Furthermore, if the ability to assess memorability in this way does not decline with age, the correlations between average predictions and average recall levels should be of the same order in the two age groups.

Accordingly, mean prediction scores and mean recall scores were calculated for each word by collapsing over cue conditions and over subjects. This procedure was carried out separately for the younger and older groups; its end result in each case was a set of 60 averaged predictions and a matched set of 60 averaged recall scores—a pair of scores for each of the 60 words used in the experiment. The results of the analyses were identical in the two age groups, $r(58) = .55$, $p < .001$, in both cases. That is, participants were quite sensitive to the inherent properties of words that influence recall, and this sensitivity showed no signs of changing with age.

Discussion

Several points can be made about the preceding results. First, the finding of no age difference in predicted memory performance confirms the observations made in previous studies (Bruce et al., 1982; Lovelace & Marsh, 1985; Rabinowitz et al., 1982). The lack of an age difference in predictions (see Figure 1) suggests that older subjects make the same predictions they would have made when they were 20 years old. The fact that the older participants gave the same predictions as their younger counterparts indicates some lack of awareness of their declining memory powers, although it should be emphasized that the present experimental situation is rather artificial and does not allow feedback about the discrepancy between prediction and later performance; also, of course, the present study was cross-sectional and not longitudinal in design. A second, related point is that the simple difference between prediction and performance is *not* a useful index of memory monitoring ability. In the present results, for instance, both younger and older adults underestimated recall in the category condition and produced a negative difference between prediction and performance. In the other conditions, the difference was positive and slightly smaller for younger adults. To say that older adults overestimate performance is therefore not always accurate. The changing pattern is understandable, however, in light of two empirical observations; first, that predicted levels do not change much with age and, second, that old and young adults are equally insensitive to the large effect of type of processing on subsequent recall. As Figure 1 illustrates, the discrepancy between mean prediction and mean performance is affected radically by the type of processing performed during encoding. Although under some circumstances young subjects may be more accurate than their older counterparts at predicting their level of performance (e.g. Bruce et al., 1982; Lovelace & Marsh, 1985; the rhyme encoding condition of the present study), this finding is apparently not a necessary one.

In line with the findings of Rabinowitz et al. (1982; see also Zechmeister & Shaughnessy, 1980), the present results suggest that people of all ages are largely unaware of memory effects associated with different mental processes but are somewhat sensitive to the effects associated with different materials. This third conclusion follows from two findings: First, both younger and older participants produced a highly significant correlation between the probability of recalling particular words and the average prediction for those words, and second, both age groups gave higher predictions to recalled than to nonrecalled items. It seems likely that both abilities stem from an assessment of the memorability of particular words, rather than from an assessment of the learner's specific mental operations on that trial. If the latter were true, then participants would have changed their predictions more radically as a function of level of processing.

The ability to discriminate between items that would or would not be recalled later was slightly greater for younger participants, but the effect was small. This result is in essential agreement with the findings of Lovelace and Marsh (1985) and of Rabinowitz et al. (1982). Unlike the results of Rabinowitz et al., however, the ability to discriminate between the two classes of items did not change as a function of type of processing—possibly because Rabinowitz et al. included a free learning condition, and this was the condition associated with relatively good predictability. The present experiment, in contrast, involved only cues presented by the experimenter; speculatively, free learning conditions may provide more scope for accurate predictions.

The present study is somewhat limited by the constraints on the processing of words provided by the study cues. It is unknown, for example, whether memory monitoring behavior is the same when processing is controlled as when it is free. That is, if memory monitoring is to benefit a person, then it should have an effect on the strategies a person selects while studying. When processing is controlled, the real effect of memory monitoring is not allowed to surface. If processing is uncontrolled by the experimenter, then strategy production deficits produced by a failure of memory monitoring are possible.

A modification of the present task could be used to examine this issue. For example, a second study–test trial, using new items, could be given to subjects after receiving feedback about performance on the first trial. If younger adults are better at monitoring memory, then their predictions should more closely match performance and still retain a good separation between predictions for recalled and unrecalled items. Alternatively, on the second study–test trial, subjects could be asked to select a cue for the later cued recall test from the three types of cue for each item. If older adults are less aware of the beneficial effects of more meaningful processing, they would be less likely than younger adults to choose the more meaningful category cues.

With regard to age differences in performance, note from Figure 1 that the present experiment provides a further example of an age-related difference in memory being reduced or eliminated under conditions of good environmental support at both encoding and retrieval (Craik, 1977, 1986; but see also Burke & Light, 1981). That is, when semantic category cues were provided at both input and retrieval, the age difference in performance disappeared.

Two major conclusions may be drawn from the present study that bear on aging research. First, as noted in a preceding section, the simple discrepancy between memory prediction and memory performance is inadequate, by itself, as an index of memory monitoring ability; the main reason for this is that the relation between these measures changes radically as a function of experimental condition. The second conclusion is that although a significant age difference in predictive ability was found in the present experiment, that difference was quite small. Also, participants in both age groups were strikingly poor at matching their predictions to the changes in performance associated with different types of processing. Although this ability represents only one aspect of the multidimensional construct of metamemory, it is reasonable to argue that it is one of the most relevant aspects for examining age differences in the use of learning strategies. The finding that adults of all ages are relatively unaware of processing effects makes it quite unlikely that age differences in memory performance can be attributed to a deficiency in the monitoring of encoding processes in older people. It seems, rather, that the memory deficits associated with aging should be attributed to age-related changes in the encoding and retrieval processes themselves.

References

Bruce, P. R., Coyne, A. C., & Botwinick, J. (1982). Adult age differences in metamemory. *Journal of Gerontology, 37,* 354–357.

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin, 90,* 513–546.

Craik, F. I. M. (1977). Age differences in human memory. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 324–420). New York: Van Nostrand Reinhold.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities. Mechanisms and performances* (pp. 409–422). Amsterdam: Elsevier.

Craik, F. I. M., & Rabinowitz, J. C. (1984). Age differences in the acquisition and use of verbal information: A tutorial review. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance, X* (pp. 471–499). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Cutting, J. E. (1975). Orienting tasks affect recall performance more than subjective impressions of ability to recall. *Psychological Reports, 36,* 155–158.

Hulicka, I. M., & Grossman, J. L. (1967). Age group comparisons for the use of mediators in paired-associate learning. *Journal of Gerontology, 22,* 46–51.

Hultsch, D. F. (1969). Adult age differences in the organization of free recall. *Developmental Psychology, 1,* 673–678.

Hultsch, D. F. (1974). Learning to learn in adulthood. *Journal of Gerontology, 29,* 302–308.

Kausler, D. H. (1982). *Experimental psychology and human aging.* New York: Wiley.

Lachman, J. L., Lachman, R., & Thronesbury, C. (1979). Metamemory through the adult life span. *Developmental Psychology, 15,* 543–551.

Lovelace, E. A., & Marsh, G. R. (1985). Prediction and evaluation of memory performance by young and old adults. *Journal of Gerontology, 40,* 192–197.

Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology, 14,* 330–345.

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology, 37,* 688–695.

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior, 19,* 736–755.

Treat, N. J., & Reese, H. W. (1976). Age, pacing, and imagery in paired-associate learning. *Developmental Psychology, 12,* 119–124.

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

Received October 26, 1987
Revision received May 20, 1988
Accepted May 23, 1988 ■